Structural studies of chlorophyll *f*-based photosynthesis

Imperial College London

Giovanni Consoli

Thesis submitted for the degree of Doctor of Philosophy

Department of Life Sciences
Faculty of Natural Sciences
2025

Statement of Originality

I, Giovanni Consoli, hereby declare that the work presented in this thesis is original and was done during my PhD at Imperial College London. The work was carried out directly by me under the supervision of Prof. A. William Rutherford, Prof. Benoit Chachuat and Dr. Andrea Fantuzzi. Contributions by collaborators are cited and acknowledged throughout the text.

Copyright Declaration

The copyright of this thesis rests with the author and its contents are made available under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND). Under this licence, you may copy and redistribute the material in any medium or format on the condition that you credit the author, do not use it for commercial purposes and do not distribute altered versions of the work. When reusing or sharing this work, ensure to make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

To my parents and my friends,
Thank you for making this life worth living

Alle volte uno si crede incompleto ed è soltanto giovane

Italo Calvino – *Il visconte dimezzato* (1952)

Acknowledgements

It still feels unreal that nearly four years have flown by since I moved to London to embark on this journey, which was made possible by three amazing people: Bill Rutherford, Andrea Fantuzzi, and Benoit Chachuat. Being able to work with you has changed my life, and I will always be grateful to you for this. You have always encouraged me to think, explore, and understand the world, in every scale and form. It is for this very approach, and your qualities, that I look up to you. Thank you, Andrea, for your humility, Benoit, for your kindness, and Bill, for your unwavering passion.

During these years I have made countless friends and travelled more than I ever had before, largely with the amazing group that joined the Digitalgaesation journey with me. To all of you, thank you for sharing those moments with me. Even when we were not travelling and actually tried to work, it never felt heavy, and I have to thank all of the colleagues of the 7th floor for that. A special thanks goes to Fiazall and Marco, for basically doing all of my work with me; to Geoff, for teaching me that solid priorities are what separates good scientists from good people; to Paul, for teaching me how to use the microscope and making it look easy; to Stefania, for answering each and every one of my stupid questions, but most of all for being a true friend; and to Ali, thank you for pushing me, even when I didn't want to, and thank you for trying to understand me.

I am infinitely grateful to all of the friends in the Nerdoni cinematic universe. I wish I could dedicate a few words to each and every one of you, but I have been told I must not go over 100,000 words. Thank you deeply for always being there to my lifelong friends Arci, Gando, Emi, Gian, Colo, Nico, Cape, Stefano, Gozzo, Simo, Marta, Ale, Giorgia, Nori, Porte and Frizz; and to the friends I have met in this new chapter: Sam, Marco, Hazel, Andrea, Silvia, Ferdi, Luca, Kate, Joe and Simone.

To Aneae, you've not been around as long, but honestly, it feels like you were always there. Your love and support have meant the world to me, you made every day feel like a field trip. Thank you, truly. And to Tomoko, thank you for treating me like I had always been there and for making me feel so incredibly at home. It has meant so much.

Finally, to my parents, Manu and Dario: thank you for your unwavering and continuous support. You have encouraged and motivated me every single day, without fail. I could never have achieved any of this without you. From the bottom of my heart, thank you.

Abstract

Photosynthesis is the central biochemical process that sustains life on Earth. It captures solar energy to drive charge separation and water oxidation, releasing molecular oxygen as a byproduct, energising the atmosphere, and enabling complex life. The reduced carbon compounds it produces support global food webs as biomass and, over geological timescales, have shaped human civilisation through the formation of fossil fuels. In recent decades, photosynthetic organisms capable of harvesting longer-wavelength photons have been identified, prompting academic interest in characterising their mechanisms and assessing their potential for use in photosynthetic bioprocesses. However, given the stringent energy requirements of canonical oxygenic photosynthesis, it is essential to assess the advantages and limitations of the evolutionary strategies that enable the same photochemical reactions to occur with less energetic photons. Understanding light harvesting, charge separation, and the spatial arrangement of long-wavelength pigments requires structural insights into the light-driven components of far-red-light-adapted photosynthetic machinery.

In this work, the structures of Photosystem I and Photosystem II from the chlorophyll *f*-containing cyanobacterium *Chroococcidiopsis thermalis* PCC 7203 were determined using single-particle cryo-electron microscopy. The locations of all eight chlorophyll *f* sites in FR-PSI, including the A_{-1B} reaction centre pigment, and all four chlorophyll *f* sites plus the chlorophyll *d* primary donor in FR-PSII, are proposed based on chemical environments, conserved sequence changes, and detailed analysis of electrostatic potential maps. This study also reports the identification and structural characterisation of a novel FR-PSII subunit encoded within the FaRLiP cluster.

The structure of the intact far-red-light-adapted allophycocyanin antenna has also been solved, and the evolutionary adaptations that allow efficient management of excitation energy transfer in the far-red-light-adapted photosynthetic apparatus discussed. Finally, the phytochrome responsible for regulating far-red light photoacclimation has been expressed and a mathematical model that describes its behaviour under varying light conditions is presented.

Table of contents

Statement of Originality i

Copyright Declaration i

Acknowledgements iii

Abstract iv

Table of contents v

List of Figures x

List of Tables xiv

Abbreviations xv

1 Introduction 1

1.1 Photosystem II 3

1.2 Photosystem I 6

1.3 Excitation energy transfer in photosynthetic pigments 7

1.4 Antenna complexes in cyanobacteria and plants 12

1.5 Thylakoid membrane organisation in cyanobacteria 14

1.6 Plant biotechnologies and microalgal cultivation 16

1.7 Long-wavelength photosynthesis 18

1.7.1 Chlorophyll *d* containing cyanobacteria 19

1.7.2 Chlorophyll *f* containing cyanobacteria 20

1.8 Cryo electron microscopy in photosynthesis and bioenergetics 23

1.9 Overview on transmission electron microscopy 24

1.10 Cryo-EM single particle analysis 25

1.11 Towards atomic resolution in CryoEM 27

1.12 The objectives of this work 30

2 Materials and methods 32

2.1 Far-red light culture growth of *C. thermalis* PCC 7203 32
2.2 Absorbance measurements 32
2.3 Isolation of FR-PSI with sucrose gradient 32
2.4 Isolation of FR-PSII and FR-APC complexes 33
2.5 Grid preparation 33
2.5.1 Collection A 33
2.5.2 Collection B 34
2.5.3 Collection C 34
2.5.4 Collection D 34
2.6 Data acquisition of FR-PSI 34
2.6.1 Collection A 34
2.6.2 Collection B 35
2.6.3 Collection C 35
2.6.4 Collection D 35
2.7 Single particle analysis 36
2.7.1 Collection A 36
2.7.2 Collection B 37
2.7.3 Collection C 38
2.7.4 Collection D 38
2.8 Model building 39
2.8.1 FR-PSI 39
2.8.2 Holo-FR-PSII 39
2.8.3 Apo-FR-PSII 39
2.8.4 FR-APC 39
2.9 De novo sequence identification of PsbH2' 40
2.10 77 K fluorescence spectroscopy of APC complexes 40
2.11 Statistical analysis of Chl substituents ESP 40
2.12 *E. coli* transformation 41
2.13 RfpA heterologous expression and purification 42
2.14 SDS polyacrylamide gel electrophoresis 42
3 Methods to investigate chlorophyll densities in Cryo-EM 43

3.1 Introduction 43
3.2 Methodology 49
3.2.1 Improvement of the cone scan method 49
3.2.2 Assessment of the quality of the cone scan for a given map/model 51
3.2.3 The sphere method, a multivariate approach in Chl classification 53
3.3 Results and Discussion 54
3.3.1 Resolution dependence of classification metrics 54
3.3.2 Further challenges with mislabelling of chlorophyll substituents 56
3.3.3 Analysis of the variance of Chl substituents ESP through PCA 58
3.3.4 Classification with SVM of the sphere scan sampling 60
3.4 Conclusions 64
4 Locating the missing chlorophyll *f* molecules in Cryo-EM structures of FR-PSI 66
4.1 Introduction 66
4.2 Results 68
4.2.1 Chl *f* A_{-1B} 74
4.2.2 Chl *f* A21 77
4.2.3 Chl *f* A23 78
4.2.4 Chl *f* B07 79
4.2.5 Chl *f* B30 80
4.2.6 Chl *f* B37 82
4.2.7 Chl *f* B38 83
4.2.8 Chl *f* L03 84
4.3 Discussion 85
4.3.1 Comparison with the structures in literature 85
4.3.2 DFT calculation of Chl A_{-1B} chemical environment 86
4.3.3 Chl *f* in the A_{-1B} position reproduces the experimental difference spectra 88
4.3.4 Implications and tuning of a Chl f as a primary donor 91
4.3.5 PsaL2 controls trimerization and L03 is the sharing pigment 93
4.3.6 Partial occupancy in far-red light photosystems 95
4.3.7 Spectral assignment of Chl *f* 97
4.4 Conclusions 99
5 A new subunit and chlorophyll *f* assignments in CryoEM structures of FR-PSII 101
5.1 Introduction 101

5.2 Results 103
5.2.1 Chl *f* C507 111
5.2.2 Chl *f* B608 112
5.2.3 Chl *f* B614 113
5.2.4 Chl *f* B617 114
5.2.5 On the other proposed Chl *f* candidates 115
5.2.6 A new FR-PSII specific subunit: PsbH2' 116
5.2.7 Phylogenetic analysis of PsbH2 and PsbH2' 117
5.2.8 Confirming B617 is the redmost Chl of WL PSII 118
5.2.9 Structural differences between broken and intact FR-PSII 119
5.3 Discussion 121
5.3.1 Function of the new Subunit PsbH2' 121
5.3.2 Chl *f* B617 in comparison with WL-PSII 124
5.3.3 Nature of the reaction centre long-wavelength chlorophyll: Chl *d* or Chl *f*? 126
5.3.4 Structural changes on the donor and acceptor side of FR-PSII 127
5.3.5 Changes in the side path and ChlZ 129
5.3.6 Energy landscape in FR-PSII dimers 131
5.4 Conclusions 134
6 Cryo-EM structure of FR-APC and evolutionary adaptations to far-red light 135
6.1 Introduction 135
6.2 Results 138
6.2.1 Absorption and low-temperature fluorescence spectroscopy 142
6.2.2 Planarity of bilin sites 143
6.2.3 Changes in FR-APC subunits lead to changes in the absorbance spectra 145
6.2.4 Absence of the distal ApcC 150
6.2.5 Dimerization interface between the core cylinders 151
6.3 Discussion 153
6.3.1 Array arrangement of PSII 153
6.3.2 Terminal emitter of FR-APC and the injection Chl *f* in FR-PSII 155
6.3.3 PsbH2' and PsbH2 as a potential stabilizing structure of FR-APC 157
6.4 Conclusions 159
7 Towards understanding RfpA dynamics 161
7.1 Introduction 161

7.2 Results 165
7.2.1 Photobioreactor setup 165
7.2.2 Heterologous expression of RfpA 166
7.2.3 His-tag purification of RfpA 168
7.2.4 Absorbance spectra and photoswitching of the GAF domain 169
7.2.5 Attempts to determine the Cryo-EM structure of full length RfpA 170
7.2.6 Mathematical modelling of the influence of RfpA on FaRLiP 171
7.3 Discussion 174
7.3.1 Towards model calibration 174
7.3.2 Further expression trial of full length RfpA 176
7.4 Conclusions 178
8 Conclusions and Outlook 180
9 Bibliography 183

List of Figures

Fig. 1.1 – Photosynthetic electron transport chain 2

Fig. 1.2 – Reaction centre of PSII 4

Fig. 1.3 – Reaction centre of PSI 6

Fig. 1.4 – Spectral overlap and reciprocal orientation for resonance energy transfer 9

Fig. 1.5 – Plant and cyanobacterial antenna complexes 13

Fig. 1.6 – Substituents to the tetrapyrrole ring in different Chl types 19

Fig. 1.7 – Structure of the FaRLiP cluster 21

Fig. 3.1 – Chlorophyll substituent ESP cones 50

Fig. 3.2 – Illustration of the area investigated by the sphere scan method 53

Fig. 3.3 – Accuracy, precision and recall metrics of available FR-PSI structures 54

Fig. 3.4 – Relationship between maximum accuracy, map resolution, and Q-score 55

Fig. 3.5 – Example of false positive in the periphery of FR-PSI 57

Fig. 3.6 – Scatterplot matrix of the first four principal components derived from PCA ... 58

Fig. 3.7 – PCA of FR-PSI chlorophyll substituents 59

Fig. 3.8 – Confusion matrices of different SVM classifiers 61

Fig. 4.1 – Local resolution of FR-PSI maps 68

Fig. 4.2 – Structure of trimeric FR-PSI and chlorophyll *f* locations. 73

Fig. 4.3 – Evidence for A_{-1B} being a Chl *f* 74

Fig. 4.4 – Structural comparison of Helix N PsaA and PsaA2 in the region A_{-1B} 75

Fig. 4.5 – Conserved changes near Chl A_{-1B} influence the position of the formyl group .. 76

Fig. 4.6 – Chl *f* A21 site 77

Fig. 4.7 – Chl *f* A23 site and variability among far-red strains 79

Fig. 4.8 – Chl *f* B07 site..... 80

Fig. 4.9 – Chl *f* B30 site..... 81

Fig. 4.10 – Chl *f* B37 site..... 82

Fig. 4.11 – Chl *f* B38 site..... 83

Fig. 4.12 – Chl *f* L03 site..... 84

Fig. 4.13 – ESP at the C2 position for the A_{-1B} chlorophylls of FR-PSI and WL-PSI..... 85

Fig. 4.14 – DFT-calculated ESP of Chl *a* and Chl *f* in the A_{-1B} site 87

Fig. 4.15 – Simulations of the $P700^+$ *minus* P700 difference spectra..... 89

Fig. 4.16 – Excitation energy transfer routes in FR-PSI 93

Fig. 4.17 – PsaL2 and PsaB2 prevent the formation of chimeric FR/WL-PSI 94

Fig. 4.18 – Wavelength assignments in FR-PSI..... 98

Fig. 5.1 – Local resolution of FR-PSII maps 103

Fig. 5.2 – Positions of the Chl *f* candidates in FR-PSII 110

Fig. 5.3 – Chl *f* C507 site 111

Fig. 5.4 – Chl *f* B608 site..... 112

Fig. 5.5 – Chl *f* B614 site..... 113

Fig. 5.6 – Chl *f* B617 site..... 114

Fig. 5.7 – Position of PsbH2' in the dimeric FR-PSII ESP map 116

Fig. 5.8 – Phylogenetic groups of PsbH2 and PsbH2' sequences. 117

Fig. 5.9 – 77 K fluorescence emission in the PsbH Thr to Ala *C. reinhardtii* mutant 118

Fig. 5.10 – Comparison between Apo- and Holo-FR-PSII 119

Fig. 5.11 – Alterations to Cyt b_{559} in Apo-FR-PSII 120

Fig. 5.12 – Potential function of PsbH2' 122

Fig. 5.13 – Cone scan analysis of the substituents of Chl_{D1} 127

Fig. 5.14 – FR-PSII donor side conserved changes 128

Fig. 5.15 – FR-PSII acceptor side conserved changes 129

Fig. 5.16 – Long-wavelength pigments landscape of FR-PSII 131

Fig. 6.1 – Local resolution of bicylindrical FR-APC map 138

Fig. 6.2 – Positions of the bilin pigments in the FR-APC 141

Fig. 6.3 – Absorption and fluorescence spectra of the bicylindrical FR-APC complex .. 143

Fig. 6.4 – Analysis of the planarity of bilin pigments in FR-APC and WL-APC 144

Fig. 6.5 – Protein environment of the ApcE2 and ApcF2 subunits 146

Fig. 6.6 – Changes in the bilin sites of ApcD2, ApcD3 and ApcD5 148

Fig. 6.7 – Clashes between distal ApcC subunit and ApcE2 150

Fig. 6.8 – Differences in the ApcB2 bilin site in the presence or absence of ApcC 151

Fig. 6.9 – Interfaces stabilizing the bicylindrical FR-APC 152

Fig. 6.10 – Mesoscale arrangement of FR-PSII and FR-APC in thylakoids 154

Fig. 6.11 – Emitter-injector distances in a PSII dimer array assembled with FR-APC ... 156

Fig. 6.12 – Potential interactions between PsbH2 and PsbH2′ with ApcD3 and ApcE2. 158

Fig. 7.1 – LEDs emission profiles ... 165

Fig. 7.2 – Dual illumination photobioreactor setup ... 166

Fig. 7.3 – Plasmids for the expression of Holo-RfpA ... 167

Fig. 7.4 – Synthesis of phycocyanobilin starting from Heme *b* ... 167

Fig. 7.5 – SDS PAGE gel of the expression of N-RfpA and F-RfpA ... 168

Fig. 7.6 – Absorbance spectra of the Pr and Pfr states and photoconversion rates ... 169

Fig. 7.7 – Low-resolution model of RfpA ... 170

Fig. 7.8 – Illustration of the three-state RfpA model ... 171

Fig. 7.9 – Predicted state populations from the RfpA model ... 174

List of Tables

Table 1 - SVM classification results of Chl *f* candidates in the FR-PSI 62

Table 2 – Data collection and processing parameters for trimeric FR-PSI (9I9L) 70

Table 3 – Data collection and processing parameters for trimeric FR-PSI (9EYS)..... 72

Table 5 – Data collection and processing parameters for dimeric FR-PSII..... 105

Table 6 – Data collection and processing parameters for monomeric FR-PSII 107

Table 7 – Data collection and processing parameters for monomeric Apo-FR-PSII..... 109

Table 4 – Data collection and processing parameters for FR-APC complex (9I1R)..... 140

Table 8 – FR-PSI Chl *f* sites assignments 213

Table 9 – FR-PSII Chl *f* sites assignments..... 214

Abbreviations

APC	Allophycocyanin
Chl	Chlorophyll
Cryo-EM	Cryo electron microscopy
Cryo-ET	Cryo electron tomography
CTF	Contrast transfer function
Cyt b_{559}	Cytochrome b_{559}
Cyt b_6f	Cytochrome b_6f
DFT	Density functional theory
EMDB	Electron Microscopy Data Bank
EPR	Electron paramagnetic resonance
ESP	Electrostatic potential
FaRLiP	Far-red light photoacclimation
FR-APC	Far-red-light-adapted allophycocyanin
FR-PSI	Far-red-light-adapted Photosystem I
FR-PSII	Far-red-light-adapted Photosystem II
GS-FSC	Gold standard Fourier-shell correlation
PBS	Phycobilisome
PCB	Phycocyanobilin
PDB	Protein Data Bank
RfpA	Regulator of far-red light photosynthesis A
ROS	Reactive oxygen species
SPA	Single particle analysis
WL-PSI	White-light-adapted Photosystem I
WL-PSII	White-light-adapted Photosystem II

1 Introduction

Oxygenic photosynthesis is the process, common to all plants and cyanobacteria, through which photons are used as an energy source and water as an electron source to generate reduced carbon species from CO_2. It is one of the most important biological processes to have ever evolved, granting oxygenic autotrophs a widespread and readily available electron source, while producing molecular oxygen as a byproduct. Molecular oxygen has energised the atmosphere, allowing complex life to evolve, while ~2.4 billion years of photosynthetic biomass stored in the form of fossil fuels has allowed the development of our societies and ultimately our modern lifestyles.

The process of photosynthesis can be split into two main phases commonly known as the light and the dark reactions of oxygenic photosynthesis. Although there are many differences between the photosynthetic processes of cyanobacteria, algae, and higher plants (such as structural organisation (1), light harvesting (2) and regulation (3)), all oxygenic photosynthetic organisms conduct the light reactions in membrane-enclosed compartments known as thylakoids and utilise the same electron transport chain. The four major components of the electron transport chain in oxygenic phototrophs are: Photosystem I (PSI), Photosystem II (PSII), the cytochrome b_6f complex (Cyt b_6f), and the ATP-synthase.

During the light reactions, photons are harvested by antenna pigments, and their energy is used to drive charge separation in the reaction centre of the two light-driven complexes of the electron transport chain, PSII and PSI. The electrons, initially stripped from water molecules by PSII, are used to reduce plastoquinones that migrate within the membrane to the Cyt b_6f complex, which catalyses electron transfer to a soluble lumenal electron carrier called plastocyanin. For every electron transferred from a reduced plastoquinone to plastocyanin, one proton is taken up from the cytosol and one proton is released into the lumen, converting the redox energy from this reaction to an electrochemical gradient. Plastocyanin acts as an electron donor for PSI that, thanks to the energy of the harvested

photons, transfers them to ferredoxin, which in turn reduces ferredoxin:$NADP^+$ reductase, leading to the formation of NADPH.

Throughout the electron transport process, protons are taken up from the cytosol and released by PSII and Cyt b_6f, leading to the acidification of the lumen and the formation of a proton motive force across the membrane, ultimately driving ATP production by ATP synthase.

The final stoichiometric balance of the light reaction of photosynthesis is the following:

$$12H_2O + 12NADP^+ + 18ADP + 18P_i \rightarrow 6O_2 + 6H_2O + 12NADPH + 18ATP$$

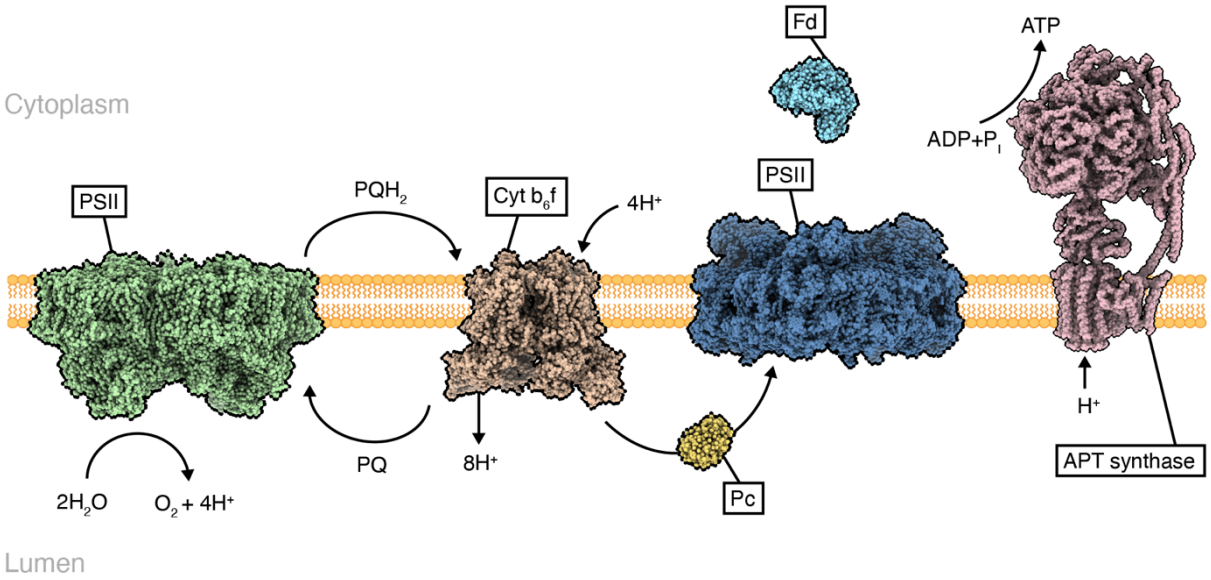


Fig. 1.1 – Photosynthetic electron transport chain

Simplified representation of the photosynthetic electron transport chain. Photosystem II (PSII) is depicted in green, Cyt b_6f in orange, Photosystem I (PSI) in blue, plastocyanin in yellow (Pc), ferredoxin in light blue (Fd) and the ATP synthase in pink.

In the dark reactions, which do not require any photons, energy in the form of ATP, and reducing power provided by NADPH, are used by Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase) to catalyse the reduction of CO_2, producing reduced carbon molecules such as sugars (4).

1.1 Photosystem II

Photosystem II (PSII), also known as the water-splitting enzyme or water/plastoquinone photo-oxidoreductase, is the first complex in the photosynthetic electron transport chain. It typically exists *in vivo* as a dimer, each monomer containing ~20 subunits. PSII typically contains 35 chlorophylls (Chl), 2 pheophytin *a*, 2 plastoquinones, 8 β-carotenes, one heme *b*, one heme *c*, one Mn_4CaO_5 cluster (also known as the oxygen evolving complex), two Cl^-, one bicarbonate, one non-heme Fe^{2+} (5,6). The core of PSII is composed by the homologous D1 and D2 subunits (the gene products of *psbA* and *psbD*, respectively). These two proteins have five transmembrane helices each, and constitute the reaction centre of PSII, in which all of the cofactors participating in the electron-transfer and water-splitting reactions are associated. More specifically, the reaction centre of PSII binds four chlorophyll molecules, P_{D1}, P_{D2}, Chl_{D1} and Chl_{D2} and two pheophytins ($Pheo_{D1}$ and $Pheo_{D2}$) (Fig. 1.2). On the acceptor side of PSII, two plastoquinone molecules are bound to D2 and D1 respectively, called Q_A and Q_B (Fig. 1.2). Between them there is a non-heme iron, to which bicarbonate is bound. On the donor side of PSII, a redox active tyrosine (Tyr_{ZD1}) participates in electron transfer reactions from the Mn_4CaO_5 cluster (Fig. 1.2).

Wrapped around the core, the CP47 and CP43 subunits are found (the gene products of *psbB* and *psbC*, respectively). CP47 and CP43 have six transmembrane helices and bind respectively 16 and 13 chlorophylls with light-harvesting function (7,8).

Despite more than three billion years of evolution in multiple ecological niches, and therefore very diverse needs in terms of bioenergetics, all organisms capable of performing oxygenic photosynthesis share the same inorganic core and a remarkably conserved sequence for most of the protein subunits (9–11). The ability to use water as a source of electrons is an evolutionary singularity and gives perspective on how complex and unique the reaction performed by PSII is. To understand it, the water splitting enzyme can be divided in three main components: the charge separating enzyme, the proton coupled quinone reduction enzyme, and the water oxidizing enzyme. These three act in concert to perform multielectron redox chemistry on both the donor and the acceptor side of PSII.

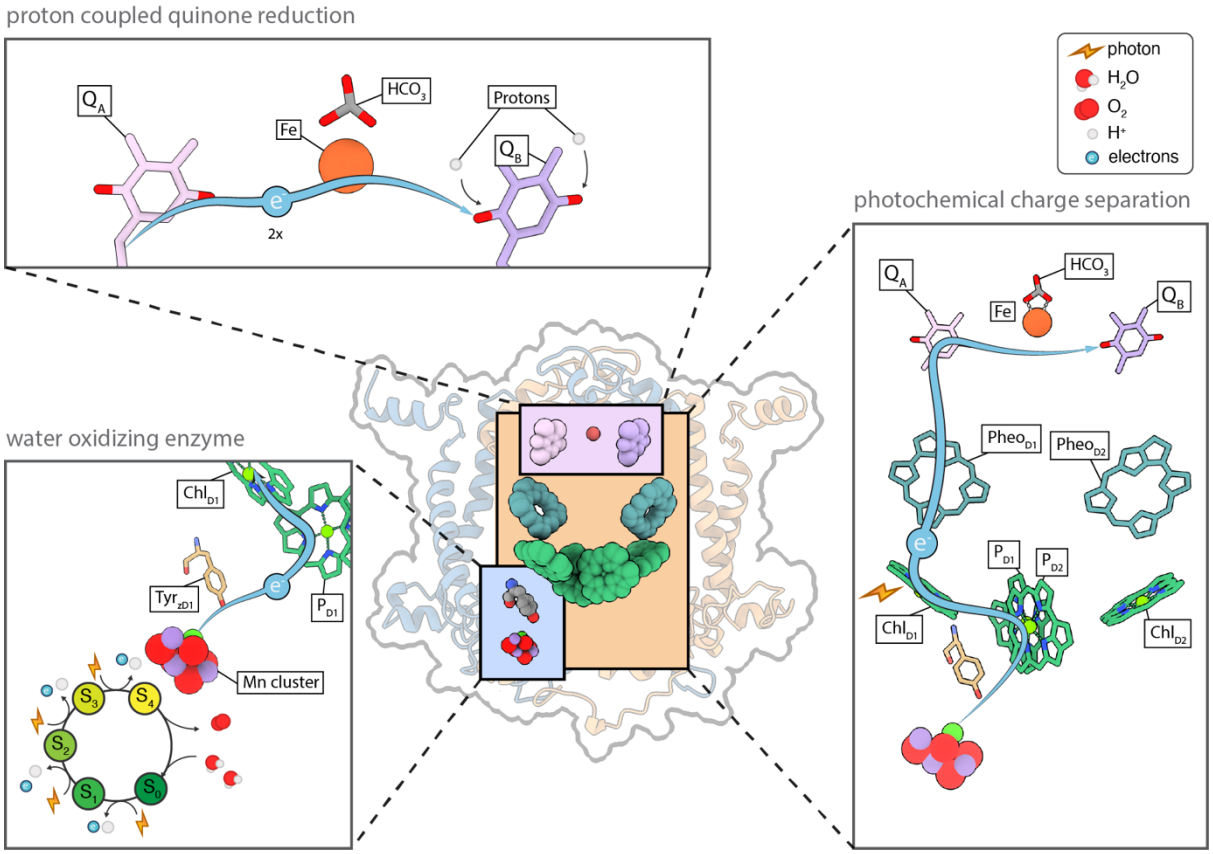


Fig. 1.2 – Reaction centre of PSII

In the centre, a cartoon representation of the key elements of PSII. The orange box outlines the elements involved in charge separation, and the path of the electrons from the donor water molecules to the final acceptor Q_B. In pink, the elements crucial for the for the proton coupled quinone reduction of Q_B. and in blue a brief schematic of the S4 cycle of the oxygen evolving complex.

Upon absorption of a photon or excitation energy transfer to the primary electron donor (P680), charge separation takes place and one electron is donated to the initial acceptor pheophytin, which subsequently transfers it to the primary quinone (Q_a) forming the stabilised radical pair P_{D1}^+ / Q_a^- in around 300 ps (12,13). The spatial separation between the electron and the hole, and the large losses in free energy at each step of the electron transfer, contribute to make the charge separated state metastable, and preventing damaging back-reactions (14,15). Q_a^- then donates the electron to the secondary (Q_b) plastoquinone acceptor, and once a second electron is donated to Q_b^-, it gets protonated form Q_bH_2 and is released allowing a new oxidised Q_b to be reincorporated in a thermodynamically favourable manner (16,17). This leads to the following acceptor side half reaction:

$$2PQ + 4e^- + 4H^+(stroma) \rightarrow 2PQH_2$$

At the same time, the hole left on $P680^+$ has to be regenerated by a nearby tyrosine residue, PsbA Tyr^{161}, known as Tyr_z, which in turn triggers the catalytic advancement of a Mn_4CaO_5 metallocofactor known as the oxygen-evolving complex (OEC) . Once four electrons have been abstracted from the Mn_4CaO_5 cluster, two water molecules release four protons and one oxygen molecule (7,18–20). This process is known as the Joliot - Kok cycle, every oxidation step of the Manganese Cluster is known as S state and is assigned a number from 0 to 4 depending on the oxidation state of the cluster (21,22). We can define the chemistry happening on the donor side with the following half reaction:

$$2H_2O \rightarrow O_2 + 4e^- + 4H^+(lumen)$$

S_1 is the dark stable state and is formed in 100% of centres in dark adapted samples. S_2 and S_3 are metastable and decay back to S_1 if given enough time. The S_4 state only exists for the time necessary for the rapid oxidation of the two water molecules bound to the Mn_4CaO_5 cluster, after which molecular oxygen is released, resetting it to the S_0 state. With the catalytic advancement of the cluster from S_0 to S_4, the protons from the two water molecules are gradually released on the lumenal side in the following ratio between the different states 1 : 0 : 1 : 2.

1.2 Photosystem I

Photosystem I (PSI) is the second light-driven protein complex in the photosynthetic electron transport chain. It catalyses light-driven oxidation of plastocyanin, a soluble copper electron carrier on the luminal donor side, and the reduction of ferredoxin on the stromal acceptor side, generating one of the most negative redox potentials in nature (23,24). PSI is found in different oligomerisation states, from monomers to tetramers, throughout oxygenic photosynthetic organisms (25). In cyanobacteria it is usually found in trimeric assemblies.

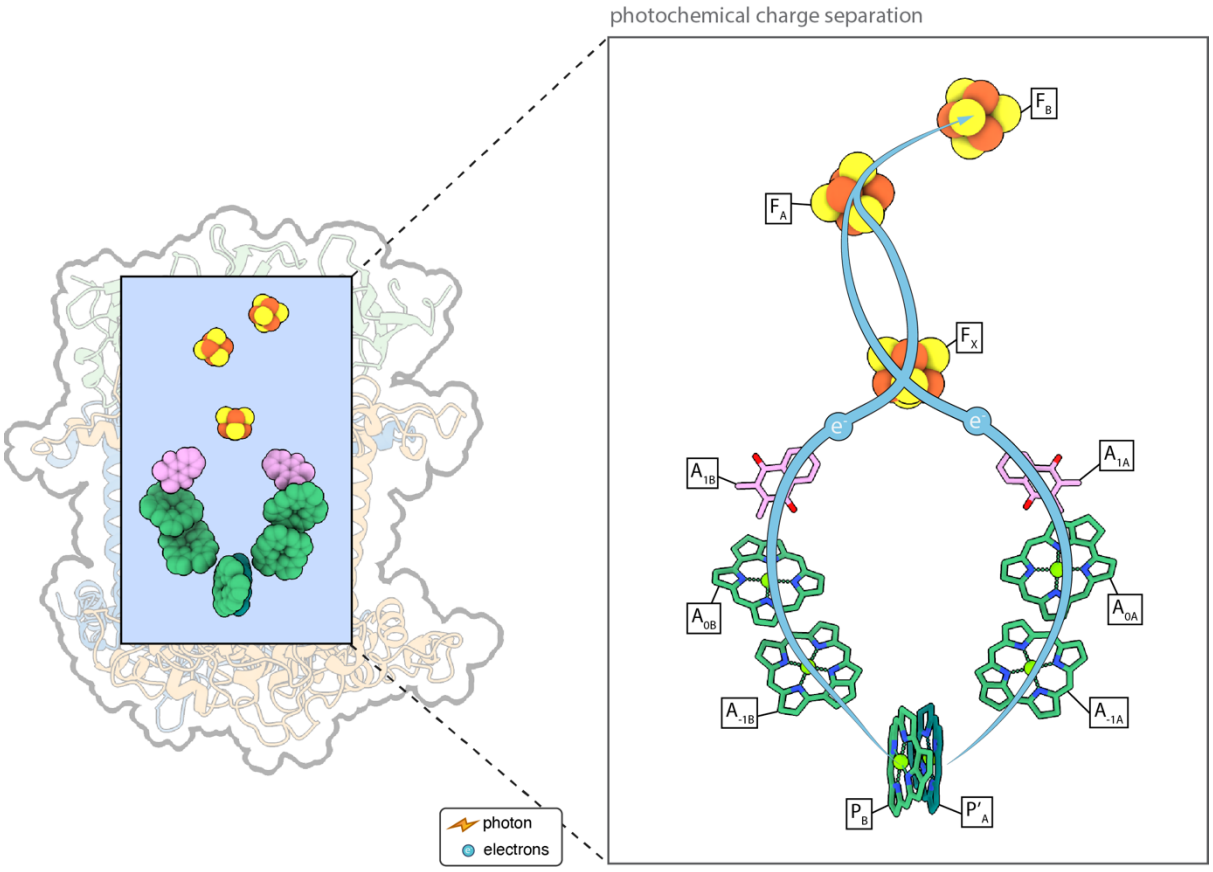


Fig. 1.3 – Reaction centre of PSI

On the left a cartoon of the PSI reaction centre in its protein matrix between PsaA and PsaB, on the right a close up on the main elements involved in charge reparation. Chl *a* molecules are depicted in green, the epimeric Chl *a*' (P'A) is represented in a darker shade of green, the phylloquinones are represented in pink and the 4Fe-4S clusters are presented in yellow (S) and orange (Fe).

A monomer is typically composed of 12 subunits and contains approximately 20 carotenoids and 90 chlorophyll molecules, making it one of the most complex networks of

excitation energy transfer in nature, as well as 2 phylloquinones and 3 FeS clusters (24). As in the case of PSII, the sequence of the reaction centre subunits is largely unchanged throughout different organisms with a high degree of structural conservation (26). PSI's reaction centre is composed of two protein subunits, A and B (the gene products of *PsaA* and *PsaB*, respectively). The reaction centre of PSI binds six chlorophyll molecules, a special pair of chlorophylls called P700, based on the absorption peak of this coupled dimer, the "accessory" chlorophylls A_{-1A}/A_{-1B}, electron-acceptor chlorophylls A_{0A}/A_{0B}. Moreover, it binds two phylloquinones known as A_{1A}/A_{1B}, and an iron–sulfur (4Fe-4S) cluster known as F_X (27). Key to the functionality of this system are the cytoplasmic subunits PsaC, PsaD, and PsaE which coordinates two other iron-sulfur clusters known as F_A and F_B (24,28).

After absorption of a photon by any of the chlorophylls in the complex, excitation is efficiently transferred to the PSI electron donor (P700). Remarkably, while P_B is a Chl *a*, P_A is the 13^2-epimer of Chl *a* (Chl *a′*), constituting a Chl *a′*/Chl *a* heterodimer (24,27). Differently from PSII, where just the D1 branch is actively involved in the primary charge separation, in PSI the electron can be donated to the two symmetric A_0 chlorophylls, coordinated by the two core subunits A and B. Charge separation is stabilised by subsequent electron transfer to one of the two phylloquinones. Subsequently, the electrons are donated to the three 4Fe-4S clusters F_X, F_A, and F_B (29,30). Due to differences in the reduction potentials of the phylloquinones, the forward electron transfer rates from A_1 to F_X differ by an order of magnitude, ~200 ns for A_{1B} ~20 for A_{1B} (31,32). The functional significance of the asymmetry is still speculative but has been rationalised as a mechanism to avoid the formation of chlorophyll triplet (^{3}Chl) and consequently of singlet oxygen ($^1O_2$) and reactive oxygen species (14,33).

1.3 Excitation energy transfer in photosynthetic pigments

Photosynthetic complexes need to specialise and adapt specific chlorophylls to play two major roles. The first adaptation is the one employed by the chlorophylls that constitute reaction centres (11,34), which are capable of converting the energy of their excited state into a charge separated state to drive photochemistry. The second equally important

adaptation is the role that most pigments have in photosynthetic complexes: harvesting photons and funnelling excitation energy to the reaction centre chlorophylls, thereby increasing their nominal absorption cross section (35–37).

The migration of excitation energy from light-harvesting chlorophylls to reaction centre chlorophylls relies on excitation energy transfer, a physical phenomenon in which the exciton, the entity constituted by an excited electron and hole pair, is transferred from an excited chromophore, referred to as donor, to another chromophore, referred to as acceptor (35). This phenomenon originates from a quantum mechanical resonance interaction between the transition dipole moments of the chromophores.

$$|D'A\rangle \rightarrow |DA'\rangle$$

As explained by Förster (38,39) in one of the earliest formulations of the theory of excitation energy transfer, the rate of energy transfer ($k^{\text{Förster}}$) between a donor and acceptor chromophore is determined by several parameters: the donor's excited state lifetime (τ_D), the donor's quantum yield of fluorescence (ϕ_D), the distance between the two chromophores (R), the relative orientation of the donor–acceptor pair (κ), and the overlap integral between the fluorescence emission spectra of the donor and the absorbance spectra of the acceptor (J_F). With N_A being the Avogadro's number and n the refractive index of the medium.

$$k^{\text{Förster}} = \frac{1}{\tau_D} \frac{9(\ln 10)\kappa^2 \phi_D J_F}{128\pi^5 N_A n^4} \frac{1}{R^6}$$

The relative orientation of the dipole moments of the two molecules can be described as:

$$\kappa = \hat{\mu}_D \cdot \hat{\mu}_A - 3\left(\hat{\mu}_D \cdot \hat{R}\right)\left(\hat{\mu}_A \cdot \hat{R}\right)$$

Where $\hat{\mu}_D$ and $\hat{\mu}_A$ represent the direction of the transition dipole moment of the donor and acceptor, and $\hat{R}$ is their mutual displacement unit vector pointing from D to A. While for two chromophores in solution the value of κ^2 averages to 2/3, in photosystems and antenna, where chlorophylls' orientations are fixed with respect to each other, this value can vary for each pair of pigments from 0 to 4 (Fig. 1.4B).

The spectral overlap can be described as the area-normalised integral of the fluorescence emission of the donor ($F_D(\lambda)$) and the absorbance spectrum of the acceptor ($\varepsilon_A(\lambda)$) (Fig. 1.4A).

$$J_F = \int_0^\infty F_D(\lambda)\ \varepsilon_A(\lambda)\,\lambda^4\,\mathrm{d}$$

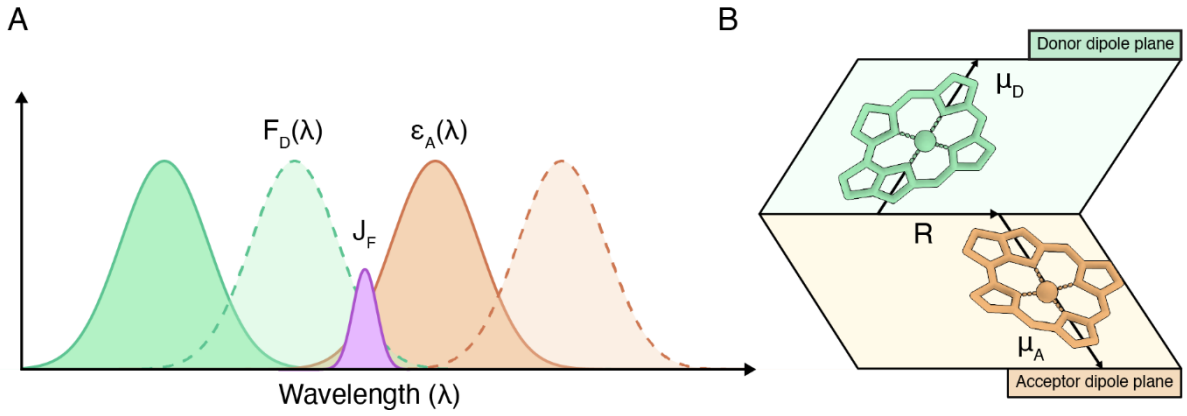


Fig. 1.4 – Spectral overlap and reciprocal orientation for resonance energy transfer

A) Visual representation of a donor molecule's absorption and emission spectra (in green) together with an acceptor molecule and its fluorescence emission spectra (in orange). The integral J_F of the spectra overlap is represented in purple. B) Visual representation of the reciprocal orientation of the dipole moments of two chlorophyll molecules.

From this formulation, we can understand important features of the photosynthetic antenna. Firstly, is that the reciprocal orientation of two pigments is crucial in determining the rates of excitation energy transfer between them. Secondly, is that excitation energy transfer happens with faster rates from chromophores that absorb at shorter wavelengths to chromophores that absorb at longer wavelengths.

Therefore in photosynthetic antenna, where a high number of pigments are present in close proximity, excitons will migrate through a series of probabilistic excitation energy transfer steps, from higher energy to lower energy pigments in a random walk. The exciton will finally be transferred to a pigment that presents rates of incoming excitation energy transfer which are much faster than the outgoing rates, which is usually referred to as a trap (37). In photosystems, the reaction centre chlorophylls act as traps, so that the rates of charge separation outcompete the rates of outgoing excitation energy transfer.

While the predominant direction of excitation energy transfer is conventionally understood as "downhill" (i.e. from higher to lower energy pigments) the phenomenon of uphill excitation energy transfer is also observed, where energy transfers from a lower-energy pigment to a higher-energy pigment (40–43). This process is governed by the Boltzmann distribution, which describes the probability (P_i) of a system being in a specific microstate (i) with energy (E_i) at a given absolute temperature (T):

$$P_i = \frac{e^{-E_i/k_B T}}{\sum_j e^{-E_j/k_B T}}$$

Although lower energy states are generally more populated, the Boltzmann distribution dictates a non-zero probability for higher energy states to be transiently occupied due to ambient thermal energy. This thermal energy, manifesting as continuous molecular motion, provides the necessary energetic headroom for an exciton to temporarily overcome relatively small energy barriers and access higher energy states.

In the context of photosynthetic complexes, excitation energy transfer rates adhere to this principle and to their Boltzmann-weighted populations. This means that while uphill transitions are less probable than downhill ones, they are not forbidden. Instead, the system continuously samples various energy states through thermal fluctuations, allowing excitons to momentarily move "uphill" before eventually relaxing towards reaction centres. This dynamic, probabilistic interplay, where thermal energy enables transient access to higher energy states, combined with specific structural arrangements that enhance entropic contributions, increases the high efficiency of photosynthetic light harvesting by ensuring effective energy funnelling to the reaction center, even from excitons on lower-energy pigments (44).

Depending on the biophysical architecture of the photosynthetic complex, reaction centre chlorophylls can present different levels of trapping, delineating two different models.

The first, which is closer to the behaviour of PSI, is referred to as a “deep” trap, in which, as the name suggests, the migration of excitation energy away from the reaction centre back to the antenna is much lower than the rates of charge separation. In this model, the

trapping of excitation is considered to be essentially irreversible, and the rate limiting step for a charge separation event to occur is the transfer to the trap (35).

The second scenario, closer to the behaviour of PSII, is referred to as a “shallow” trap. In this model, the rates of excitation energy transfer to and from the trap are more similar, allowing the excitation energy to escape and “visit” the reaction centre chlorophylls a number of times before a charge separation event finally takes place. In the “trap-limited” model, the excited-state lifetime is dominated by the rates of trapping and charge separation (45).

Finally, while thermal dynamics governed by the Boltzmann distribution explain a significant aspect of uphill excitation energy transfer, other quantum mechanical mechanisms also contribute to its feasibility and efficiency. For instance, the transfer might involve higher excitonic states of the energy-accepting molecule, effectively providing an energetically accessible pathway for the excitation to reach the desired higher energy level, even if the primary excited state of the acceptor is at a higher energy than the donor. Additionally, in certain scenarios, superexchange can play a role. This mechanism allows for a virtual, coherent transfer of excitation energy via a third, intermediate molecule, without the intermediate molecule itself becoming physically populated. This "through-bond" or "through-space" coupling can bridge an energetic gap, making uphill transfer possible by leveraging the electronic coupling within the pigment network, even when direct resonant transfer is unfavourable.

With the advancements in knowledge and increasing complexity of the theories that describe the quantum mechanical processes involved in excitation energy transfer, Förster theory of excitation energy transfer has been, at least partially, improved, in favour of models that can more thoroughly describe this phenomenon in pigment-rich protein complexes. Nonetheless, even though it cannot provide a complete description of energy transfer in photosynthetic antenna (46), Förster theory provides a good first approximation and understanding of the behaviour of chlorophylls in antenna and reaction centre pigments, suiting the scope of this thesis.

1.4 Antenna complexes in cyanobacteria and plants

To increase the amounts of photons captured and funnelled to the reaction centres of both PSII and PSI through excitation energy transfer, photosynthetic organisms have evolved an extremely diverse range of pigment-rich antenna protein complexes. The antenna complexes assemble, together with the photosystems, in structures called supercomplexes that enable the regulation of the flux of excitation energy to the reaction centres (37). Adaptation to different light environments has driven the evolution of a vast diversity of light-harvesting systems, together with several strategies to optimise the light energy input into photosynthetic membranes in response to varying light conditions (47).

In plants and green algae, the antenna complexes are constituted of transmembrane proteins, known as light-harvesting complexes (Lhcs), that bind Chl *a*, Chl *b*, and carotenoids, with members specialised in transferring excitation to PSII, called Lhcb, and to PSI cores, called Lhca (48). Lhc proteins assemble both in monomeric (i.e., CP29, CP26, and CP24 encoded by the *Lhcb4*, *Lhcb5,* and *Lhcb6* genes) and in trimeric forms (combinations of the *Lhcb1*, *Lhcb2,* and *Lhcb3* genes) (49).

In cyanobacteria, red algae and cryptophytes, light harvesting relies on phycobiliproteins (PBP), a family of water-soluble pigment-binding proteins with a close phylogenetic ancestry (50). PBP bind pigments from a family of open-chain tetrapyrroles called bilins, which are capable, thanks to minor chemical differences and tuning by the protein environment, of efficiently absorbing a broad spectrum of light that ranges from 550 nm to 710 nm. The pigments bound to the PBP are phycoerythrin (~560 nm), phycoerythrocyanin (~590 nm), phycocyanin (~620 nm), and allophycocyanin (~650–710 nm) (51).

PBP are arranged in structures called phycobilisomes (PBS) that are anchored to the stromal side of photosystems and can reach several MDa in size (52). The network of hundreds of bilin pigments in these complexes is arranged to funnel excitation energy from shorter wavelength pigments in the periphery to longer wavelength pigments at the core. Six structural classes of PBS have been described. They include structures described

as hemiellipsoidal (53), block-shaped (54), hemidiscoidal (55), bundle-shaped (56), paddle-shaped (57), and bicylindrical far-red light allophycocyanin cores (58).

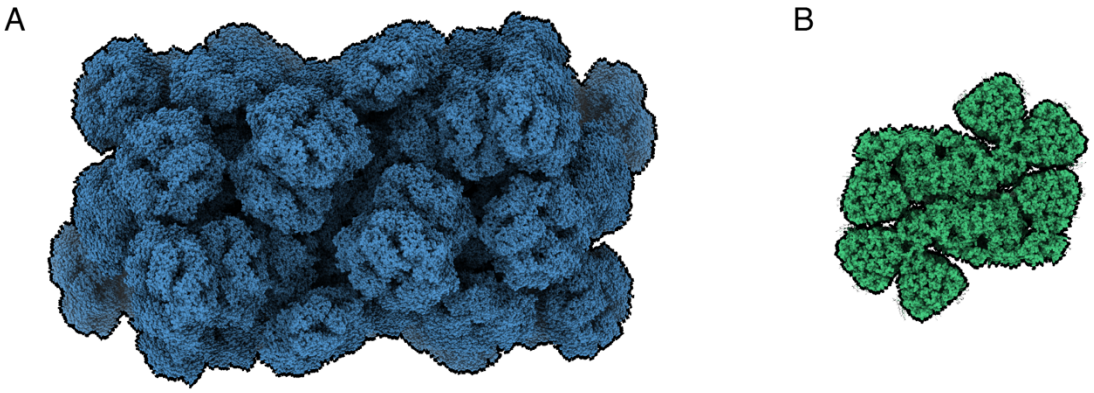


Fig. 1.5 – Plant and cyanobacterial antenna complexes

A) Hemidiscoidal phycobilisome complex from *G. pacifica* seen from the cytosolic side of the thylakoid membrane B) $C_2S_2M_2L_2$ PSII-LHCII supercomplex from *C. reinhardtii*

Spectroscopic and biochemical evidence indicates that PBSs can bind to both PSII and PSI (59,60), although direct binding of full-sized PBSs has been only demonstrated *in vivo* with PSII (61). The stable attachment of PBSs to PSII is mediated by specific linker proteins. The “PB-loop” of the linker membrane protein ApcE is considered critical for anchoring PBSs to the thylakoid membrane and facilitating stable association with PSII (62). Other linkers contribute to regulating a stable connection between the two complexes. For example Lpp1 and Lpp2, have been shown to interact with CP47 as well as the ApcB subunit of the PBS core (63), while ApcG has been demonstrated to modulate excitation energy transfer to PSII by altering its interaction with the protein (64).

Even though excitation energy from PBSs is primarily transferred to PSII, under certain conditions, energy can also be transferred to PSI, either directly or via "spillover" through PSII (65). Key models describing these transitions include the "mobile phycobilisome" model, where PBS complexes physically detach from PSII and move to PSI, and the "spillover" model, which proposes the transfer of excess energy from PSII to PSI. Phycobilisomes are known to be highly mobile on the thylakoid membrane, and their diffusion is essential for these adaptive state transitions, highlighting a crucial link between membrane fluidity, protein mobility, and photosynthetic regulation.

The differences in evolutionary choices of pigments between cyanobacteria and higher plants reflect the ecological niches occupied by these organisms as well as the balance between incident light and nutrients available (66). While Chl *a* is the most abundant photosynthetic pigment on earth, cyanobacteria and red algae accumulate vast amounts of bilins in their antenna, conferring higher activity in the green-yellow region of their action spectra compared to plants and green algae (50). In the water column, this can grant a major evolutionary advantage, since there is less competition for green light, allowing it to penetrate deeper. Compared to green algae and sea grasses, which tend to grow in shallow waters where the visible spectrum is similar to the spectrum of incident sunlight, red algae and cyanobacteria can afford to live at greater depths compared with organisms that rely solely on chlorophylls for light harvesting roles (66).

On the other hand, phycobiliproteins have a higher nitrogen content, and under conditions of limited nitrogen supply, their synthesis tends to be unfavourable. In cyanobacteria, rhodophytes, and cryptophytes, phycobiliproteins are selectively lost during nutrient stress conditions, since photosynthesis can still function in their absence (67,68). For these reasons, since most higher plants live in conditions of excess sunlight and therefore lack an evolutionary pressure to harvest different wavelengths, the use of bilin pigments for light harvesting is therefore energetically inefficient in land plants (66).

1.5 Thylakoid membrane organisation in cyanobacteria

The thylakoid membranes, not only essential for the light reactions of photosynthesis, but also for the respiratory electron transport chain, exhibit a complex and dynamic mesoscale organization which is crucial for regulation and maintenance of the photosynthetic apparatus (69,70). Unlike the organized grana stacks found in plant chloroplasts, cyanobacterial thylakoids are typically sheet-like and do not form extensive appressed regions. Instead, they appear as an ensemble of flattened cisternae positioned between the plasma membrane and the central cytoplasm, characterized by highly curved edges.

The unique architecture of cyanobacterial thylakoids is intimately linked to the presence of phycobilisomes (PBSs). These large, membrane-extrinsic light-harvesting antennae

decorate the cytoplasmic surface of the thylakoid membranes, physically spacing adjacent membranes apart and preventing the formation of tight stacks observed in plants (71).

While cyanobacterial thylakoids lack the lateral heterogeneity of their plant counterparts there is still considerable heterogeneity in their membrane landscape (72). In fact, more recently, specific regions of the thylakoid membrane have been observed having specialized functions. For instance, some areas are depleted in PBSs but enriched with ribosomes (73) or dominated by FtsH2 (74) (a protein crucial for PSII repair) have been observed, suggesting functions related to photosystem biogenesis or repair. Even the photosynthetic complexes themselves exhibit inhomogeneous distributions, forming distinct microdomains or even pseudo-crystalline arrays (75). These spatially segregated domains, sometimes hundreds of nanometres in scale, are likely to have significant functional implications for light harvesting and electron transport efficiency.

The distinct architecture of thylakoid membranes is a product of intricate interactions between lipids and protein complexes. Thylakoids possess a unique lipid composition, predominantly consisting of galactolipids (76). While the collective physical properties of the lipids composing the membrane support the formation of the characteristic flat bilayer sheets, monogalactosyldiacylglycerol, with its conical shape, has been suggested to contribute significantly to the membrane's structural plasticity, enabling the thylakoids to adopt curved and dynamic forms, potentially influencing regulatory roles (77). Beyond their scaffolding role, lipids are vital for stabilizing photosynthetic proteins and ensuring their proper function. Conversely, integral membrane proteins actively influence the lipid phase, directly contributing to the membrane's shape. For example, the oligomeric state of PSI has been shown to alter thylakoid curvature in *Anabaena* species, and mutants lacking PSI can form abnormal, tubulated membrane structures (78).

Much more still needs to be discovered on the mesoscale organisation of thylakoids, especially in non-model organisms. Nonetheless, dynamic interplay between specific lipids and proteins is crucial for the adaptability of the complex thylakoid network.

1.6 Plant biotechnologies and microalgal cultivation

In recent decades, photosynthetic microorganisms, such as cyanobacteria and microalgae, have sparked the attention of academia and industry (79). Using photosynthetic organisms to harness sunlight and convert carbon dioxide into valuable reduced carbon metabolites seems not only considered by many as a keystone to the sustainable food (80), feed (81), and fuel (82) production, but also a fruitful business opportunity to produce valuable secondary metabolites for pharmaceutical, nutraceutical, and cosmetic purposes (83). Microalgae and cyanobacterial species have other interesting characteristics compared to higher plants, not only they have higher growth rates and contain fewer degradation recalcitrant molecules (i.e., cellulose, hemicellulose and lignin) but are also less subject to seasonality. Furthermore, their cultivation can be coupled with wastewater treatment, as they are capable of thriving in phosphorous- and nitrogen-rich environments (84). Moreover, with respect to photosynthetic organisms used for first- and second-generation biofuel production, no fertile ground is necessary for their cultivation, thus creating less competition for arable land or precious, biodiverse landscapes.

However, even though large investments in terms of money and research have been made, microalgal domestication and cultivation is a field still in its infancy, and many biological and technological issues still need to be overcome to make these processes appealing. Producing biomass from microalgae and cyanobacteria remains an energy intensive process, far from having a net energy gain and meeting global demands (85,86) and as of today, there is no reported large-scale cultivation sustainable from both energetic and economic perspectives. The future of photosynthetic microorganisms for large scale production is still debated and the most optimistic forecasts are in fact derived by extrapolation of laboratory-scale data (87).

Of the biological limitations of photosynthetic biomass production, one of the most pressing needs is to adapt photosynthesis to environments which are extremely different from the ones it evolved in. In natural environments, organisms already have to cope with challenging light conditions. Light intensity and quality can vary dramatically over short periods, influencing the efficiency of the photosynthetic apparatus (47). In

photobioreactors, fluctuations in irradiance are exacerbated due to the high optical density of colonies and the continuous fast mixing regimes needed for the diffusion of nutrients and gasses. While in the outermost layers, high irradiance can cause photodamage, leading to a general loss in productivity, in the innermost layer, due to self-shading, irradiance is insufficient to saturate the photosynthetic apparatus, again lowering productivity (88). Compared to nature, the activation and inactivation of non-photochemical quenching mechanisms cannot activate and inactivate fast enough to regulate excitation fluxes to the reaction centres, exposing these organisms to photodamage. Even more detrimental for high productivity, the rapid mixing in photobioreactors leads to the movement of these organisms between the inner to the outer layers of the photobioreactors, leading to spikes in irradiance that can cause the formation of intense electric fields across the thylakoid membrane, which have been demonstrated to increase charge recombination rates in PSII, contributing to ROS production and photoinhibition (89). While better PBR design and adequate mixing regimes can help to establish higher growth rates, the overall efficiency of photobioreactors can still be greatly impaired by light fluctuations , that can lead to 4- to 5- fold reduction in growth rate (88).

For this reason, among other things, bioengineering and optimizing the light-driven reactions of photosynthesis to high density cell cultures has been suggested as one of the crucial steps in the journey towards sustainable photosynthetic biomass production. Smaller antenna sizes, as well as the introduction of faster non-photochemical quenching mechanisms can mitigate the effects of fluctuating light in photobioreactors.

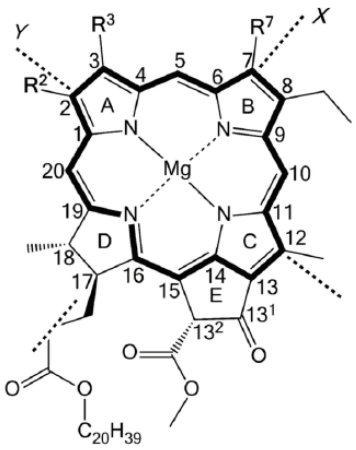
Despite their future potential, these approaches remain far from being a viable solution to present problems. “Improving photosynthesis” will require decades of fundamental research, technological breakthroughs, and substantial investments before yielding the expected results. Given these challenges, the narrative that sees photosynthetic biomass as an immediate solution to the climate crisis is misguided and risks diverting resources from more effective strategies. Efforts should instead focus on fundamental research that could provide the knowledge base for more targeted efforts and to prevent premature commercialisation of technologies that are not yet scalable or sustainable, ensuring that investments align with solutions that can deliver real impact in the near term.

1.7 Long-wavelength photosynthesis

A debated aspect regarding the improvement of bioprocess productivity is the possibility of expanding the absorption range of photosystems and antennae to limit the amount of self-shading in bioreactors and crop plants (90). Photosystems conduct extremely energy-insensitive redox chemistry, in which most of the energy absorbed from photons has to be used to handle photochemistry, trying to limit the damage incurred from side- and back-reactions and at the same time maintaining the yield of forward electron transfer as high as possible (14). Given the absorbance and action spectra of the two photosystems, 700 nm has historically been considered the "red limit" of photosynthesis (66). This prevailing dogma in the photosynthetic research community was fundamentally revised following the discovery of two types of red-shifted chlorophylls (91,92), which enable certain organisms to conduct photosynthesis with less energetic photons than previously thought possible (43).

To date, five different chlorophyll types have been identified: Chl *a*, *b*, *c*, *d,* and *f* (93). They share the same chlorin ring coordinating a magnesium, an isocyclic fifth ring and a hydrophobic phytol tail of variable length, fundamental for their insertion and stability in protein complexes (94). For this reason, they also share similar spectral properties, leading to approximately the same absorption coefficient around the Soret band (400–450 nm) and the Q_y band (~650–710 nm). Compared with Chl *a*, Chl *b* presents a formyl group instead of a methyl group at position C7 of the tetrapyrrole ring, and a blueshifted Q_y band.

On the other hand, the main difference between the longer wavelength chlorophylls (i.e., Chl *d* and Chl *f)* and Chl *a,* are the substituents to the tetrapyrrole rings. Substituents with an electron-withdrawing character along the Y axis of the Chl (from C2 to C12) cause a compression of the electron cloud toward the Y axis, ultimately leading to significant red-shifts in their absorption peaks (93,95). In particular, compared to the most abundant pigment Chl *a,* Chl *d* presents a formyl group at the C3 position, which, substituting a vinyl group, causes a red-shift in the Q_y maximum absorbance peak from 665 nm to 697 nm in methanol. Chl *f*, instead, presents a formyl substituent to a methyl group at the C2 position, also causing a red-shift to a new absorption peak of 706 nm (96,97).



	Ring substituents			Absorption peaks	
	R^2	R^3	R^7	Soret	Q_y
Chlorophyll a	CH_3	$CHCH_2$	CH_3	433	665
Chlorophyll b	CH_3	$CHCH_2$	CHO	468	652
Chlorophyll d	CH_3	CHO	CH_3	454	697
Chlorophyll f	COH	$CHCH_2$	CH_3	405	707

Fig. 1.6 – Substituents to the tetrapyrrole ring in different Chl types

Substituents in different chlorophyll types and their relative Q_x and Q_y absorptions.

From a bioenergetic perspective, the squeezed energy quota of the excited state of longer wavelength chlorophylls constitutes an energetic constraint for long-wavelength-adapted photosystems. To be able to drive the same chemistry and storing the same amount of energy in their products and electrochemical gradients, the lower "energetic headroom" must be compensated for by far-red photosystems, with a sacrifice either in photochemical efficiency or resilience to photodamage (43,89,98).

Given that light limitation is rarely a limiting growth factor during the growth of photosynthetic organisms in photobioreactors, it is worth investigating the energetic constraints that an extended absorption spectrum brings to the table. These insights are of value in designing strategies for the use of long-wavelength-adapted photosynthesis in agriculture and biotechnologies.

1.7.1 Chlorophyll d containing cyanobacteria

The first type of long-wavelength adaptation of a photosynthetic organism was discovered in 1996 in the cyanobacterium *Acaryochloris marina* (92). It was isolated in association with didemnid ascidians in tropical waters and is now known to cohabit with macrophytic red algae (99). To be able to colonize these specific ecological niches, that are perpetually shaded by other photosynthetic organisms, *A. marina* does not possess a white light version of its photosynthetic complexes but has evolved photosystems and antennae that

employ Chl *d* as the main photosynthetic pigment, enabling efficient trapping and charge separation with photons up to 740 nm (100). In particular, PSII is capable of charge separation up to ~700 nm and the nature of its primary donor is still debated (101). In PSI, Chl *d* also functions as the primary electron donor, while pheophytin *a* act as the primary acceptor (102,103), and can perform charge separation up to 740 nm (103,104). In the low-light, very stable ecological niche in which *A. marina* lives, the photosystems maintain similar absorption cross-section and forward electron transfer efficiencies to their white light counterparts. The price paid for the decreased available energy available lies in an increased sensitivity to photodamage, the pigment configuration in its photosystem II leads to a decreased energy gap between Phe and Q_a, increasing back reactions, ^{3}Chl formation and consequent photodamage (43,86,98).

1.7.2 Chlorophyll f containing cyanobacteria

The second type of long-wavelength photosynthetic adaptation was discovered by Min Chen by sampling water from Shark Bay in Australia in 2010 (91). It is now known to be present in at least 50 cyanobacterial species spread around different ecological niches (105) and, unlike *A. marina,* this is a facultative acclimation. When these cyanobacterial species are grown under increased ratios of far-red to white light, they are capable of transiently remodelling their photosynthetic apparatus, extending their absorption spectra in the far-red region in an acclimation process known as far-red light photoacclimation (FaRLiP) (106).

This extensive remodelling is controlled by a cluster composed of 21 genes, encoding the putative Chl *f* synthase gene (*PsbA4*) (107,108), alternative subunits of phycobilisomes and photosystems, and the three proteins required to sense and transduce changes in light quality. In particular, the ratio of white light to far-red light is sensed through a knotless phytochrome (regulator of far-red photosynthesis A, RfpA), a widespread family of photoreceptors sensitive to ratios between red and far-red light and is transduced by a DNA-binding response regulator (RfpB) and a cheY-like gene (RfpC) (109).

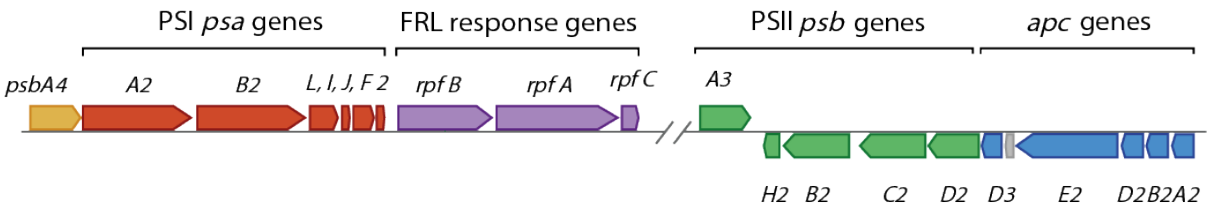


Fig. 1.7 – Structure of the FaRLiP cluster

Schematic representation of the FaRLiP gene cluster. In yellow is PsbA4, the alternative D1 subunit that allows the assembly of "super rogue" PSII. This subunit has been demonstrated to enable Chl *f* synthesis. In red, the alternative subunits of PSI, in purple, the regulatory elements of the cluster. In green the alternative subunits of PSII and in blue the alternative allophycocyanin subunits, in particular in grey a hypothetical D3 gene.

When these organisms are exposed to far-red light, their transcriptional profile is completely subverted, downregulating the expression of ~2000 genes, including those for the white light subunits of photosystems and phycobilisomes and promoting another ~900 genes and the dynamic activation of the rest of the FaRLiP gene cluster (110). The long-wavelength-adapted photosystems present some far-red specific subunits and share some subunits with their white light counterparts.

PSII contains 5 far-red specific subunits: PsbA3 (D1), PsbD2 (D2), PsbC2 (CP43), PsbB2 (CP47), and PsbH2, and incorporates four Chl *f* and one Chl *d* out of 35 chlorophylls. The reaction centre pigment Chl_{D1} has been assigned spectroscopically (43) and, more recently, structurally (111) as the single Chl *d,* and allows PSII to drive photochemistry with wavelengths up to ~727 nm (-110 meV less than its white light version). PSI contains 6 far-red light specific subunits: PsaA2, PsaB2, PsaL2, PsaI2, PsaJ2, PsaF2 and incorporates 7–8 Chl *f* (43,112). While the presence of a long-wavelength primary donor (i.e., Chl *d* in position Chl_{D1}) is accepted in Chl *f* PSII, it is still debated whether Chl *f* participates in the primary charge separation or is just used as an antenna pigment in Chl *f* PSI (43,112–118). Biophysical evidence indicates that Chl *f* containing PSI is capable of performing low-temperature far-red-light induced charge separation (43), strongly suggesting the presence of a chemical species capable of absorbing far-red light among the reaction centre chlorophylls. Nonetheless, when the structures of the trimeric Chl *f* containing PSI from two different organisms, *Fischerella thermalis* PCC 7521 (at a resolution of 3.19 Å) (112)

and *Halomicronema hongdechlroris* (at a resolution of and 2.41 Å) (116) were obtained, no evidence of a Chl *f* in the reaction centre was presented.

From a bioenergetic perspective, in contrast to Chl *d*-containing photosystems, Chl *f*-containing photosystems show no increased sensitivity to photodamage compared with their white light counterparts. In Chl *f* PSII, back-reaction routes appear similar to those in Chl *a* PSII, despite having the same amount of energy available to Chl *d* PSII. This is achieved by maintaining the midpoint potential of Phe/Phe^- and therefore the energy gap between Phe and Q_a resulting also in a smaller gap with the excited state of Chl_{D1}, favouring its repopulation and radiative routes of decay instead of those producing $^1O_2$ (98). The price to pay for Chl *f* containing photosystems is a nominal reduction in antenna size, since only a portion of the pigments employed are Chl *f* molecules.

This trade-off can be rationalised by taking into consideration the evolutionary niche in which this acclimation mechanism emerged. Indeed, far-red light photoacclimation is a temporary survival mechanism in unstable light conditions and shading coming from other cyanobacterial populations. Light fluctuations have therefore to be considered and managed accordingly, by sacrificing antenna cross section in favour of a better resistance to the photodamage caused by fluctuations in light.

Changes in the absorbance long-wavelength-adapted organisms vary from species to species, but these acclimation processes enhance light harvesting in wavelengths between 700 and 750 nm, accounting for a nominal ~19% increase in available photons. This increase has been suggested as one of the potential ways to allow an improvement to the theoretical maximum efficiency of photosynthesis (90). However, to be able to effectively design biotechnological applications of far-red photosynthesis, the photosynthetic complexes and the already known functional penalties that come with this improved harvesting capabilities (98) need to be completely characterised. Moreover, a better understanding of the acclimation process itself is needed to be eventually able to tune its triggers, rates, and overall effects on the photosynthetic apparatus for use in agricultural or biotechnological processes.

1.8 Cryo electron microscopy in photosynthesis and bioenergetics

Primary production on Earth is ultimately driven by the energy that photosystems harvest from sunlight. To unravel the underlying processes fully, the fields of bioenergetics and photosynthesis have always relied on spectroscopic and modelling techniques, and since the advent of structural biology, on high-resolution protein structures, key elements in elucidating the chemistry, mechanisms and energy fluxes of the molecular machineries of interest. Unfortunately, the challenging task of crystallizing large membrane complexes has hindered their widespread characterisation across multiple organisms. Given the time-consuming and resource-intensive nature of X-ray crystal structure determination, only a few research groups were able to pave the way in the determination of the structures of photosystems (5,119–122). This changed with the recent advances in the field of cryo electron microscopy (Cryo-EM) (123,124). It is now possible to quickly obtain medium- to high-resolution structures of very large complexes with less sample, time and optimisation than for crystallography. What was though impossible a decade ago is becoming a routine approach in the field of bioenergetics. Consequently, in the last five years, numerous structures of PSI and PSII and antenna complexes from cyanobacterial, algae and higher plants were determined (52,58,103,125–131). This has opened a Pandora's box of evolutionary variability in photosynthetic complexes, particularly in the organisation of the antenna around the highly conserved reaction centre cores (47).

A new era for structural biology has begun, in which Cryo-EM single particle analysis and Cryo-tomography will likely become dominant in the characterisation of large multi-subunit membrane complexes. Nonetheless, the contribution of crystallography, especially of serial femtosecond crystallography, will remain necessary to elucidate the transient, very short-lived states that lead to water splitting in PSII (19).

1.9 Overview on transmission electron microscopy

Transmission electron microscopy (TEM) is a type of microscopy that uses elastically scattered electrons to image thin sections of samples. In a column under extreme vacuum, an electron source (e.g., tungsten filaments, single crystal LaB_6) emits electrons that are accelerated by an electric field and focused through a series of electromagnetic lenses. The collimated radiation produced is then passed through the specimen and recorded by a detector.

The incident electron wave on the sample can be considered a plane wave, in which both the phase and the amplitude of the radiation are coherent at each given plane perpendicular to the direction of propagation. By passing through the specimen, the phase of each point in the plane gets modified by its interaction with the Coulomb potential of the atoms in its path. In the case of proteins and other "weak-phase objects", the exit wave will preserve its amplitude, and its wavefront will be spatially modulated by the electrostatic potential of the macromolecule. If the sample is thin enough, the exit wave can therefore be represented as the product of the incident plane wave with the transmittance function of the object:

$$\Psi_{exitwave}(x,y) = T(x,y)e^{-i\frac{2\pi}{\lambda}z}$$

where x and y are the coordinates of a given point in the plane of the exit wave, λ is the wavelength of the electrons and z is the direction of propagation of the electron. Other factors such as the spherical aberration and the defocus influence the measurement. Luckily, their combined effects can be represented mathematically by multiplying in Fourier space the exit wave function by a wave-aberration function ($e^{i\gamma(s)}$):

$$\gamma(s) = 2\pi\left[\frac{C_s}{4}\lambda^3 s^4 - \frac{\Delta Z}{2}\lambda s^2\right]$$

where s represents the spatial frequency, ΔZ the distance from true focus, and C_s represents the spherical aberration of the lenses. 2D image intensity can be represented in Fourier space through the following mathematical notation:

$$\mathcal{F}\{I\}(s_x, s_y) = \delta(s) - 2\mathcal{F}(s_x, s_y) sin\ \gamma(s) - \mathcal{F}\{\mu\}(s_x, s_y) cos\ \gamma(s)$$

Where $\delta(s)$ is the Dirac point spread function, $2\mathcal{F}(s_x, s_y)$ is the Fourier transform of the Coulomb potential and can be considered as if it was crystallographic structure factors that are modulated by the contrast transfer function, represented by $sin\gamma(s)$. The last term $\mathcal{F}\{\mu\}(s_x, s_y) cos\gamma(s)$ represents instead another set of structure factors that come from minor modulations in amplitude of the exit wave.

1.10 Cryo-EM single particle analysis

Cryo-EM Single Particle analysis (SPA) is an ensemble of data processing techniques that enables the determination of protein electrostatic potential maps from transmission electron microscopy images of a monodisperse solution of the protein of interest (124). Cryo-EM SPA is a continuously evolving field, and state-of-the-art algorithms keep improving the resolution achievable with this technique. Defining a standard workflow to obtain an electrostatic potential map starting from raw micrographs can be difficult, since the steps may vary according to the homogeneity and intactness of the protein sample, the image quality and various other factors. Nonetheless, it is still possible to outline the essential steps that have to be accomplished to obtain an electrostatic potential map.

The first step is to correctly compensate for beam-induced motion with motion correction techniques. When the electron beam traverses the sample, it can deform the ice, consequently altering the position of the particles in two main ways. First, at the beginning of the irradiation a “doming effect” is promoted that plastically expands the ice due to the release of stress accumulated during the freezing process. Second, the energy from the electrons breaks chemical bonds inside the macromolecules and increases the distance between atoms, expanding ice linearly with the dose. The movement generated hinders the image quality by blurring the particles during the motion. Thanks to the advent of direct electron detection cameras, the raw data is collected as a sequence of frames that allows for tracking and correcting the motion of portions of the micrographs or single particles through image reconciliation algorithms (132–134).

After this step, it is crucial to retrieve the contrast transfer function (CTF) and the envelope function of each micrograph (135). As stated before, the CTF mainly depends on the defocus of the micrograph (ΔZ) and formally determines at which spatial frequencies the images will contain signal on the image plane. The envelope function represents instead the dampening of the CTF at high spatial frequencies due to temporal and spatial aberrations in the microscope. An accurate estimation of these two functions allows the recovery of the maximum available information from the raw micrographs.

Once the raw micrographs have been pre-processed, the proteins, usually referred to as particles, are located in the micrographs and extracted with the appropriate box size. To do this, a range of possible approaches is available. Particles can be picked manually or by using rough shapes (blobs) to target differences in contrast in the micrographs. The state-of-the-art approaches involve the training of convolutional neural networks coupled with image denoising techniques, improving the clarity of the images and the quality of the particles extracted (136–138).

Once a particle stack has been extracted, the mathematical problem of Cryo-EM structure determination can be formalised as finding one or more 3D objects that best explain the 2D projections in the particles in the dataset. This ill-posed and incomplete mathematical problem (the relative orientation of the particles and the number of structurally distinct 3D volumes is not unknown) is tackled thanks to *ab initio* reconstruction algorithms in a Bayesian likelihood framework. In particular, one or multiple random initial densities are iteratively refined through a two-step procedure. In a first "expectation" step a probability distribution of the pose (relative position) of every particle is calculated. In the second step, the particles are combined in a 3D volume, which is then used to readjust the expected pose for each particle. The process can be iterated for a set number of times or until the 3D volume no longer changes.

On this very basic workflow, many different pre- and post-processing steps can be added, allowing for a more accurate accounting of the defocus and the motion of each individual particle, as well as for symmetries in the imaged protein.

1.11 Towards atomic resolution in CryoEM

Despite significant advancements in microscopes and image processing pipelines, which have facilitated the acquisition and processing of increasingly large datasets comprising hundreds of thousands of particles, simply accumulating more particles often yields diminishing returns in resolution improvement and more fundamental limitations persist. The amount of structural information extractable micrographs is ultimately limited by the extremely low signal-to-noise ratio (SNR) in raw images, a challenge that becomes more pronounced at higher spatial frequencies where noise can be orders of magnitude greater than the signal. For this reason, enhancing the SNR of the raw data is more critical for achieving higher resolution and more accurate particle orientations than larger datasets.

For many years, CryoEM routinely achieved resolutions in the 3-4 Å range, providing insights into protein architecture, and more recently structures at 2-3 Å resolution have become increasingly common. However, reaching “true" atomic resolution (139), typically defined as resolutions better than 1.5 Å, remains a complex challenge. Nonetheless, through a synergistic combination of advancements across several electron microscope hardware components, this goal was reached at the beginning of this decade by two concurrent studies that delivered apoferritin structure at 1.22 Å and 1.25 Å respectively, pushing the boundaries of the technique (140,141). The critical developments enabling these breakthroughs were primarily aimed at increasing the SNR at high spatial frequencies, addressing various hardware components of the microscope, from the electron emitter and objective lenses to the energy filter and detectors.

First, the spatial and temporal coherence and stability of the electron source proved to be a critical determinant for achieving atomic resolution. The transition from conventional Field-Emission Guns (XFEGs) to Cold Field-Emission Guns (CFEGs) was instrumental (141). XFEGs operate at elevated temperatures (~1800 K) and provide a stable beam with an energy spread of approximately 0.7-1.1 eV. In contrast, CFEGs are operated at room temperature, and generate a high-brightness electron beam with a remarkably smaller energy spread (<0.4 eV) (141). The enhanced spatial and temporal coherence provided by CFEGs directly translates to reduced chromatic aberration, particularly at higher

accelerating voltages, and consequently to improved image contrast and resolution, especially at high spatial frequences. For example, at 1.2 Å, the use of a CFEG can increase the amplitude of the CTF from 26% to 80%, leading to a substantial 9.5-fold increase in SNR (141). Modern CFEGs exhibit improved probe current stability and require less frequent "flashing" (i.e. warmup cycle to degas the materials) compared to earlier CFEGs, making them more practical for extended data acquisition sessions. The inclusion of a monochromator alongside a CFEG could further refines the electron beam by selectively narrowing its energy spread even more (to below 0.1 eV in some systems), thereby synergistically reducing chromatic aberration and enhancing overall image quality (140,142).

Another critical development was the advent of highly stable and narrow energy filters. These post-column filters are engineered to selectively remove inelastically scattered electrons, which contribute significantly to background noise and chromatic aberration (143). A key feature of these filters is their exceptional stability, enabling narrow zero-loss energy filtering, typically with slit widths <10 eV. The ability to consistently use narrow slit widths provides a significant boost in contrast and resolution.

Concurrent with energy filter advancements, a new generation of direct electron detectors revolutionised data acquisition (144). The Falcon-4 camera, for instance, represents a significant leap forward from its predecessors. While earlier detectors enabled electron counting at 40 Hz, requiring extended exposure times, the Falcon-4 operates at 248 Hz. This increased frame rate, coupled with the possibility of estimating electron events with sub-pixel accuracy, minimises event loss during counting (145). Consequently, high-quality images with typical exposure times of only a few seconds can be recorded while maintaining or even increasing the detective quantum efficiency. Furthermore, the possibility of recording movies with new formats, such as electron-event representation (EER) allows for highly flexible dose fractionation and advanced motion correction algorithms, better preserving high-resolution information that would be lost due to beam-induced motion (146).

Additionally, continuous refinement of objective lenses and aberration correction, specifically spherical aberration (Cs) correctors, typically used in material sciences, has been crucial (147). These correctors precisely compensate for optical imperfections in the electron lens system, allowing for a more coherent and focused electron beam to reach the sample, thus improving the intrinsic resolution capabilities of the microscope.

While these breakthroughs have highlighted the potential of hardware improvements in the quality of the determined structures, achieving routine atomic resolution remains a complex challenge. Many laboratories are constrained by the microscopes and associated infrastructure available to them, highlighting significant issues of accessibility and equitable infrastructure within the CryoEM community. The widespread impact of these advancements is thus limited by the availability of state of the art instruments and the specialised expertise required for their optimal operation. From a more approachable perspective, significant improvements are still required in sample preparation techniques and grid supports to ensure the formation of thin, radiation resistant, vitreous ice with optimally dispersed and oriented particles, while minimising ice contamination and preferential orientations. Further innovation in microscope hardware remains essential, but the interdependence of these components underscores that bottlenecks can shift, and continuous, synchronized advancements across the entire pipeline are necessary for CryoEM to truly reach routine atomic resolution.

1.12 The objectives of this work

The lack of high-resolution structural data on far-red-adapted photosystems limits our understanding of this new, low-energy paradigm of photosynthesis, also hindering potential biotechnological approaches aimed to introduce Chl *f*-based far-red photosynthesis in green algae and higher plants. The positions of the long-wavelength pigments in the antenna of both far-red light photosystems, as well as the nature of the primary donor of far-red photosystem I are still debated, and the intact structure of the far-red-light-adapted bicylindrical far-red light allophycocyanin is unknown. This thesis project has focused on obtaining a comprehensive structural view of the light-driven complexes and antenna proteins of the cyanobacterium *C. thermalis* PCC 7203 when adapted to far-red light conditions. In the brief introduction to each of the following chapters, a more detailed description of the outstanding questions in the literature and the techniques that were used to tackle them is provided. Here, a brief outline of the contents of each chapter is given.

In chapter 3, a new statistical approach to the analysis of the electrostatic potential of chlorophylls substituents is presented, aimed at distinguishing the different types of chlorophyll. The approach is based on methods present in literature but improves them providing a better statistical framework for the analysis of Cryo-EM maps and the assignment of different chlorophyll types in photosynthetic complexes. The method is used in the following chapters as one of the lines of evidence in the assignment of the Chl *f* positions in the two far-red light photosystems.

In chapter 4, two Cryo-EM SPA structures of far-red Photosystem I (at 1.88 Å and 2.01 Å resolution) are presented, and the 8 Chl *f* sites are assigned, among which the reaction centre chlorophyll A_{-1B}. The sites are assigned using multiple lines of evidence, i.e., the method presented in chapter 3 and phylogenetic analysis of the conserved differences between the white light and far-red light subunits of photosystem I and their consequent structural changes. The biophysical implication of a Chl *f* primary donor are discussed, as well the connectivity between the monomers in the trimeric assembly of the complex. The

photosystem I map at a resolution of 1.88 Å is, at the time of writing, the best resolution ever obtained for any Photosystem I structure.

In chapter 5, Cryo-EM SPA structures of an intact monomeric far-red photosystem II (2.35 Å), a non-intact monomeric far-red photosystem II lacking multiple subunits and the oxygen evolving complex (2.68 Å) and an intact dimeric photosystem II (2.17 Å) are presented. In the two intact maps a novel subunit, with sequence present in the FaRLiP cluster is present. The subunit has been named PsbH2' due to its interaction with the core subunit PsbH2 and its function is discussed. In these maps the 4 antenna Chl *f* sites are assigned based on the statistical analysis of the chlorophylls electrostatic potential and the conserved phylogenetic changes between the white light and far-red light versions of these two proteins. Moreover, the presence of Chl *d* as the reaction centre chlorophyll Chl_{D1} is confirmed together with the preferential orientation of the C3 formyl substituent characteristic of Chl *d*. The excitation energy transfer network of far-red photosystem II is discussed based on the assignment, highlighting the importance of considering photosystem II as a dimeric protein complex.

In chapter 6, an intact map of the far-red light adapter bicylindrical allophycocyanin (2.61 Å) is presented. The phylogenetically conserved changes that reflect the geometry of the bilin pigments in far-red light phycobiliproteins are discussed, together with the evolutionary significance of the smaller antenna complex and its broader absorption range. Moreover, based on structures present in literature, the organisation of photosystems in the thylakoid membranes after far-red light illumination is discussed. Experimental approaches are suggested to determine the structure of any PSII-APC complex. The terminal emitter of the complex is assigned to ApcE2 together with two potential excitation injection points in far-red Photosystem II.

In chapter 7, RfpA, the phytochrome that controls far-red light photoacclimation, has been heterologously expressed in *E. coli*, and its photoswitching behaviour characterised. A mathematical model of its population dynamics is presented.

This work presents significant advances in our understanding of the evolutionary costs and benefits of Chl *f*-based long-wavelength photosynthesis.

2 Materials and methods

2.1 Far-red light culture growth of *C. thermalis* PCC 7203

Chroococcidiopsis thermalis PCC 7203 was grown in liquid BG11 medium (148) at 30 °C under 750 nm LED illumination at an intensity of ~30 μmol photons m^{-2} s^{-1} (750 nm, Epitex; L750-01AU). Cultures were harvested, pelleted and flash-frozen after at least 3 months of exposure to exclusively far-red illumination. Three months of far-red light photoacclimation are necessary to allow the white light components of the photosynthetic apparatus to be completely depleted and have been chosen based on previous literature (43). The absorbance and fluorescence spectra of adapted cells were measured as well as the ratio of Chl *f*/Chl *a* to ensure consistency with data available in the literature.

2.2 Absorbance measurements

Absorption spectra of isolated complexes were recorded at room temperature using a Cary 60 spectrophotometer (Agilent Technologies).

2.3 Isolation of FR-PSI with sucrose gradient

Cells were broken with two passages in a flow cell disruptor (Constant Systems) at a pressure of 39 kPsi and then centrifuged at low speed (1000 g, 5 min at 4 °C) to remove any unbroken material. Isolation was carried out in darkness or minimal dim green light. The supernatant was then centrifuged at 40,000 rpm in a Ti45 rotor (Beckman Coulter) at 4ºC for 20 minutes to pellet the membranes. Membranes were resuspended to 1 mg mL^{-1} total chlorophyll in TM buffer (50 mM MES-NaOH pH 6.5, 5 mM $CaCl_2$, 10 mM $MgCl_2$, 1.2 M betaine) and solubilised at 4 ºC for 1 hour by the addition of β-DM to a final concentration of 0.4% (w/v) in the dark. After the removal of insoluble debris by centrifugation, the supernatant was then concentrated (100 kDa Filters, Amicon) and the solubilised membranes were loaded onto sucrose gradients (50 mM MES-NaOH pH 6.5, 5 mM $CaCl_2$, 10 mM $MgCl_2$, 500 mM Sucrose, 0.04% β-DM) and centrifuged overnight at 120,000 x g at 4º C in a SW28 rotor with maximum acceleration and deceleration. Bands

containing Photosystem I (PSI) trimers were harvested from the sucrose gradients and then diluted and reconcentrated multiple times with TM buffer, to remove most of the sucrose before being loaded on a MonoQ 5/50 anion exchange column (GE Healthcare). The column was washed with 5 column volumes of Buffer A (20 mM MES-NaOH pH 6.5, 5 mM $CaCl_2$, 5 mM $MgCl_2$, 0.03% β-DM). Fractions were eluted with a 0% to 100% gradient with Buffer B (Buffer A + 600 mM NaCl) at a flow rate of 0.5 ml/min over 50 minutes. Eluted peak fractions were pooled, washed to remove NaCl, concentrated (100 kDa Filters, Amicon) and stored at 4°C or flash frozen and stored at -80°C depending on the use.

2.4 Isolation of FR-PSII and FR-APC complexes

Cells were broken with two passages in a flow cell disruptor (Constant Systems) at a pressure of 39 kPsi and then centrifuged to remove any unbroken material at 1000x g. The supernatant was then centrifuged at 40,000 rpm in a Ti45 rotor at 4º C for 20 minutes to pellet the membranes. Membranes were resuspended in low salt buffer A (20 mM MES, pH 6.5, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 1.2 M betaine, 0.03% β-DM), adjusted to a chlorophyll concentration of 0.4 mg/ml and solubilised at 4º C for 1 hour by the addition of β-DM to a final concentration of 0.4% (w/v). Solubilised membranes were centrifuged at 42,000 rpm at 4 °C in Ti45 rotor for 30 minutes to remove unsolubilised material. The supernatant was then loaded onto a DEAE Toyopearl 650S (Tosoh Bioscience, ID: 50 mm Length: 450 mm, 500 mL CV), washed with 1 CV of buffer A and eluted with a 0%-100% linear gradient of buffer B (20 mM MES, pH 6.5, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 0.5 M NaCl, 1.2 M betaine, 0.03% β-DM) over 5 CV. Chlorophyll and bilin containing fractions were pooled, concentrated (100 kDa Filters, Amicon) and stored at 4 °C or -80 °C depending on the use.

2.5 Grid preparation

2.5.1 Collection A

C-Flat 1 μm hole size and 1 μm hole spacing with a 300-copper mesh grids (Agar Scientific) were glow-discharged for 30 s at 25 mA and 3.5 μl of FR-PSI sample (1 mg/ml of chlorophyll, equivalent to ~4.2 mg/ml of protein) were applied. Grids were blotted at 4°C

and 100% humidity in the presence of only dim green light and plunge frozen in liquid ethane after 3 seconds with a Vitrobot Mark IV (Thermo Fisher Scientific).

2.5.2 Collection B

Gold quantifoil 2 μm hole size and 1 μm hole spacing with a 300-copper mesh grids were glow-discharged for 30 s at 25 mA and 3.5 ml of PSII sample (~1 mg/ml of Chl concentration) were applied. At a temperature of 4 °C and 100% humidity and in the presence of only a dim green light, grids were blotted and plunge-frozen in liquid ethane after 3 seconds with a Vitrobot Mark IV (Thermo Fisher Scientific).

2.5.3 Collection C

Gold quantifoil 2 μm hole size and 1 μm hole spacing with a 300-copper mesh grids were glow-discharged for 30 s at 25 mA and 3.5 ml of APC+PSII sample (~1 mg/ml of Chl) were applied. At a temperature of 4 °C and 100% humidity and in the presence of only a dim green light, grids were blotted and plunge-frozen in liquid ethane after 6 seconds with a Vitrobot Mark IV (Thermo Fisher Scientific).

2.5.4 Collection D

Gold quantifoil 2 μm hole size and 1 μm hole spacing with a 300-copper mesh grids were glow-discharged for 30 s at 25 mA and 3.5 ml of APC+PSII sample (~4.2 mg/ml of protein) were applied. At a temperature of 4 °C and 100% humidity and in the presence of only a dim green light, grids were blotted and plunge-frozen in liquid ethane after 3 seconds with a Vitrobot Mark IV (Thermo Fisher Scientific).

2.6 Data acquisition of FR-PSI

2.6.1 Collection A

Initial screening of the sample was conducted on a Glacios (Thermo Fisher Scientific) operated at 200 kV and a magnification of 75000x. Particles were imaged using a Krios III (Thermo Fisher Scientific) operated at 300 kV and a magnification of 130000x. Images

were recorded on a Falcon 4i (Thermo Fisher Scientific) with a pixel size of 0.921 Å and a dose of 40 electrons per Å^2 for a total of 40 frames. Images were collected in super resolution mode with a SelectrisX energy filter with a slit width of 20 eV. The targeted defocus range was varied from –0.5 to –3.5 μm using the EPU software (Thermo Fisher).

2.6.2 Collection B

Initial screening of the sample was conducted on a Glacios (Thermo Fisher Scientific) operated at 200 kV and a magnification of 75000x. Particles were imaged using a Krios III (Thermo Fisher Scientific) operated at 300 kV and a magnification of 155000x. Images were recorded on a Falcon 4i (Thermo Fisher Scientific) with a pixel size of 0.723 Å and a dose of 40 electrons per Å^2 for a total of 40 frames. Images were collected in super resolution mode with a SelectrisX energy filter with a slit width of 20 eV. The targeted defocus range was varied from –0.8 to –2 μm using the EPU software (Thermo Fisher).

2.6.3 Collection C

The particles were imaged using a Krios 4 operated at 300 kV. Images were recorded on a Falcon 4i with a pixel size of 0.651 Å and an exposure of 50 electrons per Å^2 for a total of 40 frames, images were collected in super resolution mode with a SelectrisX energy filter with a slit width of 20 eV. The targeted defocus range was varied from –0.8 to –2 μm using the EPU software (Thermo Fisher).

2.6.4 Collection D

The particles were imaged using a Krios 3 operated at 300 kV. Images were recorded on a Falcon 4i with a pixel size of 0.723 Å and an exposure of 40 electrons per Å^2 for a total of 40 frames, images were collected in super resolution mode with a SelectrisX energy filter with a slit width of 20 eV. The targeted defocus range was varied from –0.6 to –2.4 μm using the EPU software (Thermo Fisher).

2.7 Single particle analysis

2.7.1 Collection A

A total of 8854 movies were collected from two different grids. The frames were aligned, dose weighted, and the contrast transfer function (CTF) was estimated in CryoSPARC v4.4.1 (149). Micrographs were curated by removing those with CTF fits worse than 10 Å. The subset obtained contained 7455 micrographs (84.1%) and was used to hand pick ~2000 PSI-type particles across the defocus range. Particles were 2D classified and used to template pick across the entire dataset, yielding 1,560,287 picks. Duplicated particles were removed and after multiple rounds of 2D classification and *ab initio* refinement two separate classes were identified, one corresponding to FR-PSI trimers (138,048 particles) and one corresponding to FR-PSII lacking extrinsic subunits.

FR-PSI trimers

The particle stack was polished by heterogeneous refinements and multiple rounds of per particle CTF refinement and model based motion correction, yielding a stack of 83,225 particles were used to perform non-uniform refinement imposing C3 symmetry, producing a map at a global resolution of 2.01Å based on GS-Fourier Shell Correlation at a cut-off of 0.143. The Guinier's analysis reports an estimated B-factor of 39.7.

FR-PSII monomers

The particle stack was polished by heterogeneous refinements and multiple rounds of per particle CTF refinement and model based motion correction, yielding a stack of 226,959 particles were used to perform non-uniform refinement, producing a map at a global resolution of 2.58 Å based on GS-Fourier Shell Correlation at a cut-off of 0.143. The Guinier's analysis reports an estimated B-factor of 68.1. Detailed processing workflow shown in fig. S26

2.7.2 Collection B

A total of 23311 movies were collected from a single grid. The frames were aligned, dose weighted, and the contrast transfer function (CTF) was estimated in CryoSPARC v4.4.1 (149). Micrographs were curated by removing those with CTF fits worse than 10 Å. The subset obtained 93% of the initial micrographs and was used to hand pick ~1000 PSII-type particles across the defocus range. Particles were 2D classified and used to template pick across the entire dataset, yielding 1,087,739 picks. After multiple rounds of 2D classification and *ab initio* refinement, two major classes of particles emerged, one corresponding to FR-PSII dimers and another one corresponding to FR-PSI trimers.

FR-PSII dimers:

Duplicated particles were removed from the stack and of ~20,000 particles was used for homogeneous refinement to obtain an initial map at a resolution of ~3 Å and to confirm C2 symmetry. After multiple rounds of per particle CTF refinement and local motion correction, a set complete set of 49,532 particles was used to perform non-uniform refinement imposing C2 symmetry, producing a map at a global resolution of 2.17 Å based on GS-Fourier Shell Correlation at a cut-off of 0.143. The Guinier's analysis reports an estimated B-factor of 39.7.

FR-PSI trimers:

Duplicated particles were removed from the stack and of ~30,000 particles was used for homogeneous refinement to obtain an initial map at a resolution of ~3 Å and to confirm C3 symmetry. After multiple rounds of per particle CTF refinement and local motion correction, a set complete set of 297,795 particles was used to perform non-uniform refinement imposing C3 symmetry, producing a map at a global resolution of 1.88 Å based on GS-Fourier Shell Correlation at a cut-off of 0.143. The Guinier's analysis reports an estimated B-factor of 46.6.

2.7.3 Collection C

A total of 9057 movies were collected, frames were aligned, dose weighted, and the contrast transfer function (CTF) was estimated in CryoSPARC v4.0.2 (149). Micrographs were curated by removing the ones with CTF fits below 10 Å. The subset obtained contained 93% of the total. Around ~2000 particles were hand-picked in a subset of micrographs across the defocus range, used to train Topaz (137) and pick particles across the entire dataset, yielding 534,993 picks. After multiple rounds of 2D classification and Ab-initio refinement, duplicated particles were removed and a subset of 179,581 particles was used perform non-Uniform refinement (150) imposing C1 symmetry, producing a map at a global resolution of 2.61 Å based on GC-Fourier Shell Correlation at a cut-off of 0.143.

2.7.4 Collection D

A total of 9397 movies were collected, frames were aligned, dose weighted, and the contrast transfer function (CTF) was estimated in CryoSPARC v4.0.2 (149). Micrographs were curated by removing the ones with CTF fits below 10 Å. The subset obtained contained 9224 micrographs (98%) and blob picker was used to pick particles in a subset of micrographs across the defocus range. Particles were 2D classified and used to train Topaz (137) and pick particles across the entire dataset, yielding 423,203 picks. After multiple rounds of 2D classification and Ab-initio refinement, duplicated particles were removed and a subset of 36,300 particles was used for homogeneous refinement to obtain an initial map >5Å and to confirm C2 symmetry. After multiple rounds of per particle CTF refinement and local motion correction (151), a set of 26,300 particles was used to perform non-Uniform refinement (150) imposing C2 symmetry, producing a map at a global resolution of 2.61 Å based on GC-Fourier Shell Correlation at a cut-off of 0.143.

2.8 Model building

2.8.1 FR-PSI

The FR PSI model 7LX0 (Gisriel et al., 2020) was fitted int the map, mutated to the correct sequence with Chainsaw (Phenix) and refined in Coot (152). Each subunit density was evaluated amino acid by amino acid and the model was then refined in real space in Phenix.

2.8.2 Holo-FR-PSII

The FR-PSII dimeric model 8EQM (58) was fitted to the ESP map using the Phenix software suite (153), mutated to the correct sequence with Chainsaw (Phenix) and refined in Coot (152). Each subunits density was evaluated amino acid by amino acid and the model was then refined in real space in Phenix.

2.8.3 Apo-FR-PSII

The Holo-FR-PSII model was fitted to the ESP map using the Phenix software suite (153), and the subunits that did not present density were deleted. The resulting model was evaluated amino acid by amino acid and then refined in real space in Phenix.

2.8.4 FR-APC

The far-red APC incomplete rod from 8UHE (58) was fitted to the ESP map using the Phenix software suite (153), and then the third ring was copied and fitted again with Phenix into the density of the fourth ring. The complete model was then mutated to the correct sequence with Chainsaw (Phenix) and refined in Coot (152). Each subunits density was evaluated amino acid by amino acid and the model was then refined in real space in Phenix.

2.9 De novo sequence identification of PsbH2'

The Holo-FR-PSII map presents density for a previously unreported subunits of PSII adjacent to PsbH2. Two separate strategies were employed, the subunit was built with Modelangelo, and at the same time fitted with a polyalanine helix of the correct length. amino acids retrieved with Modelangelo that presented a confident density in the map were mutated on the polyalanine helix. The resulting sequence was then blasted against the genome of *C. thermalis* PCC 7203, revealing the presence of a gene encoding for the protein in the FaRLiP cluster. The sequence was adjusted accordingly and the model refined in Phenix.

2.10 77 K fluorescence spectroscopy of APC complexes

1 ml of isolated PSII + APC complexes at an OD_{710} of 0.1 was frozen in liquid nitrogen and moved to a finger dewar also filled with liquid nitrogen in a FluoroMax 4 spectrofluorometer (HORIBA Ltd., Japan). Fluorescence emission was measured from 650 nm to 850 nm with excitation and emission slits of 2 nm and an integration time of 0.5 s over an excitation range from 400 nm to 600 nm.

2.11 Statistical analysis of Chl substituents ESP

The electrostatic potential (ESP) is interpolated on the surface of a cone with its axis extending from that of the CX – CX' bond, where X is the IUPAC number given to the chlorin carbons that have relevant substitutions, and an aperture 120° (2θ). The ESP is sampled within the range of distances 0Å to 2.5Å (d), with a step size of 0.1Å. This is repeated for every 5° rotation of the torsion angle (φ). The 90° torsion angle is defined as being parallel to the plane defined by the positions of the three carbon atoms, CX-1, CX and CX', in the direction of CX-1, where CX is the IUPAC carbon number on the chlorin ring. The direction of rotation of the torsion angle is clockwise by IUPAC convention.

Our implementation is written in Python using the GEMMI library (154). The Jupyter notebook computes the average (μ) and standard deviation (σ) of the ESP on the surface

of the cone for each of the substituents (C2, C3, C7, C8, C12). It thus provides visual information on the quality of the map and on the preferential orientation of the vinyl (C3) and ethyl (C8) substituents with respect to the chlorophyll plane. It also calculates the precision, recall and accuracy as metrics of the quality of the analysis. To classify the substituents in the C2 position, the Z-score of each sampling point is computed with respect of the average methyl substituent in position C7.

$$Z_{[d,\varphi]} = \frac{ESP^{C2}_{[d,\varphi]} - \mu^{C7}_{[d,\varphi]}}{\sigma^{C7}_{[d,\varphi]}}$$

The ESP and the Z-scores on the surface of the cones are plotted as 2D projection looking down the CX – CX' axis (fig. S1). This approach allows for the possibility that maximum differences between substituents could occur outside of the predicted length of the formyl's carbon – oxygen double bond. It thus explores a larger area compared with the original cone-scan method which sampled a single distance and gave a single line around the cone (155). Moreover, using Z-scores instead of the strict $\mu + 3\sigma$ cutoff as used previously (155,156). allows a more quantitative investigation of both positive and negative features in the environment of the substituent.

The calculations presented are obtained using a PDB in which every chlorophyll molecule was substituted with a Chl *a* and then refined as described in the methods section to prevent rotations in the Chl ring that could influence the results of the scan, using the unsharpened map. The code is available on GitHub:

(https://github.com/giovanniconsoli/Chlorophyll-subsituents-scan).

2.12 *E. coli* transformation

50 µl of *E. coli* cells (KRX strain) were incubated with the pKT271 bilin biosynthesis plasmid and the RfpA backbone containing plasmid simultaneously in ice for 20 minutes. The cells were then heat-shocked at 42 °C for 45 seconds and then kept for 5 minutes at 4 °C. The culture was then recovered in 950 µl of overgrowth medium for 30 minutes and then plated on chloramphenicol (25 µg/ml) and ampicillin (100 µg/ml) plates.

Transformants were grown in 10 ml of antibiotics supplemented LB media and aliquots stored in 10% glycerol at -80 °C.

2.13 RfpA heterologous expression and purification

The transformed strain was grown at 37 °C with shaking at 250 rpm to an $OD_{600\,nm}$ of 0.6 in 1 L of TB media containing 100 mg/L ampicillin, and 25 mg/L chloramphenicol. The culture was cooled to 18 °C and 1 mM IPTG added for overnight expression. Cells were collected by centrifugation and resuspended in 50 ml of 300mM NaCl 30mM MES pH 7.0 and broken using a cell disruptor at 4 °C. unbroken material was pelleted by centrifugation using an Optima L-100 XP ultracentrifuge with a Ti45 rotor at 121,000-xg for 1 hour. The protein soluble fraction was incubated with washed Ni-NTA beads (Qiagen) overnight at 4°C on a rotary wheel. For purification, the incubated mixture was centrifuged at 1000 g for 5 minutes and the supernatant discarded. Ni-NTA beads were washed with 15-ml 300mM NaCl 30mM MES 30mM imidazole pH 7.0 three times and the supernatants were collected. For elution, Ni-NTA beads were washed with 15 ml 300 mM NaCl 30 mM MES 150mM imidazole pH 7.0 three times. SDS-PAGE analysis was conducted, and elution fractions were pooled and concentrated using a 30 kDa VivaSpin column (Sartorius).

2.14 SDS polyacrylamide gel electrophoresis

Protein samples were prepared by mixing with 2X sample buffer (62.5 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 5% β-mercaptoethanol, and 0.002% bromophenol blue) and denatured by incubating them at 95 °C for 5 minutes. Novex 4-20%, Tris-glycine Plus WedgeWell polyacrylamide gels were loaded with 20 µl in a Mini-PROTEAN® Tetra Cell system (Bio-Rad) at 120 V for approximately 60 minutes in running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). Protein standards (Precision Plus Protein™ Dual Color Standards, Bio-Rad) were used as molecular weight markers. The gels were stained with Coomassie Brilliant Blue R-250 (0.1% w/v in 50% methanol and 10% acetic acid) for 1 hour, and excess dye was removed with deionized water.

3 Methods to investigate chlorophyll densities in Cryo-EM

3.1 Introduction

In the last decade Cryo-EM has become a prominent method in structural biology to obtain data of protein complexes in the fields of bioenergetics and photosynthesis (157). Compared to crystallography, the reduced costs and complexity of Cryo-EM allows to obtain higher resolution structures faster and to explore in greater depth the evolutionary variability of the photosynthetic machinery. The structures of both PSI and PSII are now available for a range of organisms and have allowed detailed structural analysis of different chlorophyll types and sites (103,111,112,128–131,158). In complexes containing more than one type of chlorophyll, such as the plant Chl *a* and Chl *b* containing antenna in plants and Chl *f* containing photosystems in cyanobacteria, the ability to match different ligands to chlorophyll sites has been a priority, since it has deep functional implications.

Chlorophyll types differ from one another only in the substituents to their tetrapyrrole ring (93,159). It is thought that photosynthetic organisms have evolved the synthesis of different chlorophyll types by functionalisation of the readily available Chl *a*, allowing them to alter the chemical properties of the pigment, and consequently tuning its absorption spectra by modifying their Q_x and Q_y transitions. Chl *b*, is synthesised by the enzyme CAO (Chlorophyll *a* oxygenase) and features a formyl group in position C7 instead of a methyl group (160). Chl *d* features a formyl group in position C3 instead of the vinyl group of Chl *a*, and its synthetic pathway is yet to be elucidated (101,161). Chl *f* is synthesised by a light-driven complex containing a paralogue subunit of PSII (psbA4) (107), and also differs from Chl *a* by a single atom, a formyl group in position C2 instead of a methyl group (Fig. 1.6).

These subtle structural variations require a high level of local resolution to be unambiguously detected from electrostatic potential (ESP) maps (155) and, due to the

electronegative nature of these substituents, multiple factors complicate the unambiguous detection of the formyl's oxygen atoms, even in high-resolution ESP maps:

1) Differently from X-ray crystallography, which provides information on electron density directly, Cryo-EM maps are generated from 2D projections of the Coulomb potential (ψ) of the imaged proteins according to:

$$\psi(r) = q/4\pi r\varepsilon$$

where q is the charge in Coulombs, ε is the permittivity of the medium in which the charge is situated, and r is distance from the charge. Since $\psi(r)$ linearly depends on the charge (q), the protons in each nucleus contribute positively, while the electron cloud around them contributes negatively, shielding their signal. Therefore, the more negatively polarised an atom, as in the case of the formyl's oxygens of Chl *b*, Chl *d* and Chl *f*, the lower its contribution to the overall signal of the map (155,162,163). This characteristic of ESP maps is reflected in negatively polarised side chains and in previous Cryo-EM models of Chl *f* photosystem I (112,116,129,156), where the formyl's oxygens is often undetectable, disallowing the confident assignment of some of the Chl *f* sites present.

2) The electron beam can trigger decarboxylation events during the irradiation, leading to a loss of signal corresponding to the formyl's oxygen in later frames of the movies. Signal loss related to beam damage is partially compensated during the motion correction steps, in which dose weighting techniques are employed to retrieve high-resolution information prevalently in earlier frames, while relying on later frames only for lower resolution information (164,165). Nonetheless, the sensitivity of negatively polarised side chains and formyl's oxygens to radiation damage hinders the effectiveness of dose weighting techniques (134).

3) For formyl groups that are not hydrogen bonded, the formyl's oxygen can adopt a wider distribution of torsion angles where its energy will be minimised, therefore diluting the signal over the two different positions in plane with the tetrapyrrole ring of the chlorophyll.

These issues became dominant when the first Cryo-EM structures of Chl *f* containing PSI and PSII were published (111,112,116,155). In these complexes, knowing the position of Chl *f* sites is crucial to relate structure to function. The debated presence of long-wavelength pigments among the reaction centre chlorophylls (43,112,156), for example, can alter the interpretation of spectroscopic and biochemical data, and thus our understanding of how this low energy paradigm of photosynthesis works (118,166). Moreover, the correct assignment of Chl *f* to specific sites is needed for modelling approaches that rely on calculating excitation energy transfer rates between pigments.

Due to the conserved structural changes in long-wavelength-adapted photosystems and the fact that Chl *f* absorbs at longer wavelengths compared to Chl *a*, Chl *f* sites are expected to be extremely conserved in each photosystem to avoid slow and inefficient excitation energy transfer to the reaction centre. At room temperature (293 K) the Boltzmann constant allows only for an uphill energy transfer of around ~25 meV, therefore Chl *f* must be placed in strategic positions in the architecture of a photosystem to be able to harvest long-wavelength photons and avoid low energy trap in transferring these photons to the reaction centre.

Therefore, being able to confidently assign the Chl *f* pigments to specific sites has fundamental importance, but this assignment cannot be informed solely by the sole direct observation of the ESP map. Other aspects, such as the chemical environment, energetic and structural constraints as well as completely separate lines of evidence need to be taken into consideration when claiming the presence or absence of a Chl *f* at a specific site. Chl *f* assignments in the literature have so far been guided by consideration of the specific conserved changes in the far-red paralogues of photosystem subunits relative to their homologues in white light-absorbing Chl *a* photosystems and by the chemistry of Chl *f* (43,93).

In particular, two features that can help with the identification of potential Chl *f* sites are the presence of i) a hydrogen bond partner for the formyl group to be stabilised, and ii) a non-histidine axial coordination of the central Mg^{2+}. Changes of substituents on a tetrapyrrole chlorophyll ring, influence not only the absorption properties of the molecule,

but also the coordination chemistry of the central Mg^{2+} atom. While Mg^{2+} in Chl *a* is preferentially coordinated by an histidine, the formyl group in Chl *f* exerts a stronger pull of electrons on the Mg^{2+}, making it behave more like a Lewis acid, by favouring electrostatic bonds with side chains or molecules with a fixed strong dipole, such as water molecules (93). Taking into consideration all the previous assignments in literature, the conserved far-red light changes around the sites, hydrogen-bond partners for the C2 formyl and the central Mg^{2+} ligand, a table that encompasses all the Chl sites in FR-PSI and FR-PSII can be found in Appendix A.

Considering the issues relative to the contribution of electronegative atoms to Cryo-EM maps, careful evaluation of the experimental ESP of Chl substituents is also crucial for a reliable assignment. In the literature two main methods have attempted to quantify the ESP of the chlorophyll's C2 substituent, aimed at assigning Chl *f* sites in FR-PSI maps.

The first method, presented by Kato et al. (116), relied on computing a difference maps between the experimental map of FR-PSI and a synthetic map generated from the refined FR-PSI model in which every chlorophyll is assigned to Chl *a*. A second difference map was used in order to exclude pseudo-positive signals, by subtracting an experimental WL-PSI map to the original FR-PSI map. All the maps were normalised based on the respective root-mean-squared density values, and to account for differences in Chl positions each site in WL-PSI was superimposed both rotationally and translationally with the corresponding Chl site in FR-PSI based on the two atomic models. Chl *f* sites were then assigned in presence of a positive ESP signal in both difference maps with a minimal threshold of 5σ.

This approach originates from X-ray crystallography methods, in which it is common to search for small differences between structures by comparing two isomorphous data sets. Nonetheless, the method has issues related to the nature of Cryo-EM maps and the differences between WL-PSI and FR-PSI. In particular, while electron density (ED) maps can be scaled relatively easily, methods to scale ESP maps have only recently started to appear (167) and are unable to deal with local resolution differences, leading to problematic spurious positive peaks when generated from independent experimental maps. Moreover, since the ESP has both positive and negative components, the

interpretation of putative positive ESP difference features also differs from ED maps, since they cannot be interpreted as additional or missing atoms as they may result from differences in local resolution or B-factor (155). The scale of ESP maps is also influenced by solvent, box size and the processing/refinement algorithms used, complicating further the difference between maps originating from different datasets. Second, local differences in the chemical environment of Chl sites between WL- and FR- PSI structures are massive. Changes in amino acid sequence, as well as changes in the planarity of the chlorophyll molecules and different ligands to the Mg^{2+} can influence the refinement and consequently the difference maps produced.

The second method, similar to other more common methods available in Cryo-EM, such as EMringer (168), was developed by Gisriel et al. (155,156), who provided the first statistical framework to analyse and classify chlorophyll substituents based on controls internal to the ESP map of interest. In the method, for each chlorophyll molecule, the ESP is measured on the perimeter of the circular base of a cone with axis extending from that of the CX– CX' bond (where CX is a Chl substituent), at the expected bond length (1.23 Å) and angle (120°) of a formyl oxygen. The rings of ESP of the methyl substituent in position C7, common between Chl *a* and Chl *f*, are binned into set resolution classes and used to compute the average ESP of a methyl substituent and its standard deviation in a map of interest for each resolution. ESP rings obtained from the C2 substituents are then compared to the standard methyl distribution and declared statistically different when exceeds the 3σ confidence interval, at an angle in which energetic constraints obtained by DFT are minimized.

The main advantage of the "cone scan" method compared to the "difference maps" method is that it provides an internal reference to the map being analysed, removing the need to determine the structure of two isomorphous proteins and avoiding artifacts related to data processing when comparing ESP maps obtained from different samples, grids, or processing workflows. The presence of a large number of chlorophyll sites in a single photosystem allows for an average methyl substituent normal distribution to be confidently estimated.

Nonetheless, given the complexity of classifying chlorophyll substituents, the “cone scan” method still presents some challenges: i) the results are influenced by the goodness-of-fit of the chlorophyll model; ii) the lack of an internal calibration standard does not allow for quantification of the performance of the map/model combination; iii) the spread of resolutions of different chlorophyll sites within the map may require for a portion of the chlorophyll sites to be discarded or segmented into different resolution bins, making the method prone to being biased by the user choice; and iv) the $\mu + 3\sigma$ methyl distribution cutoff used as a classification metric does not necessarily yield the best classification accuracy.

The “cone-scan” and “difference map” methods, when applied to the same map (6KMX) have produced conflicting assignments. The latter could classify 8 chlorophyll sites as Chl *f*, while the former only managed to positively label only 5 of them. Only 3 assignments are common between the two sets (i.e., B07, B37, A23).

This chapter presents the different avenues explored to improve the classification of chlorophyll substituents in Cryo-EM maps of photosynthetic complexes, with a view to highlighting their strengths and weaknesses as well as providing information on features characteristic of Chl *f* sites. As part of the methodology, a promising new statistical framework to evaluate the nature of chlorophyll substituent in Cryo-EM maps was developed further. The framework was first compared with other methods available in the literature and then used to analyse the available maps on the EMDB of Chl *f* containing photosynthetic complexes, providing a comprehensive analysis of the characteristic features of different chlorophyll sites, as well as common pitfalls in the assignment of chlorophyll cofactors.

This method, together with other lines of evidence has driven the assignment of the Chl *f* sites in the maps obtained for FR-PSI and FR-PSII in subsequent chapters (Chapters 4 and 5).

The Phyton code implementing this method can be retrieved from:

https://github.com/giovanniconsoli/Chlorophyll-subsituents-scan

3.2 Methodology

3.2.1 Improvement of the cone scan method

The cone scan method (155,156) provides the statistical framework to quantify the ESP of chlorophyll substituents. However, by measuring the ESP only on the ring at the expected bond length of the formyl's oxygen, this approach fails to consider the full breath of information that a Cryo-EM map can provide. For this reason, a more complete approach has been developed, taking into account more of the chemical environment of the chlorophyll substituents. This modified approach is used to analyse chlorophyll sites in the following chapters (Chapters 4 and 5).

Specifically, the electrostatic potential (ESP) is interpolated on the entire surface of the cones with axis extending from that of the CX – CX' bond, where X is the IUPAC number given to the chlorin carbons that have relevant substitutions, and an aperture of 120° (2θ) in a distance range between 0 Å and 2.5 Å (d) (Fig. 3.1A). Then, the average (μ) and standard deviation (σ) of the ESP is computed on the surface of the cone for each of the substituents (C2, C3, C7, C8, C12), providing visual information on the quality of the map and on the preferential orientation of the vinyl (C3) and ethyl (C8) substituents with respect to the chlorophyll plane.

The strict cutoff to the normal methyl distribution at $\mu + 3\sigma$ presented in Gisriel et al. (155) only took into account positive variations in the ESP. However, the chemical environment of different moieties can influence the ESP both positively and negatively. Therefore, our modified approach takes into account both positive and negative changes in the ESP of the substituent and normalizes data by expressing the results in terms of Z-scores relative to the C7 methyl distribution:

$$Z_{[d,\varphi]} = \frac{ESP^{CX}_{[d,\varphi]} - \mu^{C7}_{[d,\varphi]}}{\sigma^{C7}_{[d,\varphi]}}$$

where d is the distance on the cone lateral surface from the CX' atom, φ is the torsion angle of the bond according to the IUPAC convention, $\mu^{C7}_{[d,\varphi]}$ is the average methyl ESP, and $\sigma^{C7}_{[d,\varphi]}$ is the average methyl ESP's standard deviation.

To visualise the 3D space of the cone surface, the cones are represented as 2D projection looking down the CX – CX' axis. This approach allows for the differences between substituents outside of the predicted length of the formyl's C=O double bond to be visualised (Fig. 3.1B), exploring more thoroughly the ESP of the protein environment compared with the original cone-scan method.

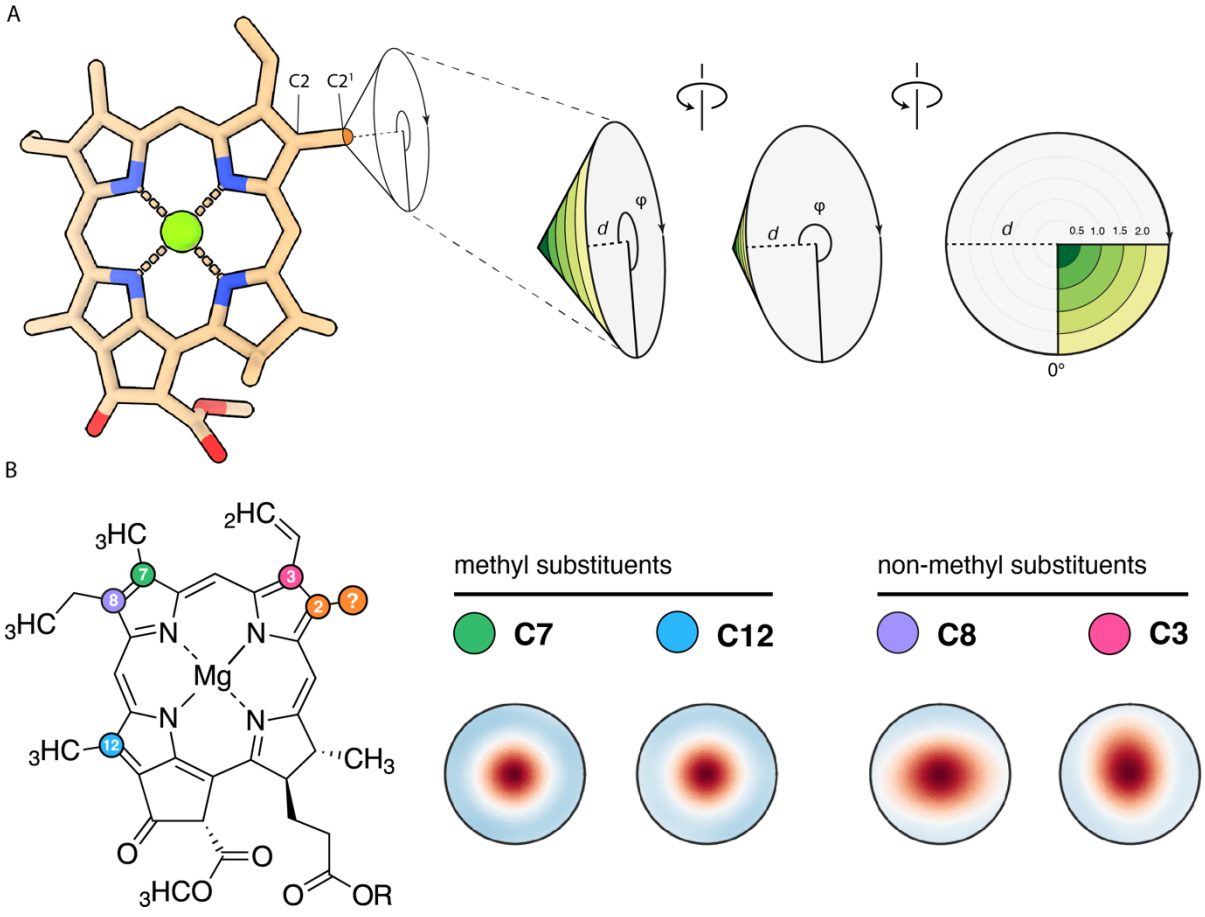


Fig. 3.1 – Chlorophyll substituent ESP cones

A) Illustration of the 2D projection of a cone with apex at the C2 substituent of a chlorophyll molecule, highlighting the distance from the C2' carbon (d) and the torsion angle (φ). B) Guide on the substituent numbering on the chlorophyll ring and a representation of the average substituent of methyl and non-methyl substituents. While the ESP of methyl substituent is round, the ESP of non-methyl substituent is higher in intensity and elongated. C3 vinyl substituents present a clear preferential orientation in plane with the chlorophyll ring, while C8 ethyl substituents have two preferential orientations off plane to the chlorophyll ring.

3.2.2 Assessment of the quality of the cone scan for a given map/model

To assess the quality of the analysis for a given map and PDB model combination, and to find an appropriate standard deviation cutoff to classify a substituent as significantly different from a methyl distribution, confusion matrices for a Z-score classification cutoff range from 0 to 5 are calculated as well as the precision, recall and accuracy metrics. The Z-score cutoff at which accuracy is maximised for a map/model combination can be used as an indication of the ideal cutoff to be used, and to help understand the nature of given substituents.

To compute these metrics, moieties in common between different types of chlorophylls, are used. In both Chl *a* and Chl *f,* C3 and C8 are non-methyl substituents, and are therefore expected to present a variation in their ESP compared to a methyl distribution. On the other hand, C7 and C12 are methyl substituents and therefore no variation in their ESP is expected compared to the methyl distribution.

For the scope of the analysis presented, true positives (TP) are defined as the non-methyl substituents that present ESP values outside of the methyl normal distribution for a given Z-score threshold. False positives (FP) are defined as the methyl substituents that present ESP values outside of the methyl normal distribution for a given Z-score threshold. True negatives (TN) are defined as the methyl substituents that do not present ESP values outside of the methyl normal distribution for a given Z-score threshold. False negatives (FN) are defined as the non-methyl substituents that do not present ESP values outside of the methyl normal distribution for a given Z-score threshold.

Precision represents the percentage of correctly assigned non-methyl substituents over the total number of positive assignments, so a high precision (close to 100%) is indicative of a low number of FP assignments:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall represents the percentage of correctly assigned non-methyl substituents over the total amount of non-methyl substituents, so a high recall is indicative of a low number of FN assignments:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Finally, accuracy represents the percentage of correctly classified substituents for a given Z-score threshold and therefore high accuracy means a low number of misclassified substituents (FP+FN):

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}}$$

3.2.3 The sphere method, a multivariate approach in Chl classification

One way to make the representation of the Chl substituent ESP more comprehensive is to analyse a section of a sphere around the chlorophyll substituent with respect to the CX – CX' as $ESP^{C2}_{[d,\theta,\varphi]}$. The information-rich 3D data generated does not allow for an intuitive representation or understanding of differences and calls for a multivariate approaches to classify Chl substituent and visualise their deviation for a given map/model combination. Two different methods are considered to analyse the 3D data.

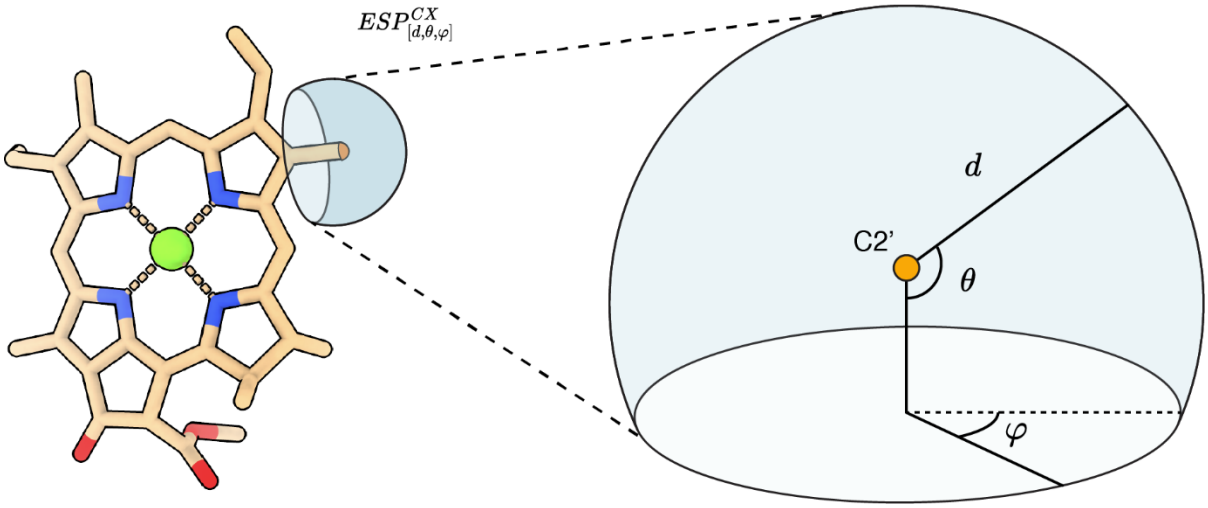


Fig. 3.2 – Illustration of the area investigated by the sphere scan method

Visual representation of the space sampled for each substituent during the sphere scan analysis. The centre of the sphere with a conical wedge removed is on the CX' Chl substituent, θ is the aperture angle from the CX-CX' axis, and d and φ represent the distance and torsion angle of the point sampled.

The first method is principal component analysis (PCA), a dimensionality reduction technique that allows to represent a high-dimensional spaces by transforming the coordinates of the system in the directions (principal components) that capture the largest variance in the data. Here, PCA was applied on a dataset composed of the hemispheres extracted from each substituent of each chlorophyll for a given map/model combination, demonstrating how, in principle, it is possible to visualise, in high-resolution maps (< 2.5 Å) two distinct groups of hemispheres, corresponding to the ESP originating from methyl and non-methyl moieties, as well as their local resolutions.

The second method is support vector machine (SVM), a supervised machine learning technique used for binary classification of data points by maximizing the margin of the data from the hyperplane used for their separation. This approach provides a labelled approach to classifing Chl substituents in Cryo-EM maps.

3.3 Results and Discussion

3.3.1 Resolution dependence of classification metrics

To understand the relationship between the cutoff value and the quality metrics of this analysis, maps of FR-PSI photosystems available on EMDB were analysed with the improved cone scan technique (Section 3.2.1). The analysis shows that for low cutoff values, the method is prone to misclassifying methyl groups as non-methyl groups, while for high cutoff values it is prone to misclassifying even ethyl (C8 in Chl *a*) and vinyl (C3 in Chl *a*) groups as methyl substituents (Fig. 3.3).

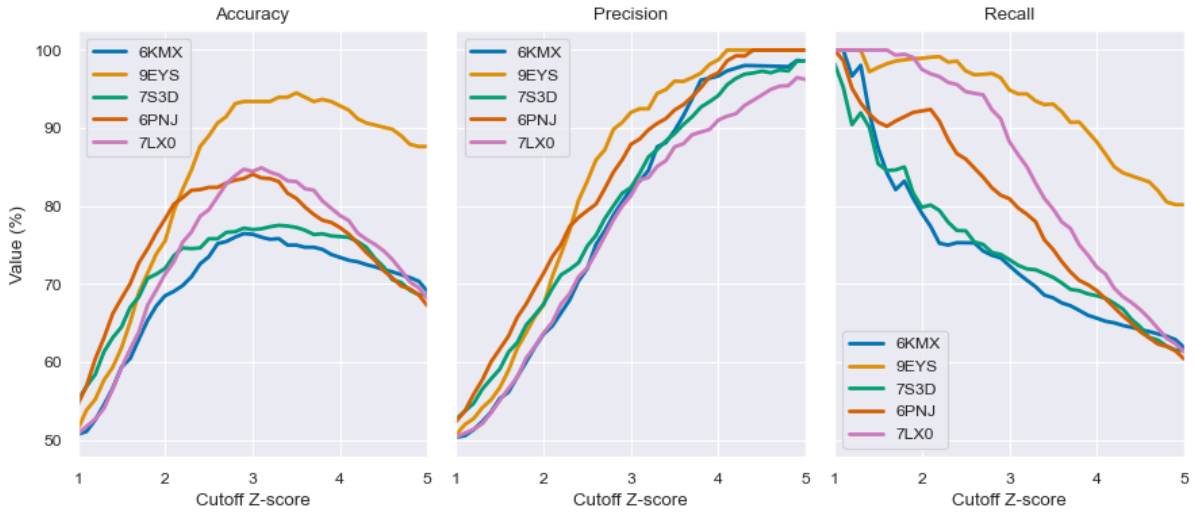


Fig. 3.3 – Accuracy, precision and recall metrics of available FR-PSI structures

The line plot represents the values in percentage of the three metrics when changing the Z-score cutoff value of the analysis for different structures available on the PDB.

The analysis demonstrates the influence of improved resolution on the ability of the improved cone scan to classifying chlorophyll substituents. In the high-resolution FR-PSI structure (2.01 and 1.88 Å) analysed in the following chapters, both the accuracy and recall metrics improve for Z-scores cutoff values larger than 2, while the precision metric seems

to be comparable across different cutoffs. As expected, higher resolution decreases the mislabelling of non-methyl substituents, leading to more robust assignments of formyl groups in the analysis performed in the following chapters.

However, improvements in the resolution of the ESP map do not necessarily result in an increased classification accuracy. This behaviour could be due to improper model building and fitting of chlorophyll molecules in lower resolution structures. The sets of restraints commonly used in the real space refinement of models of photosynthetic complexes does not fully reflect the geometry and distortions of the tetrapyrrole ring of Chl molecules when higher resolutions are achieved, leading to poor fits and consequently mislabelling in the cone scan analysis. This highlights the importance of accurate model building for a successful analysis.

Fig. 3.4 – Relationship between maximum accuracy, map resolution, and Q-score

Scatterplots represent each of the structures available in EMDB for photosynthetic complexes containing chlorophyll in terms of their GS-FSC resolution, Q-scores and maximum achieved accuracy in the improved cone scan method. PSI structures are represented in orange, while PSII structures are represented in green. The linear fitting of the datapoints (in black), predict what would be the resolution, and Q-scores needed to achieve a perfect classification of methyl and non-methyl substituents in the analysis.

By analysing most of the photosystem maps present in the EMDB, for resolution better than 3 Å, and plotting the maximum accuracy values against their GS-FSC resolution and Q scores (a measure of the resolvability of single atoms in a Cryo-EM map) a trend that

correlates resolution to the maximum accuracy achieved by the analysis can be found (Fig. 3.4). In particular, structures with lower resolutions (left panel) and worse Q-scores (right panel) tend to score worse for the maximum accuracy.

Through linear regression, it is also possible to extrapolate the values of resolution and Q-scores at which the improved cone scan would predict without false negatives or false positives the presence of an increased ESP for chlorophyll substituents. The expected resolution is around 1.44 Å, which is extremely uncommon for routinely determined structures. The expected Q-score of 0.84, is currently achieved by 0.4% of the structures deposited on EMDB only.

Overall, this analysis suggests that, with current methods, it may not be possible to unambiguously detect the presence of a single atom in ESP maps in the average ranges of resolution. Once again, this shows the importance of coupling multiple lines of evidence for the assignment of Chl *f* sites.

3.3.2 Further challenges with mislabelling of chlorophyll substituents

At the resolutions at which Cryo-EM maps of photosystems are routinely determined, the improved cone scan method fails to classify chlorophylls substituents in the ESP map with 100% accuracy. This is largely due to the changes in local resolution present in ESP maps and the quality of the model building and fitting of the chlorophyll rings in the ESP map.

Moreover, interference with ESP arising from other atoms in the range or close to the range of the cone sampled can be detrimental to the classification performance. To analyse where false positives occur, multiple map/model combinations were investigated. The survey indicates that false positive results, caused by methyl substituents being incorrectly classified as having increased ESP compared to the average methyl distribution are usually chlorophylls in which the cone lateral surface is perpendicular to that of another chlorophyll ring (Fig. 3.5). The presence of external atoms, unrelated to the chlorophyll of interest, in the range in which the cone is sampled can bias the analysis by increasing the ESP recorded by the cone scan (Fig. 3.5B, C). On the other hand, false negatives often originate from chlorophylls located at the periphery of the complexes, where local

resolution is lower and local B-factor increases. The compound effect causes the ESP of single atoms to flatten and spread over larger areas, diffusing peaks in ESP coming from single atoms and therefore not being statistically different from that of a methyl group.

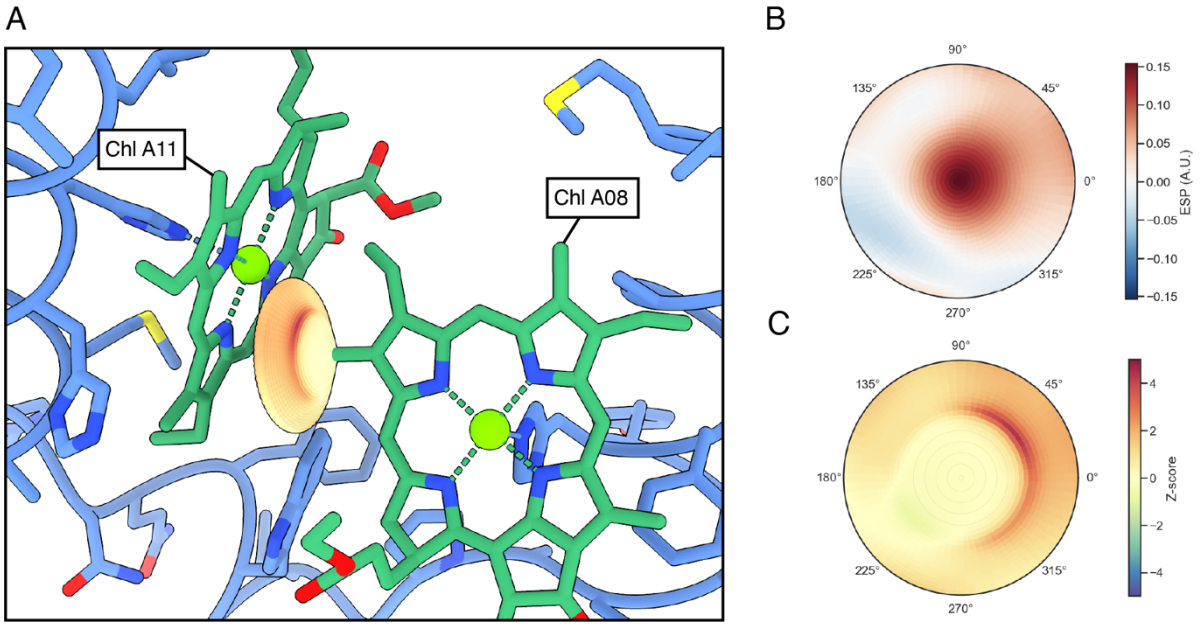


Fig. 3.5 – Example of false positive in the periphery of FR-PSI

A) Chemical environment of the Chl site A08. Chl are represented in green and PsaA2 in blue. At the C2 position a representation of the Z-score cone scan is present. B) ESP cone scan C) Z-score cone scan.

3.3.3 *Analysis of the variance of Chl substituents ESP through PCA*

The hemispheres obtained from each of the substituents of the Chlorophylls present in the FR-PSI structure were used to compute and plot four principal components. The first component explains ~55% of the variance, the second ~13%, the third ~9%, and the fourth ~7% (Fig. 3.6).

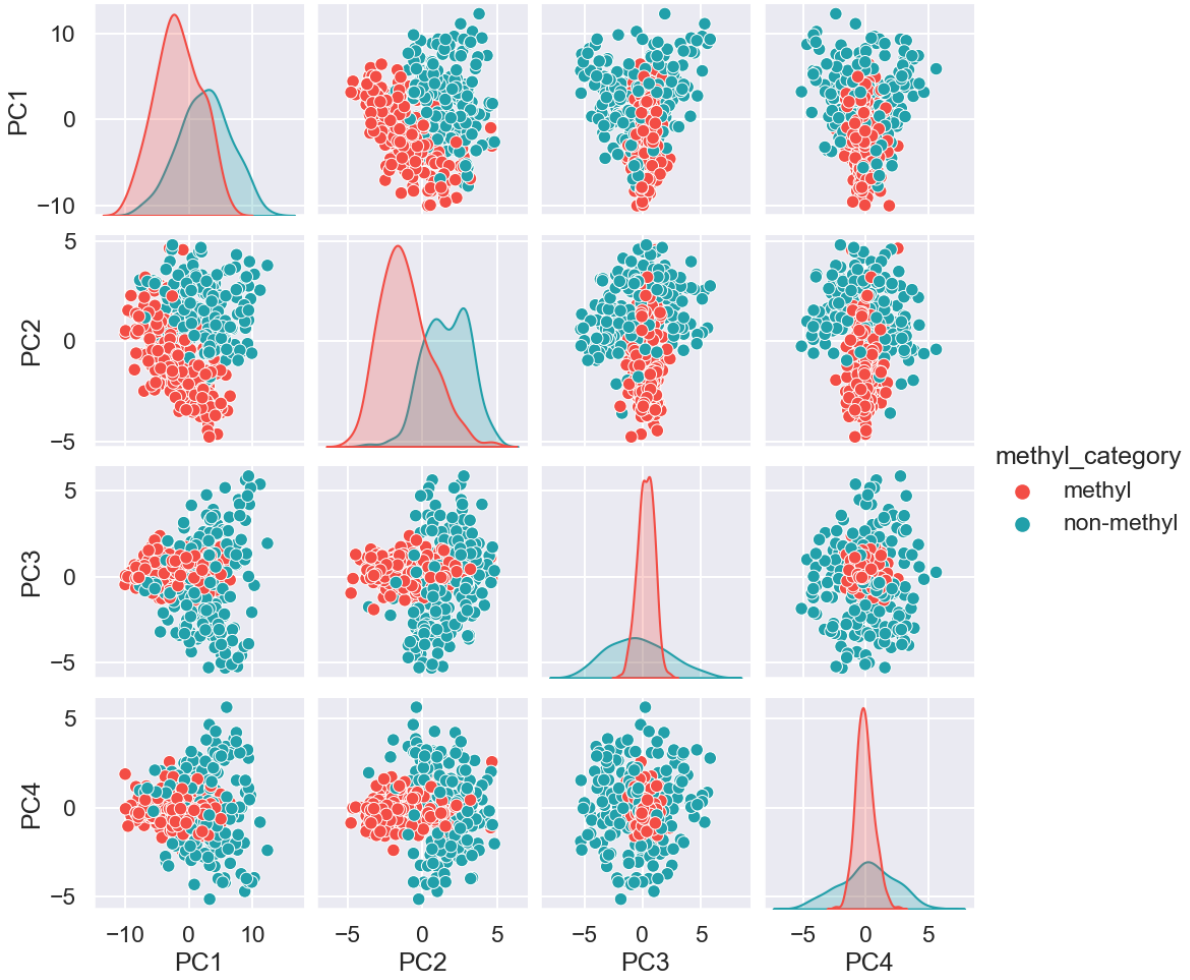


Fig. 3.6 – Scatterplot matrix of the first four principal components derived from PCA

Each panel displays the pairwise relationship between two of the four principal components, while the diagonal plots represent the distribution of "methyl" (blue) and "non-methyl" (orange) data points for each component. The scatter points are coloured by their methylation status, highlighting distinct patterns between the two categories along principal component directions. Diagonal density (KDE) plots illustrate the relative distribution projected along the direction of each principal component.

While along the direction of PC1 the two substituent types are mostly overlapped, along the direction of PC2 it is possible to distinguish two peaks in the distribution of the non-methyl substituents, suggesting that PC2 better captures the shape of the ESP of non-

methyl substituents. The two peaks are likely to correspond to ethyl and vinyl groups, that different ESP both in terms of intensity and orientation. PC3 and PC4 present a very narrow peak for the methyl groups, and a much broader distribution for non-methyl substituents, but both centered around the same value.

Closer inspection of PC1 and PC2 (Fig. 3.7) reveals a distinct pattern among different chlorophyll substituents. Along the diagonal spanning from positive values of PC1 to positive values of PC2, the substituents align in accordance with their local resolution. On the other hand, perpendicular to this diagonal, the substituents segregate into two defined clusters: methyl groups at lower PC values and non-methyl groups at higher PC values. Notably, the separation between these groups becomes sharper at higher resolutions, supporting the observations detailed in the preceding analyses of this chapter.

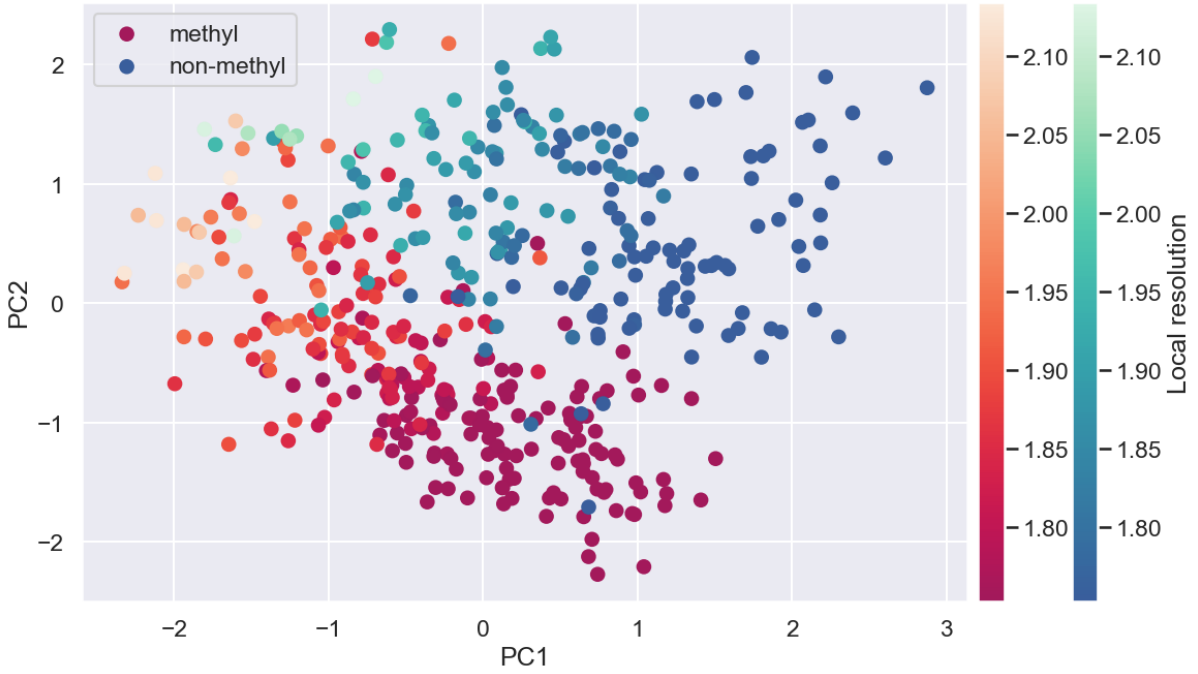


Fig. 3.7 – PCA of FR-PSI chlorophyll substituents

Principal Component Analysis (PCA) plot of chlorophyll substituents in the FR-PSI structure. The x-axis and y-axis represent the first and second principal components (PC1 and PC2), respectively. Substituents are categorised into two groups: methyl groups (blue) and non-methyl groups (red). The datapoints are further coloured based on their local resolution values with a gradient scale from -2.2 to -1.7, in two distinct gradients depending on whether they are methyl or non-methyl. The clustering and separation of substituents highlight their distinct orientations and the influence of resolution.

The analysis suggests that the effects of resolution and of the type of substituent on the ESP distribution of methyl and non-methyl substituents are deeply intertwined, and

therefore relying on low-resolution structural data to assign chlorophyll substituents might lead to data overinterpretation. Techniques used to assign specific chlorophyll substituents should aim to provide an indication on the nature of the substituents, that needs to be corroborated and taken together with different lines of evidence (i.e., phylogenetic conservation and chemical environment) to confirm the assignments.

3.3.4 Classification with SVM of the sphere scan sampling

Support vector machines (SVMs) offer several key advantages for classification tasks, particularly in information rich volumetric datasets as the ones derived from the sampling of chlorophyll substituents in Cryo-EM maps with the sphere method. SVMs are highly effective in situations where the number of dimensions exceeds the number of samples, as the algorithm focuses on finding the optimal hyperplane that maximises the margin between classes. This makes them well-suited for complex datasets where the decision boundaries are unclear. Moreover, SVMs are inherently robust to overfitting, since they maximise the margin and penalise overly complex models. The versatility of SVMs is enhanced by their ability to use different kernels, such as linear, polynomial, and radial basis (RBF) functions. This flexibility allows SVMs to handle not only linearly but also non-linearly separable data effectively, adapting to the underlying structure of the data without requiring extensive data manipulation. The above mentioned kernels were all applied to train SVM classifiers. The models were trained using 80% of the substituents of known nature (C3, C7, C8, C12) and then cross-validated on the remaining 20%. This process was repeated five times to obtain averaged accuracy precision and recall metrics. The average accuracy of the linear SVM is of 99.8%, of the third order polynomial 97.4% and of the RBF of 98.4% on the test set.

The higher accuracy of the linear kernel compared to more complex ones like the polynomial and RBF can indicate two distinct behaviours. On one hand, the better performance can indicate that the data is linearly separable, allowing for a straight hyperplane to effectively classify the data collected from the ESP sphere of each substituent. On the other hand, given the relatively small size of the dataset, together with the highly dimensional feature space of the ESP sphere the linear model might be

presenting a degree of overfitting, but this is unlikely since the cross validation presents high accuracy metrics.

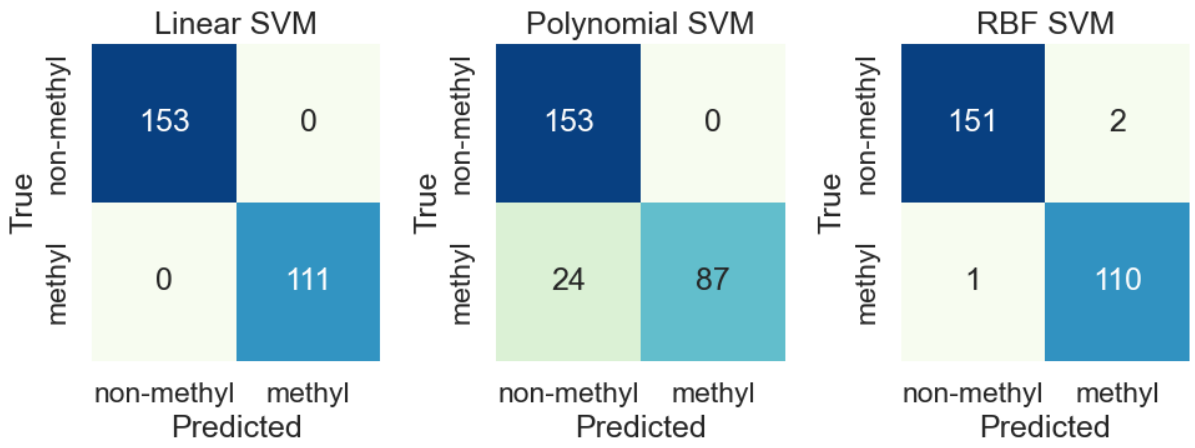


Fig. 3.8 – Confusion matrices of different SVM classifiers

Each heatmap represents the classifier's performance during cross-validation: linear SVM, polynomial SVM (degree 5), and RBF SVM. The x-axis indicates the predicted substituent type, while the y-axis represents the expected substituent type. Top-left corner of each matrix represents true positives (non-methyl predicted as non-methyl). Bottom-right corner represents true negatives (methyl predicted as methyl). Top-right corner shows false negatives (non-methyl misclassified as methyl). Bottom-left corner indicates false positives (methyl misclassified as non-methyl). Darker cells correspond to higher counts, showing the distribution of predictions.

All of the three models were applied to the final test set composed of all of the hypothesised Chl *f* candidates that will be presented in the next chapter (Table 1). All of the three models perform well on the test set, where they correctly classify the formyl substituents at the C2 position of the hydrogen bonded substituents, including the more debated non-hydrogen bonded formyl oxygens such as A_{-1B} and in some cases L03 and B38.

	Ring position	Local resolution	expected substituent	Linear SVM	Poly SVM	RBF SVM
A-1B	C2	1.75	formyl	1	1	1
A-1B	C3	1.75	vinyl	1	1	1
A-1B	C7	1.75	methyl	0	0	0
A-1B	C8	1.75	ethyl	1	1	1
A21	C2	1.86	formyl	1	1	1
A21	C3	1.86	vinyl	1	1	1
A21	C7	1.86	methyl	0	0	0
A21	C8	1.86	ethyl	1	1	1
A23	C2	1.79	formyl	1	0	1
A23	C3	1.79	vinyl	1	1	1
A23	C7	1.79	methyl	0	0	0
A23	C8	1.79	ethyl	1	1	1
B07	C2	1.75	formyl	1	1	1
B07	C3	1.75	vinyl	1	1	1
B07	C7	1.75	methyl	0	0	0
B07	C8	1.75	ethyl	1	1	1
B30	C2	1.75	formyl	1	1	1
B30	C3	1.75	vinyl	1	1	1
B30	C7	1.75	methyl	0	0	0
B30	C8	1.75	ethyl	1	0	1
B37	C2	1.75	formyl	1	1	1
B37	C3	1.75	vinyl	1	1	1
B37	C7	1.75	methyl	0	0	0
B37	C8	1.75	ethyl	1	1	1
B38	C2	1.75	formyl	1	0	0
B38	C3	1.75	vinyl	1	1	1
B38	C7	1.75	methyl	0	0	0
B38	C8	1.75	ethyl	1	1	1
L03	C2	1.75	formyl	1	0	0
L03	C3	1.75	vinyl	1	1	1
L03	C7	1.75	methyl	0	0	0
L03	C8	1.75	ethyl	1	1	1

Table 1 - SVM classification results of Chl *f* candidates in the FR-PSI

Classification results of the different SVM models (Linear, Polynomial, and RBF) for predicting the Chl substituent type. Each row corresponds to a specific substituent of a Chl, with the local resolution, the expected substituent type (ground truth) and the predictions from the three models. The substituent type is encoded as follows: 0 represents methyl, and 1 represents non-methyl. The table highlights the models' ability to correctly classify substituents compared to the expected outcomes.

The results obtained with this method are promising but classifying substituents with a high degree of accuracy would require higher resolution maps compared to the cone scan method. Moreover, the results of the classification are not as interpretable as the cone scan methods presented previously, providing no information on the ESP or the chemical environment of the substituents. While using the improved cone scan method the variation in ESP can be assessed even when not significant, providing information on whether a higher resolution structure might be of help in determining the nature of a specific substituent. By contrast, the SVM approach does not provide this information.

Methods to obtain calibrated probabilities and confidence intervals for the classification predictions could provide further insight into the classifier's uncertainty, also the incorporation of Platt scaling (169) could improve the interpretability of the results for specific chlorophyll substituents. However, exploring such methodologies is beyond the scope of the current thesis.

3.4 Conclusions

With the increasing number and improving resolution of photosynthetic complex Cryo-EM structures available, better frameworks to classify different type of chlorophylls from their ESP density are needed. In this chapter, two methods that aim to achieve the classification of Chl substituents have been presented. The first method provides information on the chemical environment of Chl substituents by sampling the ESP around them and using the average ESP distribution of a methyl group as a reference. The second one, samples an hemisphere of density around the substituents that can then be analysed through dimensionality-reduction techniques such as PCA and supervised classification algorithms such as SVM.

The first method, referred to as improved cone scan, has demonstrated to be a solid starting point in guiding the assignment of different types of chlorophyll. The statistical framework, based on determining the average methyl electrostatic potential introduced previously (155) allows to adopt the map directly as a refence for the classification and for evaluating the performance of the analysis case by case. Nonetheless, given the reliance of the cone scan method on the quality of the fit of chlorophyll molecules in the Cryo-EM maps, it is important that the model/map combination analysed are refined with care and the correct sets of restrains and that the assignments obtained are corroborated by different lines of evidence (e.g. phylogenetic conserved changes, spectroscopy, functional and biophysical measurements) .

According to Gisriel et al. (155) the resolution considered necessary to directly observe formyl substituents of chlorophyll molecules in Cryo-EM maps is around ~2.3 Å depending on map quality. This empirical observation, based on the maps available at the time, fails to reflect the complexity of the assignment. While a local resolution of ~2.5 Å is enough to confidently assign hydrogen bonded formyl groups, non-hydrogen bonded formyl groups present on average a much lower positive ESP, and a much higher resolution (< 2 Å) is required. The analysis performed with the cone scan method on maps available in EMDB clearly confirms this, indicating that an unambiguous classification of a Chl substituent could require resolutions as high as ~1.4 Å. Since non hydrogen bonded

formyl groups present on average a much lower maximum positive Z-score in resolutions ranging between 2 and 3 Å, the absence of an increased ESP does not rule out the presence of a formyl group in the site analysed, and the visual and statistical observation of an increase ESP potential should be weighted accordingly.

The second method presented, referred to as sphere scan, aims to build an analytical workflow that is less reliant on the quality of fit of the chlorophyll molecules and on human intervention. It has performed well on hydrogen bonded formyl groups, but lacks the interpretability and the information on the chemical environment of the Chl substituents provided by the improved cone scan method. The two methods could act in concert to interpret the ESP of chlorophyll substituents, on one hand providing a statistical comparison with the potential of a methyl group and, on the other hand, providing a supervised classification method that could robustly classify chlorophyll substituents automatically in cases where the resolution of the ESP map allows it.

The use of the sphere scan coupled with PCA to understand what factors can influence the ESP of Chl substituent has highlighted how the shape and intensity of the ESP potential in presence or absence of an increased potential arising from an extra atom is tied to the local resolution and the chemical environment of each Chl substituent independently. Nonetheless, the lack of interpretability of the SVM sphere method does not allow for the ESP density to be visualised and the method does not provide information on the chemical environment of each Chl substituent. Even though it is a promising approach towards the automatic classification of Chl substituents, it is still not competitive with the improved cone scan.

The cone scan method has efficiently aided the assignment of Chl *f* sites in the structures of FR-PSI and FR-PSII that are presented in the following chapters. Moreover, it has allowed direct comparison of the quality of our map with the ones presented in literature as well as assessing the goodness-of-fit of different map model combinations that were generated throughout the refinement of structures.

4 Locating the missing chlorophyll *f* molecules in Cryo-EM structures of FR-PSI

4.1 Introduction

Photosystem I is the second light-driven complex of the photosynthetic electron transport chain. It catalyses a single-turnover oxidoreduction reaction, oxidising plastocyanin on the lumenal donor side of the thylakoid membrane and reducing ferredoxin on the stromal side (24). Canonical white light photosystem I (WL-PSI) is found in different oligomerisation states in oxygenic photosynthetic organisms (25) (i.e., monomers, dimers, trimers and tetramers). Each monomer is composed of 12 subunits that coordinate over 90 chlorophyll molecules, most of which act as antenna pigments to funnel excitation energy to the 6 reaction centre Chl *a* molecules that perform charge separation (27,28). The FaRLiP cluster contains 6 far-red-adapted paralogous subunits of PSI (PsaA2, PsaB2, PbaI2, PsaJ2, PsaL2, and PsaF2) that assemble, together with six white-light-shared subunits (PsaC, PsaD, PsaE, PsaK, PsaM, and PsaX) to form FR-PSI (112). The complex contains approximately the same number of total chlorophylls as WL-PSI (91 in FR-PSI and 96 in WL-PSI). Chlorophyll quantification analyses report 0 Chl *d* and 6 to 8 Chl *f* molecules per monomer but vary significantly depending on the level of intactness of the samples analysed, and possibly the species of origin (43,170).

Originally, the long-wavelength pigments were thought to play only a light-harvesting role (171), with heat from the environment being required for uphill excitation energy transfer to the higher-energy, photochemically active, Chl *a* primary donor. Nevertheless, subsequent biophysical studies on FR-PSI provided the opposite evidence, demonstrating that the complex is capable of efficient far-red-light-driven charge separation at cryogenic temperatures (T = 4 K) (43), suggesting the presence of a long-wavelength pigment in the reaction centre. In the same work, it was also suggested that this long-wavelength pigment could not occupy the A_{0A} or A_{0B} position due to the unsuitability of Chl *f* in a such a low-potential role and could not occupy P_A or P_B due to the presence of conserved electrochromic features around 700 nm in the $P700^+$ *minus* P700 difference spectra,

leaving as possible candidates only one or both A_{-1} pigments. When Cryo-EM structures of FR-PSI were released, however, they showed no evidence of structural changes or extra electrostatic potential associated with a formyl group in the C2 position among the redox-active chlorophylls (112,116,129,156). This, together with the lack of phylogenetically conserved residues specific to the FR-PSI reaction centre, was used to argue against the presence of a photochemically active Chl *f*. Subsequent spectroscopic studies based their interpretations on this conclusion (118,166,172) and, to reconcile the low-temperature, long-wavelength photochemistry with the absence of a redox-active Chl *f*, a model was proposed in which excitation energy is transferred from Chl *f* antenna pigments to a putative enhanced long-wavelength charge-transfer absorption band assigned to a Chl *a* pair in the reaction centre (166). Nonetheless, the resolution obtained for the three Cryo-EM structures available for FR-PSI (112,116,129) does not allow the direct detection of a non-hydrogen-bonded formyl group, and this is reflected in these conflicting assignments in these structures, which share only one Chl *f* site and classify at best only five of them. For this reason, it was suggested that, aside from the five hydrogen-bonded sites, the other pigments were present in the system only with partial occupancy in peripheral antenna sites (156). Given the intrinsic difficulties in detecting Chl *f* in Cryo-EM ESP maps, higher resolution structures are needed to obtain reliable structural assignments for all of the 8 Chl *f* sites and to resolve the key mechanistic controversy over the presence of a long-wavelength pigment among the redox-active chlorophylls of FR-PSI.

To address these questions, two independent high-resolution Cryo-EM maps of trimeric FR-PSI from *C. thermalis* PCC 7203 at resolutions respectively of 2.01 Å and 1.89 Å were obtained and analysed with the methods presented in Chapter 3, providing evidence for the location of all the 8 Chl *f* sites (170). Among newly assigned Chl *f* sites, the reaction centre chlorophyll A_{-1B} is proposed as a candidate for the long-wavelength primary electron donor in PSI (34,43,170,173,174), while L03 is proposed as an excitation energy sharing pigment among different monomers (170,175). Finally, a tentative assignment of the low-temperature absorption peaks to specific Chl *f* sites is proposed, and the implications of a Chl *f* primary donor are explained together with the importance of the trimeric oligomerisation state for FR-PSI.

4.2 Results

FR-PSI complexes were isolated from *C. thermalis* PCC 7203 with two distinct methods as described in the Materials and Methods section. Two Cryo-EM ESP maps were generated through SPA at a global resolution of 2.01 Å and 1.89 Å (Fig. 4.1). The difference in the resolution obtained from the two samples is related to the pixel size, particle density, and number of micrographs collected, rather than heterogeneity or intactness due to the two different purification methods. The two maps present defined ESP backbones to fit all of the 12 FR-PSI subunits, comprising six far-red-light-specific subunits: PsaA2, PsaB2, PbaI2, PsaJ2, PsaL2, PsaF2 and six white-light-shared subunits PsaC, PsaD, PsaE, PsaK, PsaM, PsaX. Each monomer coordinates the following cofactors: 22 β-carotene, 83 Chl *a*, 8 Chl *f*, 2 phylloquinones, 3 iron sulfur clusters, and various lipids.

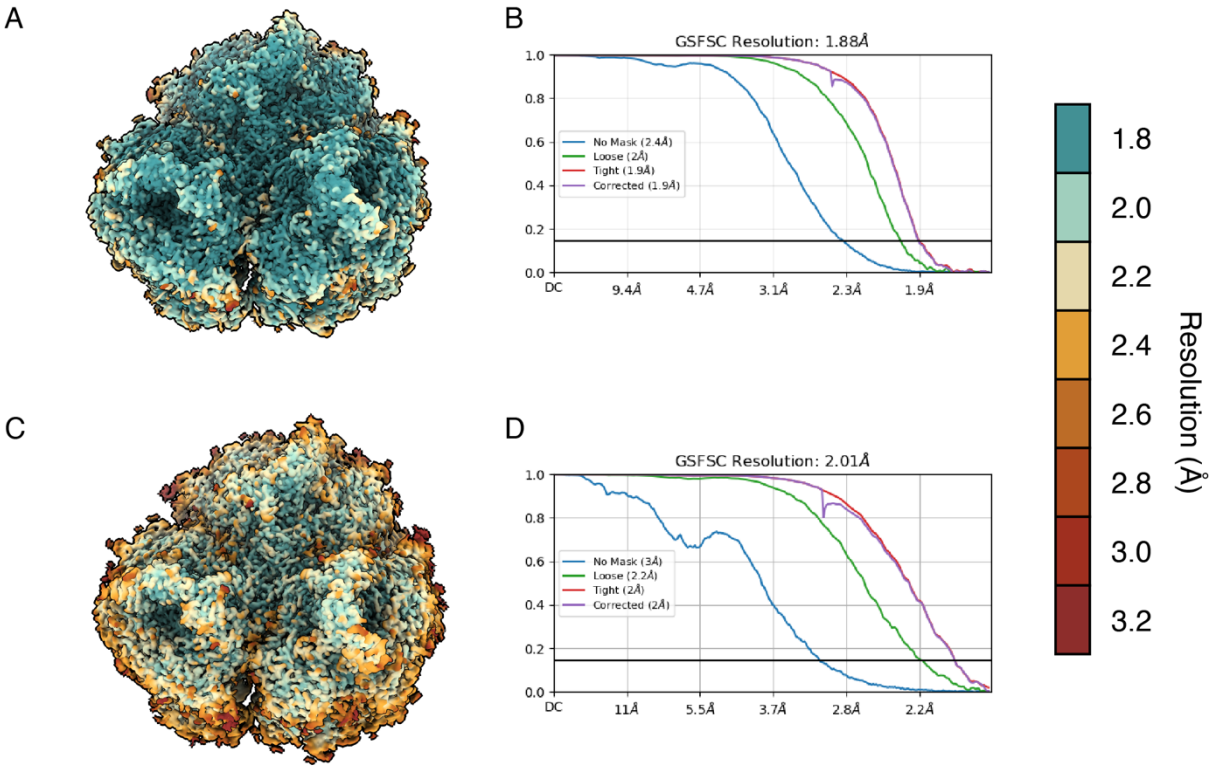


Fig. 4.1 – Local resolution of FR-PSI maps

Visual representation of the local resolution and GS-FSC of the structures discussed in this chapter, the colour-map is provided on the right. A) Trimeric FR-PSI isolated by anion exchange chromatography (PDB ID: 9I9L) B) GS-FSC of trimeric FR-PSI isolated by anion-exchange chromatography (PDB ID: 9I9L) C) Trimeric FR-PSI isolated by sucrose gradient (PDB ID: 9EYS) D) GS-FSC of trimeric FR-PSI isolated by sucrose gradient (PDB ID: 9EYS)

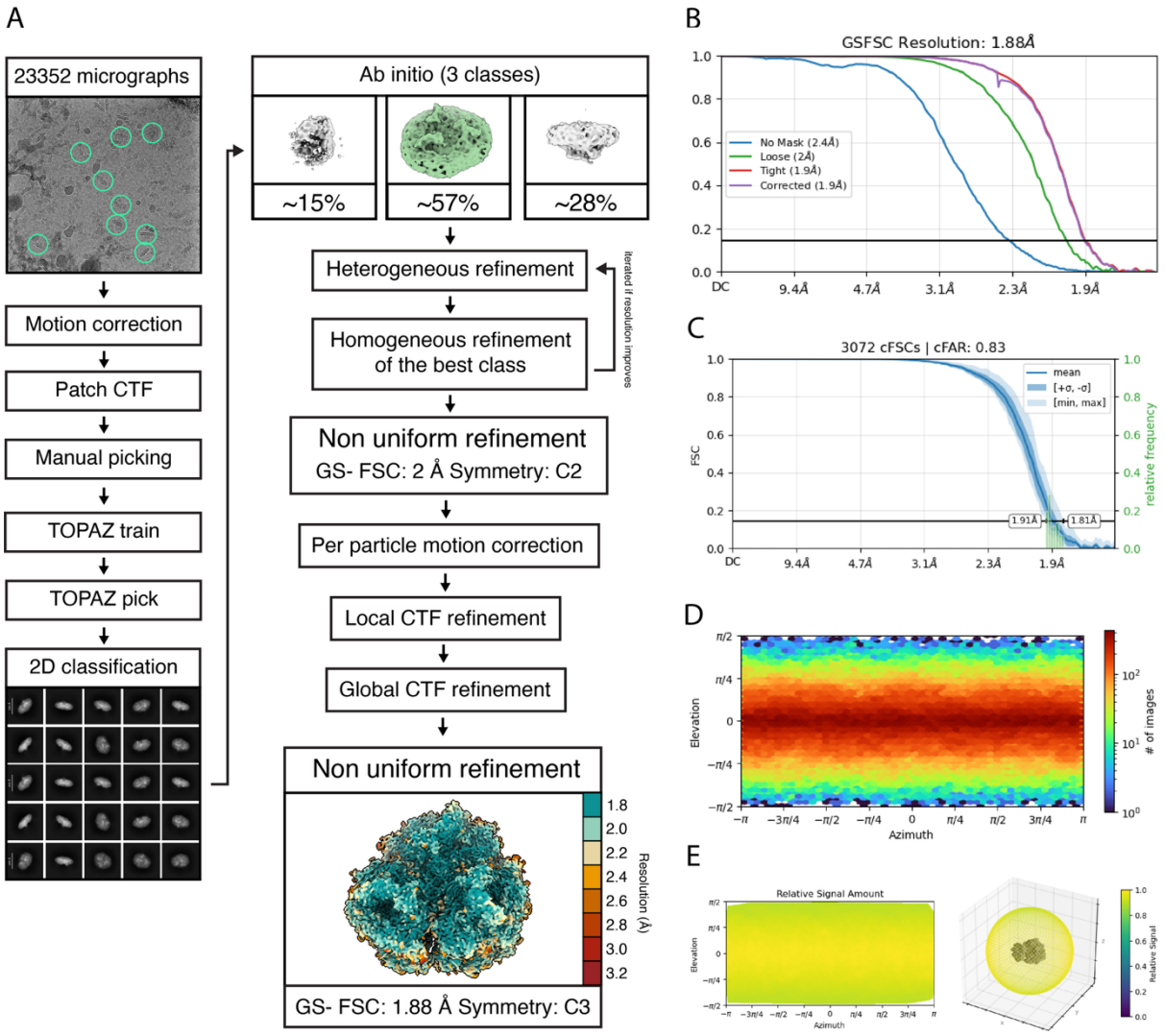


Fig. 4.2 – Processing workflow and reconstruction diagnostic plots for trimeric FR-PSI (9I9L)

A) processing workflow to obtain the final reconstruction of the trimeric FR-PSI. B) GS-FSC curve for the trimeric FR-PSI. C) cFSC curves for the trimeric FR-PSI complex. D) Viewing direction distribution of the particle stack. E) relative signal amount for each of the viewing directions.

Data collection

Microscope	Krios III
Camera	Falcon 4i
Magnification	155000x
Voltage (kV)	300
Electron exposure (e-/Å^2)	40
Defocus range (μm)	-0.8 to -2
Pixel size (Å)	0.723
Energy filter	Selectris (20 eV)
Exposures	23331
Image format	EER

Data processing

Box size	600 px
Initial particles (no.)	1229384
Final particles (no.)	297795
Symmetry	C3
Map resolution (Å)	2.51
Map sharpening *B* factor	-47.8

Model Refinement

Refinement package	PHENIX
Initial model used	9EYS
Real/reciprocal space	Real Space
Resolution cutoff	1.90

Model Validation

MolProbity score	1.28
ClashScore	5.00
Bond length R.M.S.D. (Å)	0.001
Bond angles (°)	0.560
Poor rotamers	0.00%
Favored rotamers	99.89%
Ramachandran outliers	0.00%
Ramachandran favored	98.17%

Table 2 – Data collection and processing parameters for trimeric FR-PSI (9I9L)

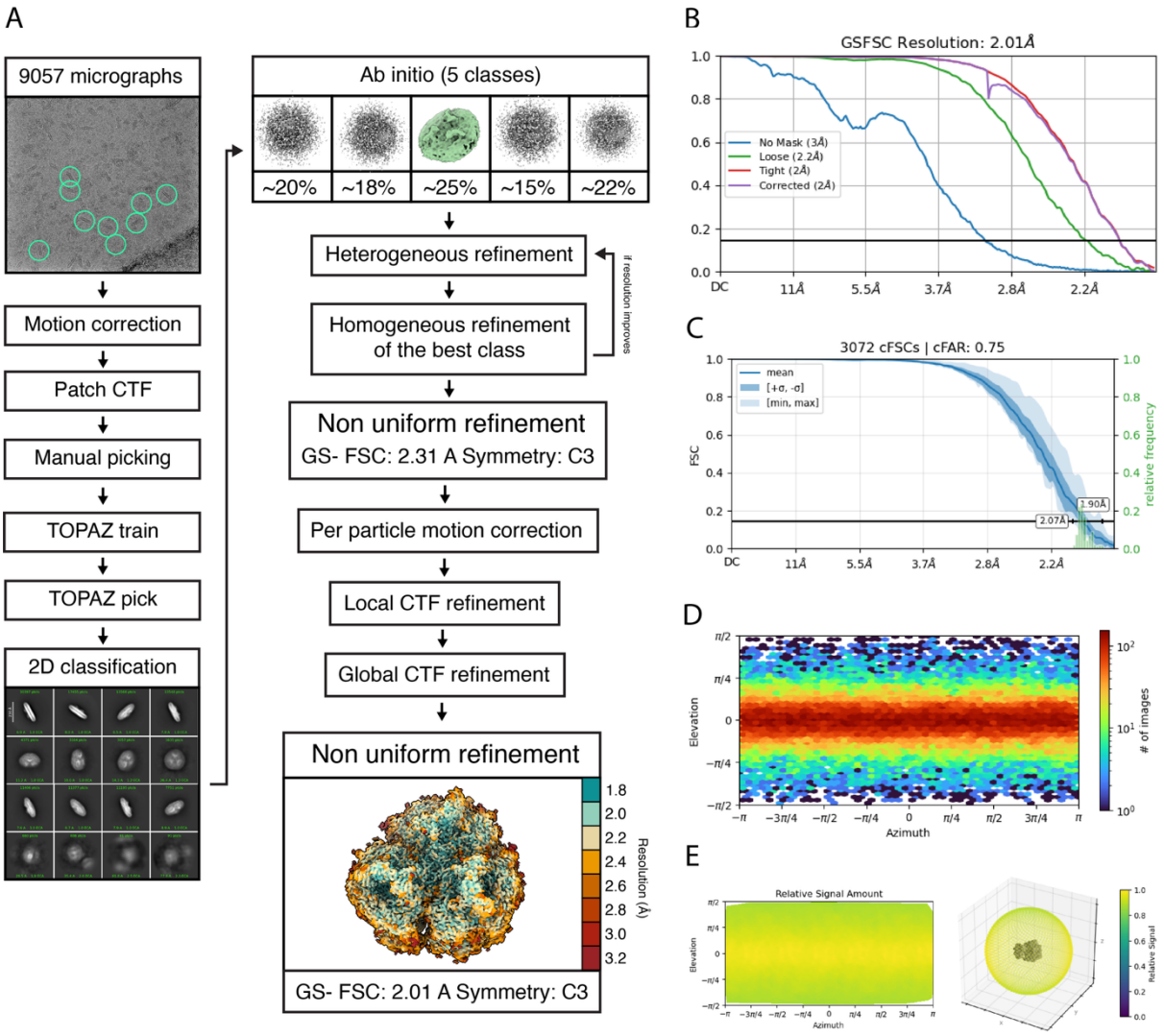


Fig. 4.3 – Processing workflow and reconstruction diagnostic plots for trimeric FR-PSI (9EYS)

A) processing workflow to obtain the final reconstruction of the trimeric FR-PSI. B) GS-FSC curve for the trimeric FR-PSI. C) cFSC curves for the trimeric FR-PSI complex. D) Viewing direction distribution of the particle stack. E) relative signal amount for each of the viewing directions.

Data collection

Microscope	Krios III
Camera	Falcon 4i
Magnification	130000x
Voltage (kV)	300
Electron exposure (e-/Å^2)	40
Defocus range (μm)	-0.5 to -3.5
Pixel size (Å)	0.921
Energy filter	Selectris (20 eV)
Exposures	8.854
Image format	EER

Data processing

Box size	600 px
Initial particles (no.)	138048
Final particles (no.)	77852
Symmetry	C3
Map resolution (Å)	2.01
Map sharpening *B* factor	-39.6

Model Refinement

Refinement package	PHENIX
Initial model used	7FIX
Real/reciprocal space	Real Space
Resolution cutoff	2.10

Model Validation

MolProbity score	1.52
ClashScore	9.98
Bond length R.M.S.D. (Å)	0.003
Bond angles (°)	0.560
Poor rotamers	0.00%
Favored rotamers	99.89%
Ramachandran outliers	0.00%
Ramachandran favored	98.17%

Table 3 – Data collection and processing parameters for trimeric FR-PSI (9EYS)

The improved resolution and methods used have allowed the detection of the formyl groups from the ESP map for all eight Chl *f* molecules in *C. thermalis* FR-PSI, that occupy the following sites: A_{-1B}, A21, A23, B07, B30, B37, B38, and L03 (Fig. 4.4). Five of the Chl sites (A21, A23, B07, B30, and B37) are in agreement with the most recent assignments published in the literature (112,116,129).

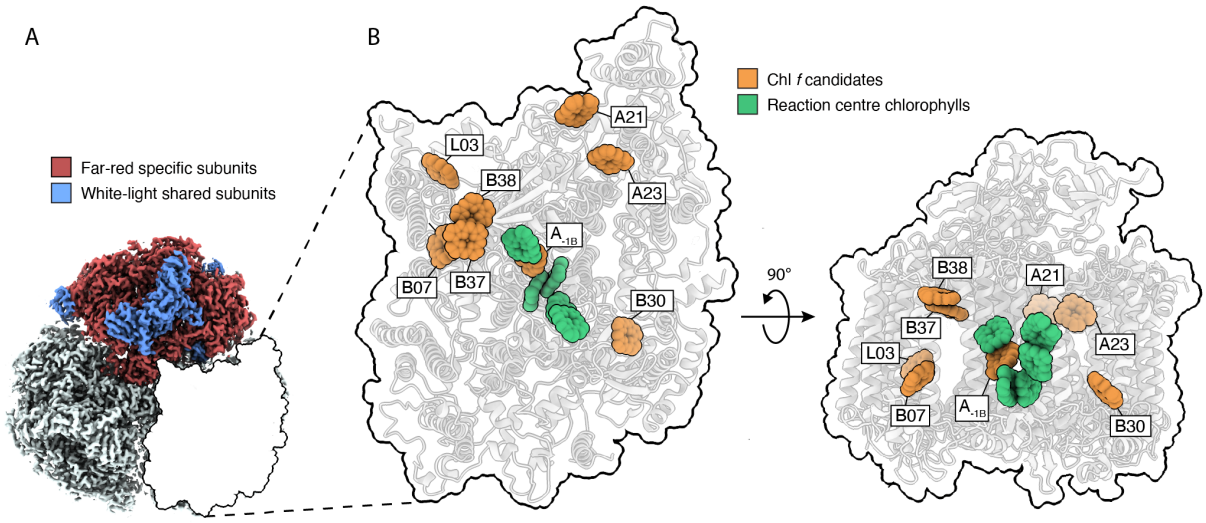


Fig. 4.4 – Structure of trimeric FR-PSI and chlorophyll *f* locations.

A) View of trimeric FR-PSI map from the cytoplasmic side. The map presents clear ESP for all the known PSI subunits: the six far-red light specific subunits (PsaA2, PsaB2, PsaF2, PsaI2, PsaJ2, PsaL2) are represented in red, while the six subunits in common with white light-PSI (PsaC, PsaD, PsaE, PsaK, PsaM, PsaX) are represented in blue. B) Location of the 8 candidate Chl *f* molecules (in orange) in a FR-PSI monomer, as viewed from the cytosolic face and membrane plane. The 6 redox active chlorophylls are shown in the centre of the monomer (in green when Chl a), Chl f is in the A-1B site while the 5 others are Chl *a*.

The presence of a C2 formyl Hydrogen-bond in five of these cases (A21, A23, B07, B30, and B37) likely contributes to the ESP signal by fixing the possible orientations of the formyl group and by diminishing the charge on the negatively polarised formyl oxygen atom (see chapter 3 for further discussion). In the following section, an in-depth analysis of each Chl *f* site is presented, followed by a discussion on the functional significance of the sites, and a list of structural features that diverge from the canonical WL-PSI.

4.2.1 *Chl f* A_{-1B}

The reaction centre chlorophyll A_{-1B} was suggested, based on biophysical evidence, to be the (or one of the) long-wavelength pigments that drives low energy charge separation in FR-PSI. In both WL-PSI and FR-PSI, the site features a water molecule coordinating the central Mg^{2+} atom, as preferred by Chl *f* molecules, but does not present potential hydrogen-bond partners in proximity to the C2 substituent. This apparent lack of conserved amino acid changes near the A_{-1} sites in FR-PSI has been used as an argument against the presence of a Chl *f* at this location (112). However, a conserved change, from Ile/Leu/Val (depending on the species) to PsaA2 Cys^{463}, is in fact present in all the available sequences of FR-PSI (Fig. 4.5D). The conserved sulfhydryl group of PsaA2 Cys^{463} and the C2 formyl of Chl *f* A_{-1B} are separated by the benzyl ring of PsaA2 Phe^{459}, suggesting a novel and unusual type of interaction between this residue and A_{-1B} (Fig. 4.5A).

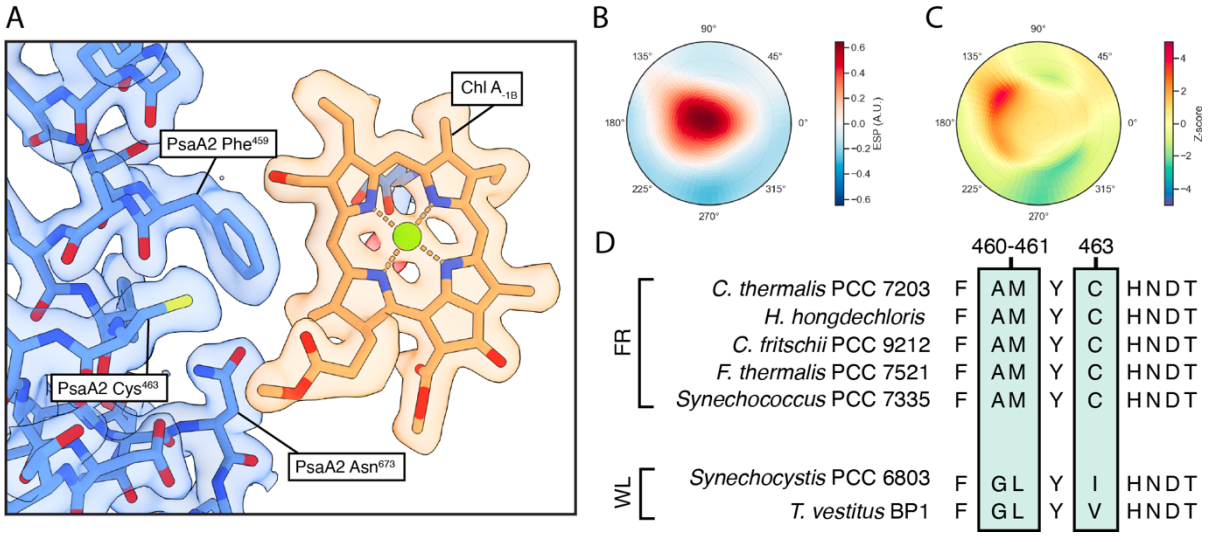


Fig. 4.5 – Evidence for A_{-1B} being a Chl *f*

A) The chemical environment of the A_{-1} Chl on the B branch of the FR-PSI reaction centre. The Chl density is in orange, the density of the amino acids from PsaA2 in blue. B) The ESP polar plot showing the presence of an increased positive potential at 145° torsion angle around the C2 substituent. C) The Z-score polar plot highlighting the presence of a statistically significant (Z-score > 3) increase in positive (145° torsion angle) and negative (270° torsion angle) ESP features around the C2 of the A_{-1B} Chl relative to the averaged ESP environment around a methyl substituent (fig. S1). D) The relevant part of the multiple sequence alignment of FR and WL PsaA, showing the far-red specific changes associated with the A_{-1B} site.

The influence of PsaA2 Cys^{463}, as well as the presence of the formyl oxygen on the C2 substituent, was analysed using the improved cone scan method presented in chapter 3.

The positive increase in ESP can be attributed to the presence of the Chl *f* formyl group, while the negative ESP feature present at ~270°, at the interface between the formyl carbon-oxygen double bond and the benzyl ring of PsaA2 Phe^{459} (Fig. 4.5), suggests an influence of the SH–π motif on the reaction centre Chl *f* formyl group, which could affect its binding affinity to allow its site-specific insertion and its redox tuning as a primary donor. Moreover, other conserved changes with possible functional significance are present near the A_{-1B} site (Fig. 4.5D, Fig. 4.6).

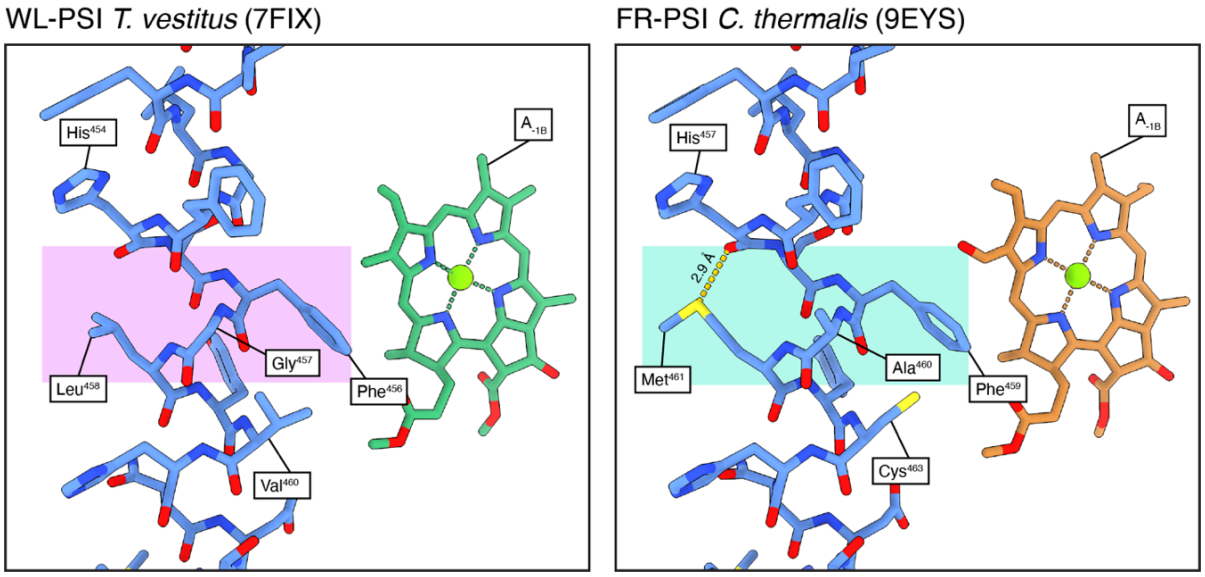


Fig. 4.6 – Structural comparison of Helix N PsaA and PsaA2 in the region A_{-1B}

The panels show a structural comparison of the PsaA and PsaA2 in high-resolution structures of WL-PSI (7FIX) and FR-PSI (9EYS) respectively. PsaA and PsaA2 are represented in blue, the destabilised helix turn in WL-PSI is highlighted in pink and for FR-PSI the same region with the additional stabilizing interaction, is highlighted in cyan, Chl *a* is represented in green and Chl *f* is represented in orange.

In both forms of PSI, the helix turn that contains the Phe^{456}/Phe^{459} facing Chl A_{-1B} is distorted, due to the pull exerted by the central Mg^{2+} of Chl A31 (not shown) on His^{454}/His^{457}. In WL-PSI, this results in the backbone carbonyl of PsaA His^{454} being too distant from the amide backbone of PsaA Leu^{458} to be involved in an α-helical hydrogen bond (~5 Å). While in WL-PSI this specific turn is likely to be flexible due to the lack of the α-helical hydrogen bond and the presence of PsaA Gly^{457}, in FR-PSI, two of the conserved changes, PsaA2 Ala^{460} and PsaA2 Met^{461}, seem to stabilise the helix turn. PsaA2 Met^{461} interacts with the carbonyl of PsaA2 His^{457}, re-establishing a connection between the two helix turns. This could have consequences on the rigidity of this portion of the

protein, especially during the electron transfer process, and for holding in place the two residues, PsaA2 Phe^{459} and PsaA2 Cys^{463}, which are proposed here to be important for the function of Chl *f* A_{-1B} as the primary donor.

Another conserved far-red change near Chl A_{-1B} is PsaA2 Asn^{673} (Ile in WL-PSI). This amino acid is in proximity of PsaA2 Cys^{463}, and the torsion angle of its side chain is different in the two maps of FR-PSI presented in the chapter. In 9EYS, PsaA2 Asn^{673} is turned away from the sulfhydryl moiety of PsaA2 Cys^{463}, while in the higher resolution structure it is facing directly towards it and it is in hydrogen bonding distance. This behaviour could indicate that this residue is important to regulate the redox properties of the Chl *f* A_{-1B} by readjusting the ESP environment of the primary donor upon charge separation.

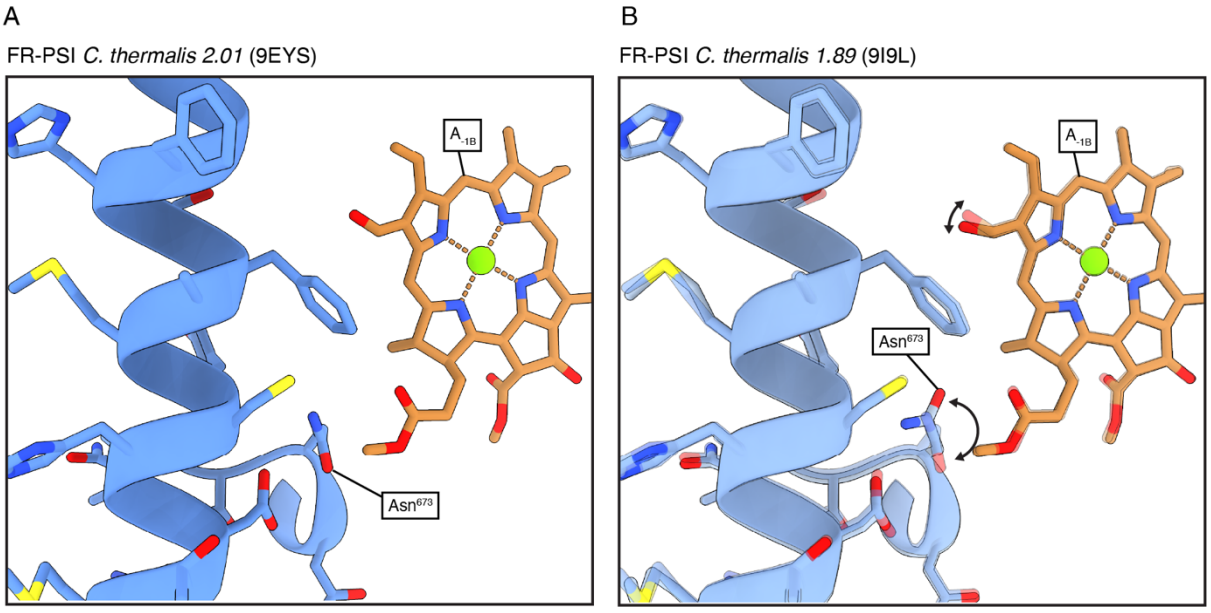


Fig. 4.7 – Conserved changes near Chl A_{-1B} influence the position of the formyl group

Structural comparison of the PsaA2 in the two FR-PSI models presented in the chapter. A) In 9EYS the side chain of PsaA2 Asn^{673} presents a torsion angle in which the nitrogen atom is close to PsaA2 Cys^{463}. B) In 9I9L the side chain presents a torsion angle in which the oxygen and nitrogen atoms are equally distant from the sulfhydryl group of PsaA2 Cys^{463}.

4.2.2 Chl f A21

The ESP map around Chl A21 suggests that this site is occupied by a Chl *f*, with its formyl group hydrogen bonded (~2.8 Å) by the backbone amide of PsaA2 Leu^{333} (Fig. 4.8A) confirming the previous assignment (112,116,156). Statistical analysis of the local ESP around the C2 substituent confirms the presence of a formyl group (Fig. 4.8B, C). Leu^{333} is part of a far-red specific loop insertion of 8 amino acids in PsaA2 (Fig. 4.8D). In addition, the loop contains a conserved Met^{332}, the sulfur atom of which is located at 3.45 Å from the formyl oxygen, suggesting possible effects on the site energy of this pigment. A water molecule acts as the axial ligand for the Mg^{2+} and forms an hydrogen bond network that involves PsaA2 Gln^{323}. In WL-PSI, A21 is coordinated by PsaA His^{332} (28). The insertion disrupts this site by changing the ligand of Chl A21 from His to a water and stabilising the environment of the formyl group. The ESP map from *Synechococcus sp.* PCC 7335 and *H. hongdechloris* also shows indications of the expected formyl group (Gisriel et al., 2021; Kato et al., 2020).

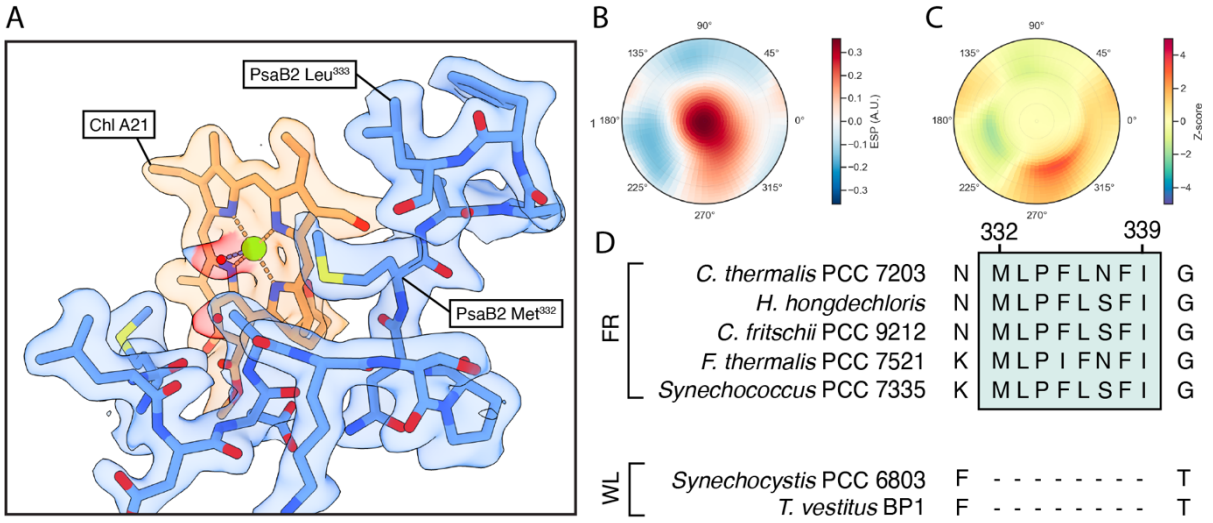


Fig. 4.8 – Chl *f* A21 site

A) Chemical environment of Chl *f* A21. The Chl density is in orange, the density of the amino acid residues from PsaA2 in blue, the density of water molecules in red. Hydrogen bonds are indicated as dotted lines. The C2 formyl group is within hydrogen-bonding distance to the peptide bond amide group of PsaA2 Leu^{333} and is likely to interact with PsaA2 Met^{332}. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent at an angle of ~270°. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent at an angle of ~270°. D) Multiple sequence alignment of FR and WL PsaA, showing the far-red specific loop insertion of 8 amino acids.

4.2.3 Chl f A23

Chl *f* A23 is identified along with its hydrogen bonding ligand, PsaA2 Gln^{365} (Fig. 4.9A, B, C). The same assignment has been made in *H. hongdechloris* based on structural and phylogenetic arguments (116,156). It was noted that among Chl *f*-producing cyanobacteria species, approximately half lack this Gln and instead have the non-hydrogen-bonding Met, leading to the assumption that this is a species-specific Chl *f* site (116,156). Other conserved changes are identified in the helix that coordinates A23 in the species that have Gln hydrogen bonded to the Chl *f*. These changes may be important for the insertion of a Chl *f* into this site (Fig. 4.9D). PsaA2 Thr^{368} coordinates the oxygen of PsaA2 Gln^{365} keeping the amide group in a position that favours the formation of the hydrogen bond with the C2 formyl oxygen. At the same time, two additional changes, PsaA2 Asn^{358} and PsaA2 His^{362}, move the water that coordinates the Mg^{2+} of Chl A23, shifting the chlorophyll to a position that optimises the interaction of the C2 formyl group with PsaA2 Gln^{365}.

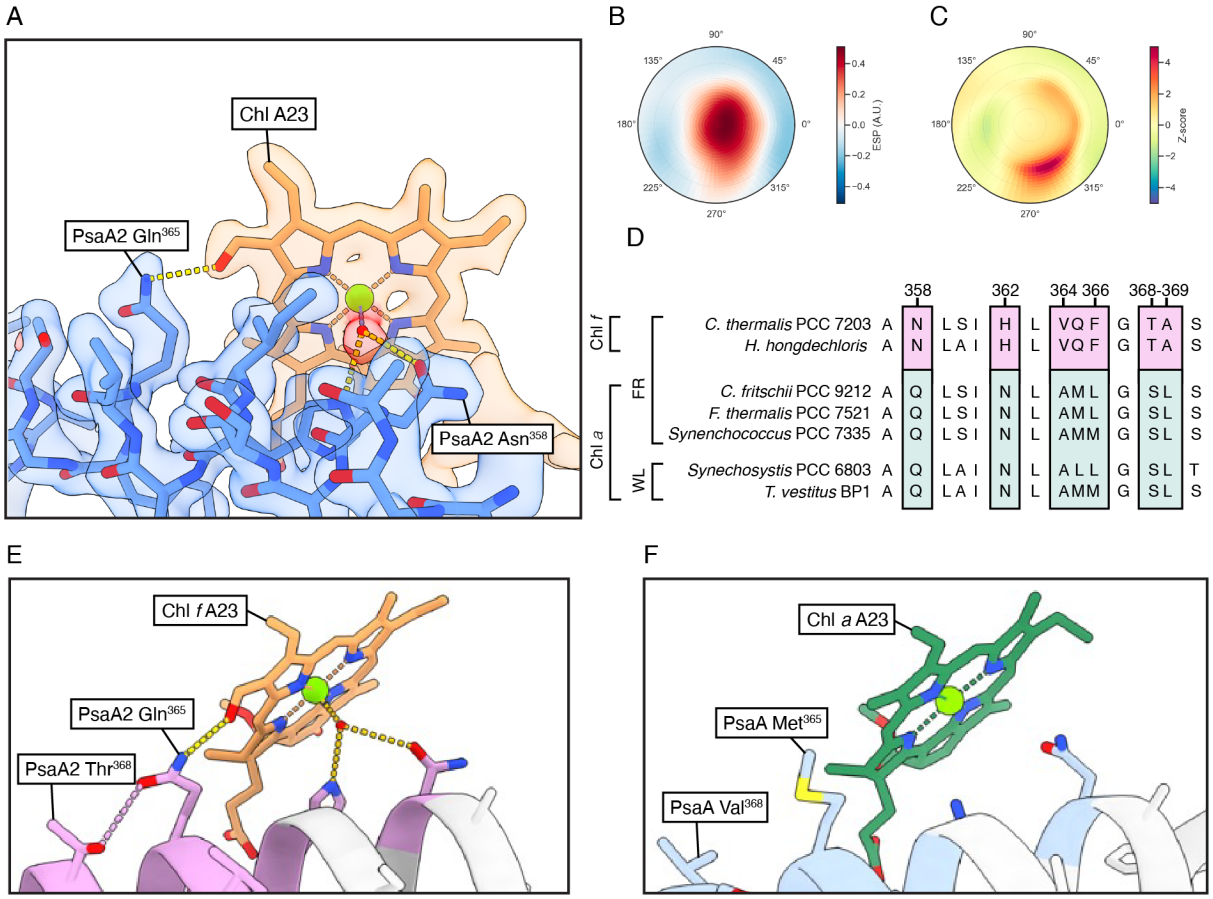


Fig. 4.9 – Chl *f* A23 site and variability among far-red strains

A) Chemical environment of Chl A23. The Chl density is in orange, the density of the amino acid residues in blue, the density of water molecules in red. Hydrogen bonds are indicated as yellow dotted lines. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. D) Multiple sequence alignment of FR and WL PsaA, showing the differences that allow the binding of a Chl *f* at the A23 site. In pink, changes associated with the presence of a Chl *f*, and in light blue the equivalent amino acids in WL-PSI and in those FR species that appear to lack Chl *f* in this site. E) Chl A23 site in presence of the mutation that allow the insertion of a Chl *f*. F) FR-PSI from *F. thermalis*, in which the changes are not present and presents a Chl *a* at the A23 site.

4.2.4 Chl f B07

The ESP map indicates an hydrogen bond from the backbone amide of PsaB2 Ala^{94} to the C2 formyl group of Chl *f* in the B07 site (Fig. 4.10A, B, C). PsaB2 Ala^{94} is present in most far-red species instead of a conserved proline in WL-PSI. The Mg^{2+} of the Chl *f* is coordinated by the amide of PsaB2 Gln^{95} (Fig. 4.10D). B07 was identified as a likely candidate for Chl *f* in earlier work, based on sequence-based modelling indicating a non-

His axial ligand (43) and Cryo-EM data (112,116,129,156). In *Synechococcus* sp. PCC 7335 (129), the presence of the WL-specific Pro suggests that the formyl group is oriented differently. Instead, a water molecule is instead present in the Cryo-EM structure of FR-PSI from this organism, accompanied by an increased ESP potential from the C2 formyl substituent.

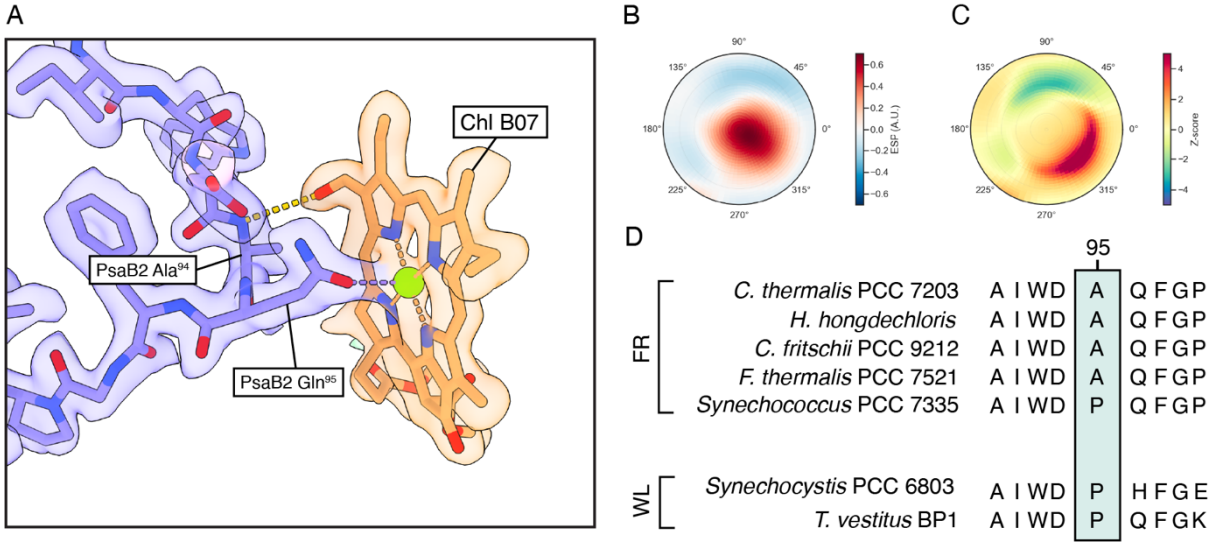


Fig. 4.10 – Chl *f* B07 site

A) ESP map and atomic model around the B07 site. The Chl density is in orange, the density of the amino acid residues in PsaB2 is in violet. Hydrogen bonds are indicated as dotted lines. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. D) Multiple sequence alignment of FR and WL PsaB, showing Pro to Ala substitution in position 95 that is specific to the far-red sequence, although not present in all.

4.2.5 Chl f B30

The ESP map shows an hydrogen bond to the formyl group of B30 from Tyr^{38}, a conserved change present in PsaJ2 (Fig. 4.11A, D). Local analysis of the ESP results in a clear single peak oriented towards the PsaJ2 Tyr^{38} side chain (Fig. 4.11B, C). Tyr^{38} is part of the C-terminal extension of PsaJ2 that displaces three Chl *a* molecules and a carotenoid from this region in the canonical WL structure (CLA 1301,1302,1303, and BCR 1306 in Jordan et al., 2001). B30 also shows an unusual ε-His ligand from PsaB2 His^{443}. Calculations suggest that ε-His coordinated chlorophylls have slightly lower site-energies than those with the more common δ-His ligand (176). Chl B30 was previously identified as a Chl *f*

candidate given the potential hydrogen bond from the conserved FR-specific PsaJ2 Tyr^{38} (112). When the cone scan was applied, indications for a distribution between two possible orientations were reported in *F. thermalis* (156), but this evidence was lacking in *H. hongdechloris* as the PsaJ2 subunit was absent from the structure (116). The lower resolution in *F. thermalis*, the two possible positions for the proposed hydrogen bond, and the anomalously long His-Mg^{2+} distance, all indicate that in the *F. thermalis* FR-PSI this site was perturbed. Neither the ESP map nor the cone scan for *Synechococcus* sp. PCC 7335 showed evidence for the formyl in B30 (112). This was interpreted as low occupancy of the site, perhaps due to loss of PsaJ2 as seen in *H. hongdechloris* in both FR- and WL-PSI. Here, in intact FR-PSI map with higher resolution, this site is fully occupied. Moreover, the Chl B30's 13^1-keto group is hydrogen bonded by the conserved far-red specific change in PsaF2 Tyr^{50}, which is a Thr in WL-PSI. Assuming that the site energy of Chl *f* molecules is influenced in the same way Chl *a* molecules are when this hydrogen bond is present, B30 can be expected to be among the more red-shifted Chl *f* sites (177).

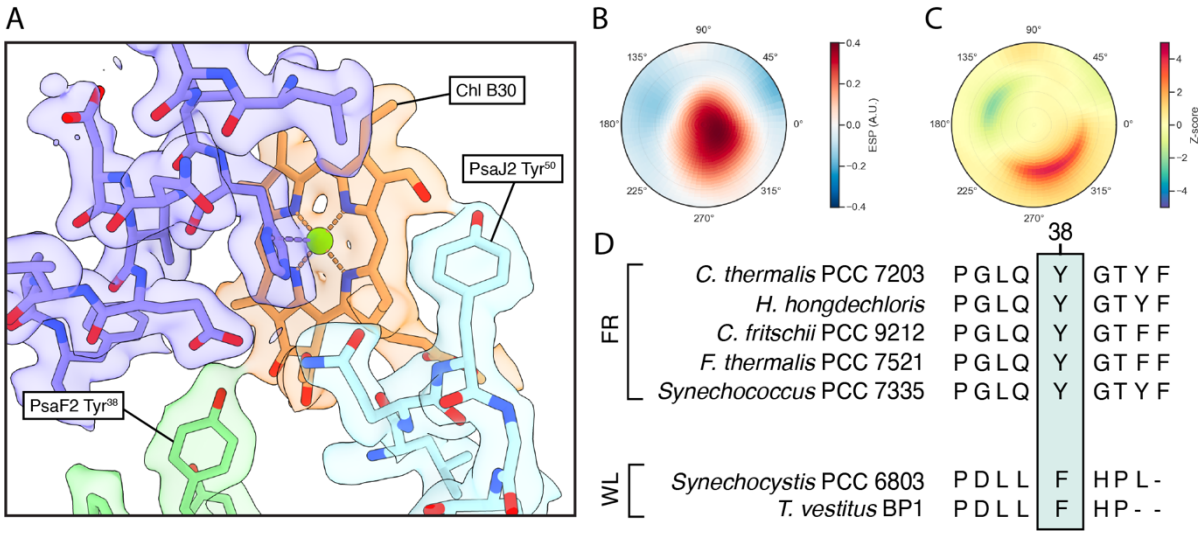


Fig. 4.11 – Chl *f* B30 site

(A) ESP map and atomic model around the Chl B30 site. The Chl density is in orange, the density of the amino acid residues from PsbB2 in violet and those from PsaJ2 in cyan. Hydrogen bonds are indicated as dotted lines. (B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. (C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. (D) Multiple sequence alignment of FR and WL PsaJ, showing the far-red specific PsaA2 Tyr^{38}.

4.2.6 Chl f B37

An hydrogen bond from the backbone amide of PsaB Gly^{697} to the formyl oxygen of B37 (Fig. 4.12A) is evident in the ESP polar plot of the C2 substituent (Fig. 4.12B, C). PsaB2 Gly^{697} is specific for FR-PSI, replacing either Val or Ile in PsaB of WL-PSI (Fig. 4.12D). The Mg^{2+} of B37 has water as its axial ligand. Based on these characteristics, B37 is confirmed as being a Chl *f*, in agreement with previous assignments (43,112,116,166). Moreover, another far-red-conserved change in the sequence of PsaB2 could be important for the function of Chl B37, or in the dimeric interaction with the Chl *f* in position B38. The 13^1-keto group of Chl *f* B37 is hydrogen bonded by a the far-red-conserved residue PsaA2 Asn^{452} (Ile in WL-PSI). Also in this case, the expected effect could be a red-shift of the absorption spectra of this chlorophyll (177), which is in accordance with the Chl *f* dimer B37/B38 extremely red-shifted absorption at 800 nm.

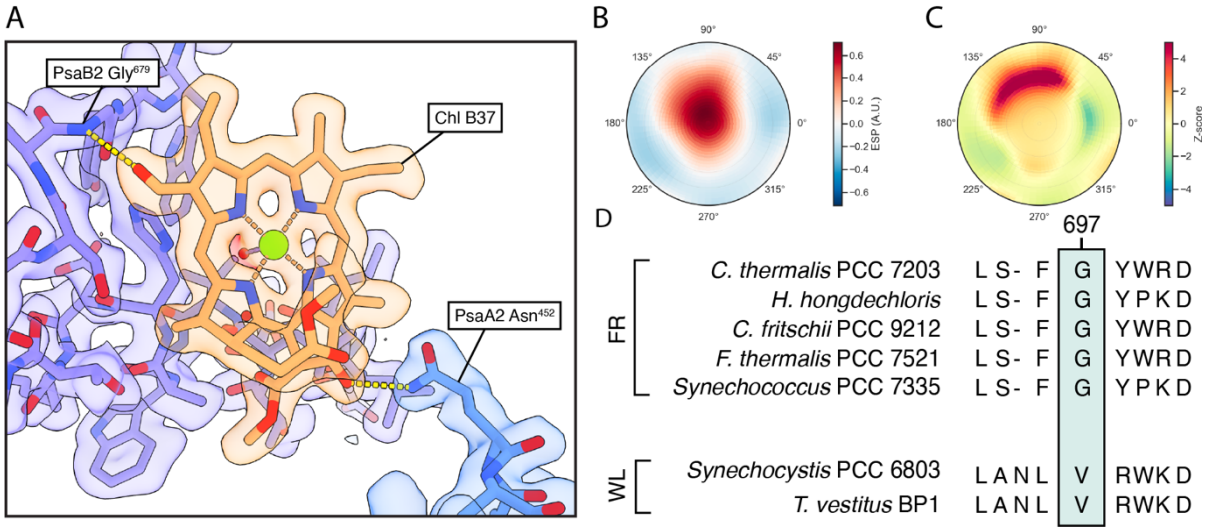


Fig. 4.12 – Chl *f* B37 site

(A) ESP map and atomic model around the B37 site. The Chl density is in orange, the density of the amino acid residues in violet and the density from water in red. Hydrogen bonds are indicated as dotted lines. (B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. (C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. (D) Multiple sequence alignment of FR and WL PsaB, showing the FR-specific Gly in position 697.

4.2.7 *Chl f B38*

Together with B37, B38 forms the low energy Chl *f* coupled pair absorbing at 800 nm . The central Mg^{2+} of Chl B38 is coordinated by a water that forms a distal hydrogen bond with PsaB2 Asp^{701}, but its formyl oxygen lacks a hydrogen bond partner (Fig. 4.13A, D). Nevertheless, a positive feature, with peak Z-score > 3 is present in the ESP polar plot of the C2 substituent. The torsion angle of the feature suggests that the formyl group is parallel to PsaB2 Trp^{22}, a residue that is also conserved in PsaB in WL-PSI. Compared to other sites, Chl B38 presents a lower intensity of the signal, reflecting the weak character of this interaction, and once again confirming that higher electronegativity and a broader spread of torsion angles can diminish the formyl oxygen's ESP signal. Tros et al. and Gisriel et al. (129,166) pointed to a conserved change in some FR-PSI sequences, where PsaB Trp^{22} is replaced with a Phe, which, being smaller, allows the insertion of a water molecule to bridge the formyl oxygen of Chl B38 to the carbonyl backbone of Thr^{18}. Nevertheless, this change is not present in *C. thermalis* PCC 7203, and there is no ESP density indicating the presence of a water molecule.

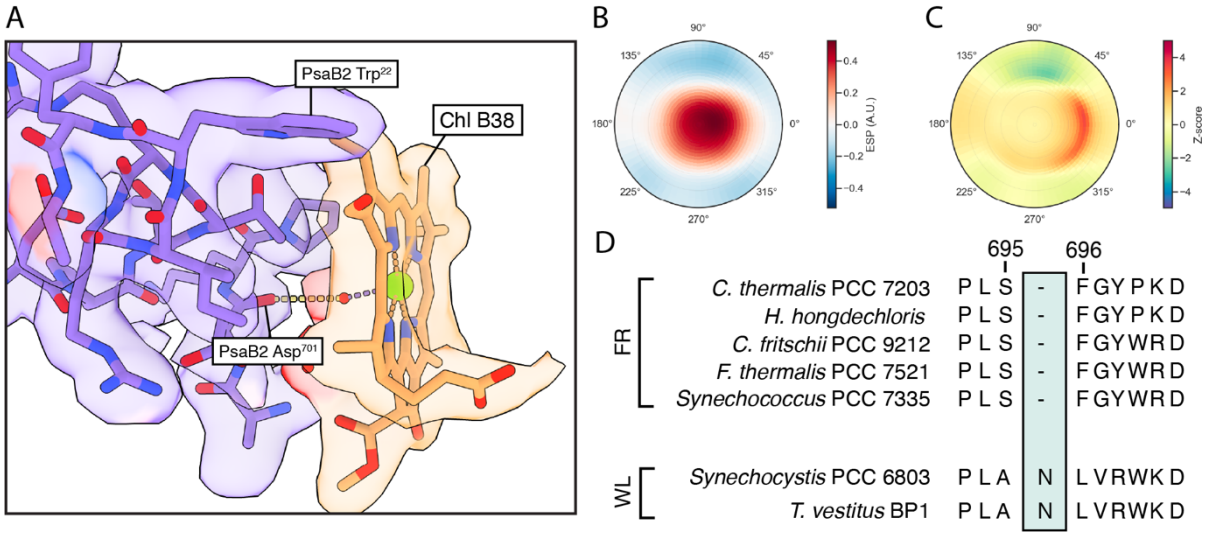


Fig. 4.13 – Chl *f* B38 site

A) ESP map and atomic model around Chl B38 site. The Chl density is in orange, the density of the amino acid residues in violet and the density from water in red. Hydrogen bonds are indicated as dotted lines. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. D) Multiple sequence alignment of FR and WL PsaB.

4.2.8 Chl f L03

The ESP polar plots of Chl L03 show a clear density for a formyl group in the C2 position that does not appear to be hydrogen bonded. The chemical environment of the site is similar in both WL-PSI and FR-PSI in that it contains a water molecule coordinating the central Mg^{2+} atom of the chlorophyll; however, it also exhibits a far-red-specific conserved amino acid change, with PsaL2 Val^{81} replacing the Lys/Leu/Ile of PsaL, generating the space necessary for the formyl group. The increased ESP density is present at a polar angle that suggests a possible interaction with a diglyceride present both in WL- and FR-PSI (Fig. 4.14A, B, C).

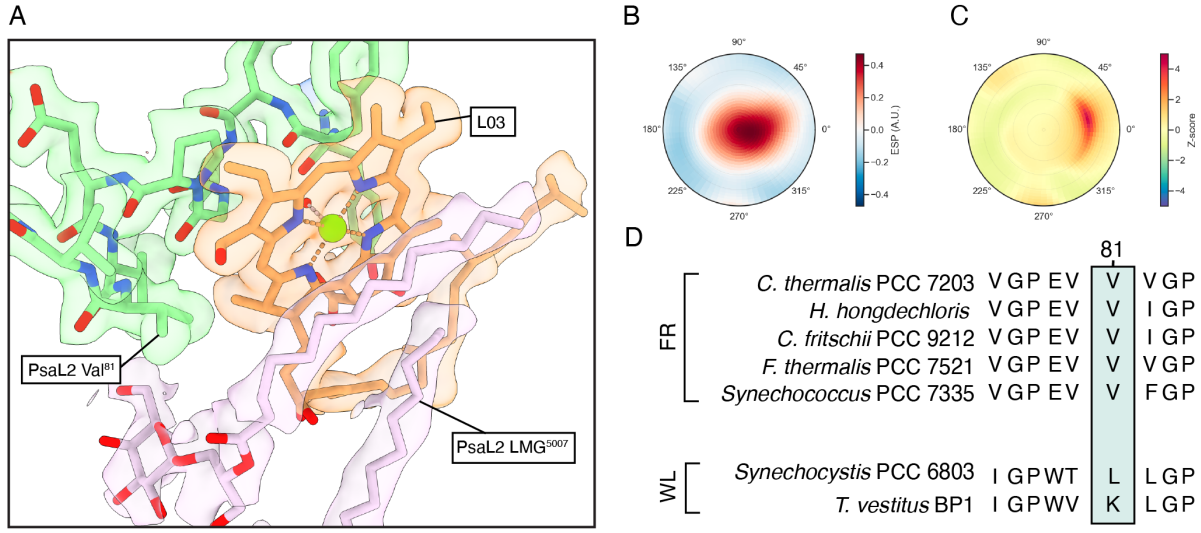


Fig. 4.14 – Chl *f* L03 site

A) Chemical environment of Chl L03. The Chl density is in orange, the density of the amino acid residues from PsaL2 in green, the density of the diglyceride in pink, and the density from water in red. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. D) Multiple sequence alignment of FR and WL PsaL.

4.3 Discussion

4.3.1 Comparison with the structures in literature

A survey of high-resolution structures of FR-PSI and WL-PSI in the literature shows that the negative ESP feature at the C2 substituent of Chl *f* A_{-1B} is present in a similar location in other FR-PSI structures, i.e., *Halomicronema hongdechloris* (116), *Synechococcus* sp. PCC 7335 (129), but absent from the equivalent site in WL-PSI maps of comparable resolution (Fig. 4.15A).

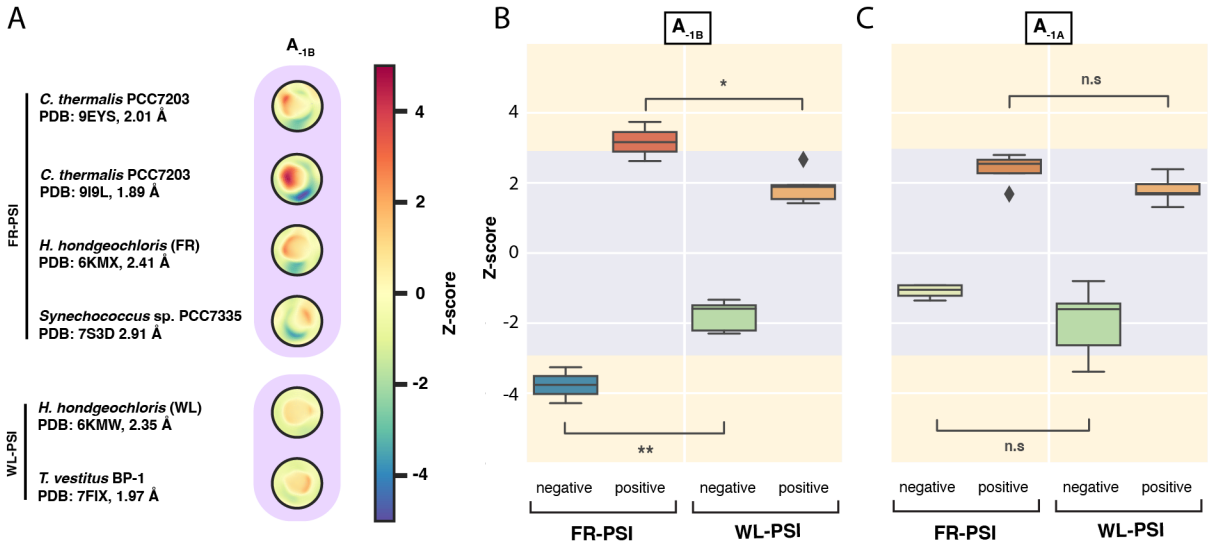


Fig. 4.15 – ESP at the C2 position for the A_{-1B} chlorophylls of FR-PSI and WL-PSI

A) Comparison of plots of ESP of the C2 substituent on the A_{-1B} chlorophylls calculated using the high-resolution structures for FR-PSI and WL-PSI available from the PDB. The plots show the Z-scores coloured according to the scale bar, with statistically significant positive and negative regions of ESP present only in FR-PSI. B) Box plot displaying the mean values and standard deviation of maximum and minimum Z-scores across different structures of FR-PSI and WL-PSI for the C2 substituent of the A_{-1B} site. C) Box plot displaying the mean values and standard deviation of maximum and minimum Z-scores across different structures of FR-PSI and WL-PSI for the C2 substituent of the A_{-1A} site. The yellowed portion highlight Z-score values that above 3 and below -3. Asterisks indicate the statistical significance of differences between the column pairs, as determined by a two-sample t-test: * indicates $p < 0.05$, ** indicates $p < 0.01$.

The maximum and minimum Z-scores in the cone scans from the C2 substituent of the A_{-1B} and A_{-1A} reaction centre Chl, of maps available in literature for WL-PSI and FR-PSI with comparable resolution were analysed. A clear statistical significance is present in for the maximum and minimum ESP Z-score of the Chl A_{-1B} site, confirming the assignment of a

Chl *f* in this location and the robustness of the method towards small structural changes in the PDB. The same analysis performed on the A_{-1A} site, shows no significant difference between WL-PSI and FR-PSI, providing further evidence of the presence of a Chl *a* in this location.

4.3.2 DFT calculation of Chl A_{-1B} chemical environment

Among the evidence presented in favour of the presence of A_{-1B} in the reaction centre, a set of DFT calculations was performed in collaboration with Nick Rew, Daniel Medranda and Prof. Jenny Nelson from the Department of Physics of Imperial College London, to identify and distinguish the origins of the electrostatic effects contributing to the increased negative ESP features observed around the C2 formyl of Chl *f* in the A_{-1B} site.

The calculations were performed on the A_{-1B} and A_{0B} chlorophylls and the surrounding protein environment using the structural model with and without an additional optimisation step. In the following, the qualitative results apply to both optimised and non-optimised structures. Density Functional Theory (DFT) calculations of the electrostatic environment of A_{-1B} and its neighbouring amino acids are shown as isosurface contours of the ESP (Fig. 4.16). A negative region (in blue) is present between the formyl group and the benzyl ring of PsaA2 Phe^{459} when Chl *f* occupies the A_{-1B} site. This negative region appears to correspond to the negative ESP feature detected at a polar angle of 270° in the ESP and Z-score polar plots (Fig. 4.16B, C)

With no hydrogen bond to the formyl group, the negatively polarised formyl oxygen remains un-neutralised, exhibiting a negative well of ESP (Fig. 4.16D). This feature is present irrespective of the presence of PsaA2 Phe^{459}/Cys^{463}. When PsaA2 Phe^{459}/Cys^{463} is present, the negative ESP well is spatially confined to the region between the Chl *f* and the PsaA2 Phe^{459} (Fig. 4.16B), and the depth of the well is estimated to decrease from -0.0667 to -0.0432 atomic units (~0.640 eV). With Chl *a* in the far-red A_{-1B} site there is no negative well between the C2 methyl group of Chl *a* and the PsaA2 Phe^{459} (Fig. 4.16A). Thus, the presence of the negative ESP feature in the Cryo-EM is additional evidence of Chl *f* being in the A_{-1B} position.

The presence of a formyl group at the C2 site of Chl *f* leads to a redistribution of ESP along the y-axis of the Chl A_{-1B} (Fig. 4.16E), and this is directed by the interaction of the C2 formyl bond and the π-electrons of PsaA2 Phe^{459}.

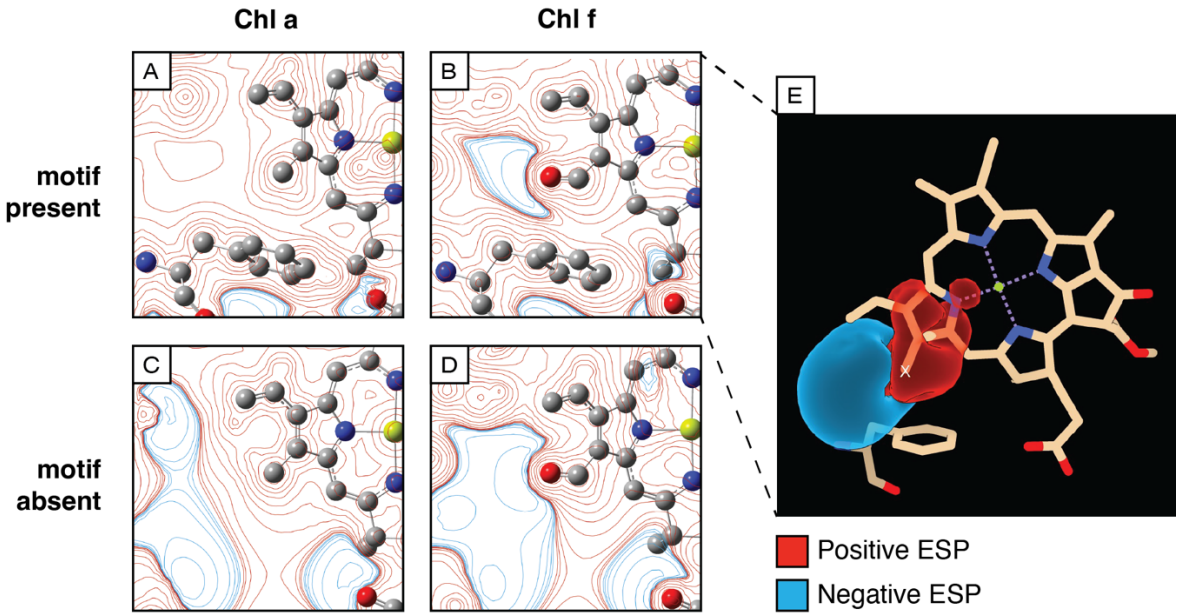


Fig. 4.16 – DFT-calculated ESP of Chl *a* and Chl *f* in the A_{-1B} site

DFT-calculated ESP contours in a plane defined by the formyl group (C=O) at the C2 position of Chl A_{-1B} and the γ-carbon of PsaA Phe^{459}. Contours are shown for ESP values, red indicates positive ESP, whilst blue indicates negative ESP. A) Chl *a* A_{-1B} with PsaA2 Phe^{459}/Cys^{463} present. B) Chl *f* A_{-1B} with PsaA2 Phe^{459}/Cys^{463} present. C) Chl *a* A_{-1B} in absence of PsaA2 Phe^{459}/Cys^{463} absent. D) Chl *f* A_{-1B} in absence of PsaA2 Phe^{459}/Cys^{463} absent. Chl *a* A_{0B} and the other amino acids adjacent to the C2 carbon were included in the calculation but are not shown in the figure. E) Representation of Chl *f* A_{-1B} with PsaA2 Phe^{459}/Cys^{463} present as isosurface contours.

The presence of the negative ESP feature in the region of the C2 formyl oxygen suggests that the sulfhydryl-π-formyl motif has functional and evolutionary significance in tuning the binding and the redox potential of the Chl *f* in the A_{-1B} site. Chl $f^{-\bullet}$ and Chl f^{*} are less reducing than Chl $a^{-\bullet}$ Chl a^{*} and so their redox potentials must be lowered if Chl *f* is to play redox roles in PSI like those of Chl *a* (178). These chemical requirements provide insight on the absence of a conserved hydrogen-bonding group to increase the affinity of this site for Chl *f*. This also explains why the positive ESP signal arising from the C2 formyl oxygen of the Chl *f* in the A_{-1B} site is weaker in comparison with other Chl f formyl group in this the structure.

4.3.3 Chl f in the A_{-1B} position reproduces the experimental difference spectra

As a separate line of evidence for the presence of a Chl *f* in position A_{-1B} in FR-PSI, we collaborated with Michael Hofer and Prof. Thomas Renger from the Institute for Theoretical Physics, Johannes Kepler University, Linz, Austria, to calculate the $P700^+$ *minus* P700 difference spectrum of FR-PSI, considering the presence of Chl *f* in different sites.

In the calculated spectra, the ground and excited states of the Chl *a* and Chl *f* were represented by atomic partial charges obtained from a fit of the electrostatic potential of their respective charge densities, which were obtained with time-dependent density functional theory. Compared to previous literature, accurately describing the electrochromic shifts required a higher level of theory compared to the point-dipole approximation, as well as consideration of the specific physicochemical properties of Chl *a* and Chl *f*, was required. The $P700^+$ induced shifts on either Chl *a* or Chl *f* in the A_{-1B} site were found to be 9 cm^{-1} and -50 cm^{-1} respectively. This sign difference between the two types of chlorophyll was not previously predicted (43,179).

The experimental $P700^+$ *minus* P700 difference spectrum is almost quantitatively reproduced when considering the Chl *f* positions assigned in the Cryo-EM map presented here, including A_{-1B} assigned with a site energy at 747 nm (Fig. 4.17A). On the other hand, when excluding the presence of a Chl *f* in the A_{-1B} position, the bleaching corresponding to this site in the calculated spectrum around 750 nm disappears, demonstrating that this Chl *f* is responsible for this band in the experimental spectrum (Fig. 4.17B). The small negative peak at around 760 nm arising from the Chl f antenna pigments is still present, suggesting that this feature arises from one of the antenna Chl *f* molecules. When ignoring coupling effects, the experimental spectrum cannot be reproduced in the 650 nm to 720 nm region, where the changes are attributable to effects on Chl *a* (Fig. 4.17C).

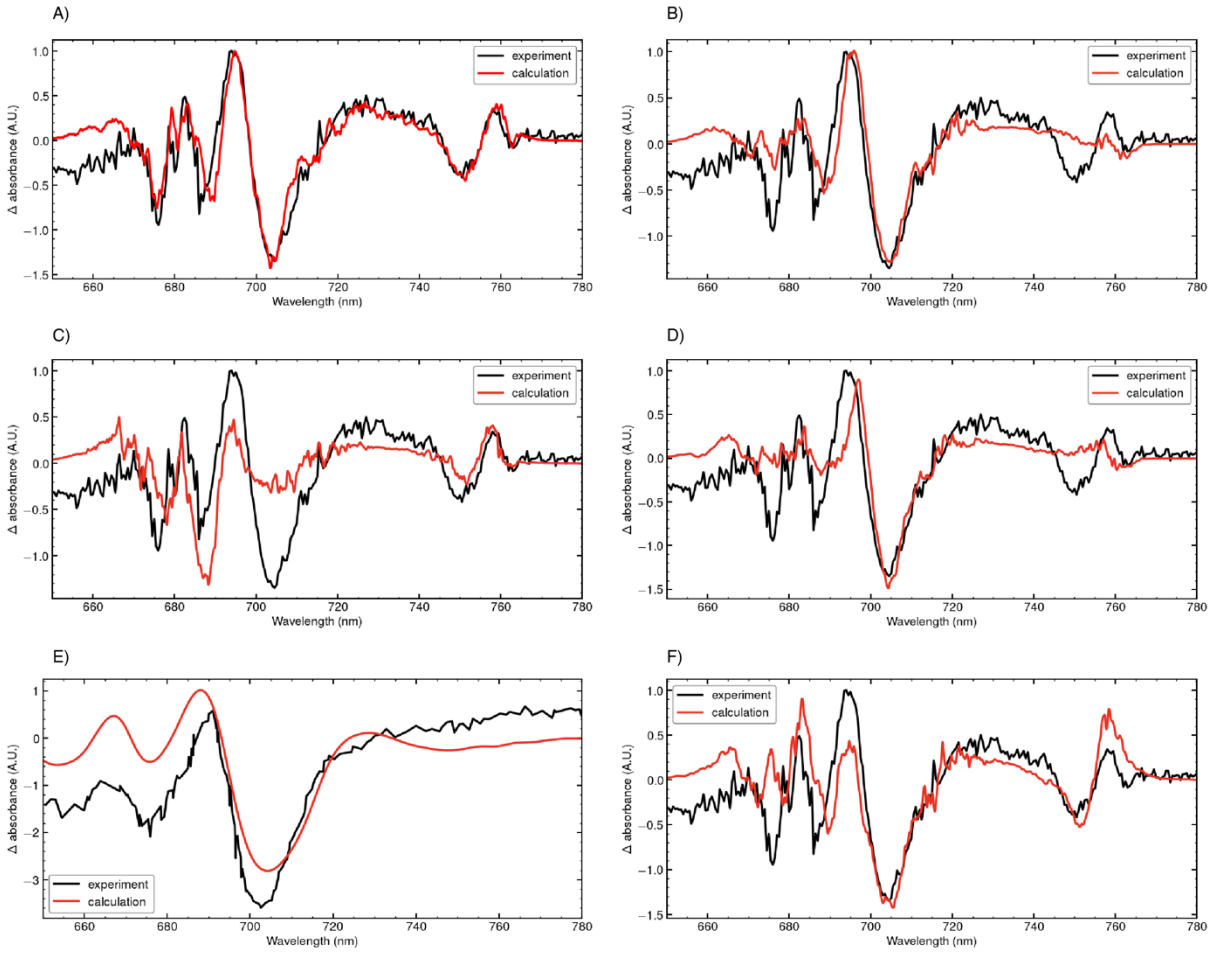


Fig. 4.17 – Simulations of the P700$^+$ *minus* P700 difference spectra

A) Difference spectra considering the Chl *f* positions presented in the chapter. B) Difference spectra considering only the antenna sites presented in the chapter. C) Difference spectra excluding the contributions due to coupling. D) Difference spectra as in panel B but considering the presence of a Chl *f* in the antenna site with the highest electrochromism (B24) E) difference spectra as in panel A but calculated at room temperature F) Difference spectra considering the presence of a Chl *f* in A_{-1A} rather than in A_{-1B}.

As a further proof of concept, the P700$^+$ *minus* P700 difference spectrum was also calculated with Chl *f* only in antenna sites, but assuming Chl B24, the antenna pigment with the largest electrochromic shift, as a Chl *f* (Fig. 4.17D). In this case, a small peak is visible in the Chl *f* region now, the magnitude is too small to explain the experimental spectrum. As B24 is now lacking in the Chl *a* spectral region, the fit is worse in this region too, since B24 is excitonically coupled to the RC pigments, in particular to P_A and P_B. The only possible way to obtain an electrochromic shift in the antenna that is large enough to explain the lineshape around 750 nm would be to consider all of the pigments that present the strongest electrochromism in the antenna (i.e., B39, A40, B24) as Chl *f* with exactly the same site energy. This assumption is inconsistent with the present Cryo-EM model. Previous claims

that there is no Chl *f* in the reaction centre of FR-PSI can be ruled out based on the theoretical modelling of the spectroscopic data.

Moreover, the simulation explains why previous reports described a weak (43) or absent (116) Chl *f* bandshift in room-temperature difference spectra; this observation can be explained by the temperature dependence of the dielectric constant. Whereas the high-energy Chl *a* side of the difference spectrum is influenced both by the excitonic couplings and the electrochromic shifts in site energies, the low-energy Chl *f* region is dominated by electrochromic effects because Chl *f* at A_{-1B} is off-resonant to the other reaction centre pigments, which are Chl *a* molecules (Fig. 4.17). The $P700^+$ *minus* P700 difference spectrum measured at room temperature showed only weak features (Nürnberg et al 2018) or no sign (116) of bandshifts attributable to Chl *f*. This is because at room temperature, the effective dielectric constant (ε_{eff}) needs to be increased to 8, compared to the value 2-3 at $T < 170$ K. The decrease of ε_{eff} at low-temperature is due to the freezing out of many conformational degrees of freedom.

Finally, since A_{-1A} was also originally proposed as a candidate Chl *f* in the reaction centre and is the site with the strongest electrochromic shift in the reaction centre outside the special pair, the simulated $P700^+$ *minus* P700 difference spectrum with Chl *f* at position A_{-1A} instead of A_{-1B} (Fig. 4.17F) was also calculated. The experimental data are only semi-quantitatively described, but the fit is not as good as that obtained for Chl *f* at A_{-1B} (Fig. 4.17A).

Overall, the set of calculations presented, which involves a higher level of theory compared to previous literature, demonstrates that the only pigments that can generate a bandshift of a magnitude equal to the experimental data are the ones in the reaction centre. Specifically, the fit of the experimental data agrees with the high-resolution structures presented in this chapter, pointing towards A_{-1B} as the Chl *f* in the reaction centre.

4.3.4 Implications and tuning of a Chl f as a primary donor

The ESP maps presented in this chapter show an increase in the electrostatic potential for the C2 substituent of A_{-1B} (Fig. 4.5A), the presence of conserved changes in proximity of the site (Fig. 4.5D), the DFT calculations of the C2 environment (Fig. 4.16) and the $P700^+$ *minus* P700 difference spectra simulations (Fig. 4.17), all provide additional evidence that Chl *f* occupies the A_{-1B} site, which adds to the one present in literature (43).

The intrinsically less reducing excited state of Chl *f* compared to that of Chl *a,* makes it a less suitable pigment to perform effectively the role of primary donor in FR-PSI. This behaviour would be exacerbated by the presence of a hydrogen bond to the C2 formyl group, explaining why phylogenetically there are no changes that provide a stabilizing interaction to the C2 of A_{-1B}. The presence of the Phe/Cys pair, on the other hand, provides an area of more negative ESP around the formyl group of A_{-1B} (Fig. 4.5B, C) which is expected to destabilise the LUMO of the Chl *f* and make its redox potential more negative (180). This may tune the redox potential of Chl *f* in A_{-1B}, thereby restoring the reducing power needed for the PSI primary donor (178,180). The proposed charge separation route in FR-PSI would involve the charge separation event originating on A_{-1B}, followed by the donation of an electron to A_{0B} while the hole moves to Chl *a* dimer P_A P_B.

In FR-PSI, excitation with far-red light will result in charge separation preferentially on the B-side of the reaction centre, with a more marked effect at low temperatures. When excited by visible light, the Chl *a* antenna will mainly pass the excitation onto the Chl *f* pigments, as seen in ultrafast kinetic studies (118,181). Visible light may also excite redox-active Chl *a* molecules (A_{-1A}, P_A and P_B), especially when they are targeted directly, and these species are potentially capable of charge separation on the A-side (43,118,181) if primary electron transfer can compete with energy transfer to the low energy Chl *f* at A_{-1B}. This predicted complex mixture of kinetic pathways may be unravelled by appropriate use of the excitation wavelength selectivity.

This also provides insight on the debated topic of the origin of charge separation in WL-PSI. Due to the extremely short time regimes in which it takes place (25 - 70 ps), the sequence of events that lead to the charge separated state in PSI is not fully understood.

The first hypothesis on the nature of the primary electron donor and charge separation in PSI is that, upon excitation, the P700 pair is the primary electron donor (182–184). The second, is that the two A_{-1} Chl molecules are the primary electron donor, generating an initial charge separated state with the pair of A_0 Chl molecules (A_{-1}^{+}/A_0^{-}), to then be regenerated by an electron being transferred from P700 to the oxidised A_{-1}^{+} Chl molecule (43,174,181,185,186).

The confirmed presence of a Chl *f* in this position provides a strong point in favour of the second hypothesis (i.e., Chl A_{-1} are the primary donors in WL-PSI). From a phylogenetic point of view, mutations that provide an evolutionary advantage by allowing the insertion of a longer wavelength pigment will have a higher probability of being localised in proximity of the primary donor. This does not exclude the possibility that the nature of the charge separation of WL-PSI and FR-PSI is different, and that the primary donor of WL-PSI might be P700.

The presence of a Chl *f* in the reaction centre does not exclude the possibility that changes in the coupling between reaction centre pigments, and the consequent appearance of a long-wavelength charge transfer absorption band, helps drive far-red light charge separation. In fact, even if no changes are seen in the vicinity of the A-side quinone, A_{1A}, a specific far-red conserved change around the lower potential quinone A_{1B} is present: the substitution of PsaB2 Ser^{673} with Thr^{673} (187). The structure around the quinone appears to be unaffected by this change. However, PsaB2 Thr^{673} is involved in a steric interaction with A_{0B}, and this could limit any potential dynamic changes associated with quinone reduction. As pointed out recently, the PsaB2 Ser to Thr^{673} change, together with two other conserved amino acid changes (PsaB Val/Ile^{666} and Ser^{667}), contributes to a 5° rotation of Chl A_{0B}, changing its aromatic overlap with A_{-1B} and thus potentially modifying their coupling (187).

4.3.5 *PsaL2 controls trimerization and L03 is the sharing pigment*

FR-PSI exists as trimers *in vivo*, whereas WL-PSI is found as dimers or tetramers in *C. thermalis* (188). Most of the antenna Chl *f* molecules are located on the B-side, closer to A_{-1B}, and this reflects the greater number of conserved amino acid differences in PsaB2 compared to PsaA2 (43). FR-PSI trimerization positions this side close to the interface between individual PSI monomers, resulting in a core enriched in Chl *f* (Fig. 4.18).

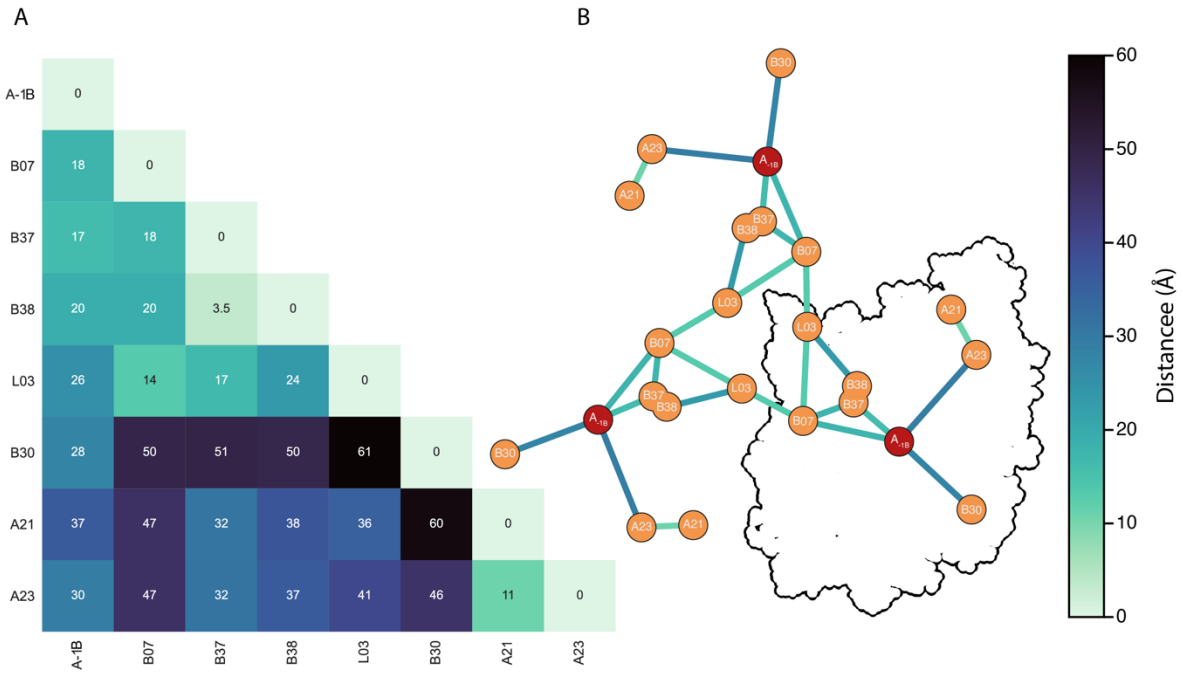


Fig. 4.18 – Excitation energy transfer routes in FR-PSI

A) Chl *f* edge-to-edge distance matrix colour coded according to the colorbar on the right. B) Shortest distances among Chl *f* sites showing the potential exciton routes towards the reaction centre and between the monomers in FR-PSI trimer colour coded according to the colorbar on the right. FR-PSI monomer outline is shown in black, antenna Chl *f* are in orange and reaction centre A_{-1B} site in red.

PsaL2 mediates trimer formation and Chl *f* L03 is located between Chl *f* B07 of the same complex and that of the neighbouring one. Therefore, Chl *f* L03 is likely to play an important role in low energy excitation energy connectivity in FR-PSI trimers. Within this core, the distances between long-wavelength pigments could allow efficient excitation energy transfer between Chl *f* molecules and between FR-PSI monomers in the trimer (Fig. 4.4, Fig. 4.18). Excitation energy sharing between individual PSI complexes in PSI trimers has been suggested to occur in WL-PSI via specific routes, one of which involves the chlorophylls of PsaL (175). Even though FR-PSI, like its white light counterpart, acts

as an efficient trap and performs rapid charge separation, the trimeric arrangement offers distinct advantages. Excitation energy sharing among Chl *f* molecules within the trimer significantly increases the chances of charge separation, especially when one or both other monomers are unable to perform charge separation due to damage or a lack of electron donors and acceptors. This network of long-wavelength pigments, extending to adjacent monomers, effectively expands both the number of long-wavelength pigments and the physical antenna size for each reaction center. Beyond this aspect, the trimeric architecture is also essential to complete and stabilize of some of the Chl *f* pigments (i.e. L03, B37 and B38). Furthermore, the trimeric form of PSI canonically presents a more "compact" form, which is critical for packing more photosystems in conditions of scarcity of photons. Finally, the trimeric assembly could be important for the formation of supercomplexes with FR-PSII and FR-APC, and to the general mesoscale organization of the thylakoid membrane in far-red light conditions.

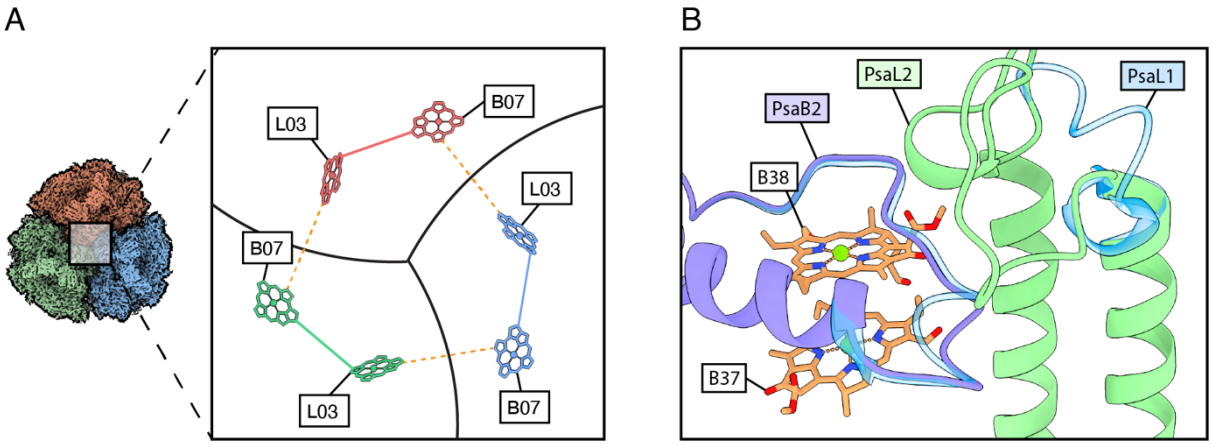


Fig. 4.19 – PsaL2 and PsaB2 prevent the formation of chimeric FR/WL-PSI

A) Trimeric FR-PSI from the cytoplasmic side, each monomer coloured in red, green and blue. In the same colour scheme, the chlorophylls closer to the connection interface (i.e., B07 and L03). Solid lines represent intra-monomeric distances (11 Å) and yellow dotted lines represent intra-monomeric distances (16 Å). B) Protein environment of the B37/B38 Chl *f* dimer and PsaL2/PsaB2 interface. PsaB2 is represented in violet, PsaL2 in green, Chl *f* molecules in orange and the WL version of the two subunits in blue.

The importance of inter-trimer Chl *f* connectivity is also highlighted by another structural feature: the far-red specific B37/B38 loop of PsaB2, which has extensive interactions with another far-red specific loop of PsaL2 (Fig. 3F). These interactions could stabilise the Chl *f* dimer B37/B38, favour the association of PsaB2-PsaL2, and prevent the formation of a

chimeric WL-FR-PSI by obstructing the interactions between PsaL1 and PsaB2. The formation of WL-FR-PSI complexes would hinder their ability to efficiently use far-red light to perform charge separation

4.3.6 Partial occupancy in far-red light photosystems

One of the debated aspects of Chl *f* containing photosystems is whether Chl *f* is present with partial occupancy in some of the sites (155). This behaviour is considered to be present in Chl *b* containing antenna from higher plants, in which some of the sites are more promiscuous, and can bind either Chl *a* or Chl *b*, making their assignment both in crystallography and Cryo-EM more complicated. In the same way, it has been argued that in FR-PSI and FR-PSII, it is possible that some of the Chl sites are capable of binding both Chl *a* and Chl *f* and therefore will contain one or the other in sub-stoichiometric amounts.

This view has been reinforced by several observations, including: first, the capacity of Chl *f* to bind in a small proportion to Chl sites in WL-PSI, demonstrated in *Synechococcus sp.* PCC 7002 capable of synthesising Chl *f* (189); second, the absence of a clear assignment of all expected Chl *f* sites in earlier structures of FR-PSI; and third, contradictory reports on the number of Chl *f* pigments in bulk pigments analyses, which range from 6 to 8 per monomer.

In the work from Kurashov and collaborators (189), it is demonstrated that WL-PSI complexes from *Synechococcus sp.* PCC 7002 are capable of incorporating a small amount of Chl *f* molecules (~3.8) per monomer. In this organism, WL-PSI incorporates naturally 1 to 4 Chl *f* pigments upon the expression of the Chl *f* synthase gene (*ChlF*), acquiring the capability to harvest far-red light and perform some degree of charge separation.

The fact that the “Chl *f* doped” WL-PSI is capable of using far-red light excitation is to be expected, since the heat from the environment can provide the thermal energy needed for relatively efficient uphill excitation energy transfer from Chl *f* to Chl *a*. Nonetheless, the presence of clear fluorescence emission peaks in these samples provides evidence that, even in an organism that possesses only the machinery to traffic Chl *a* and no specific adaptation for any of the chlorophyll sites, there is a preference for the insertion of Chl *f*

in specific sites compared to others. Upon FaRLiP, the absence of WL specific subunits, together with the (still uncharacterised) machinery used to transport Chl *f*, is likely to ensure the insertion of Chl *f* molecules only in specialised functional sites in far-red-adapted photosystems. Moreover, low-temperature absorption spectra from FR-PSI present clear, reproducible peaks, which change between different strains but are consistent across different preparations of FR-PSI from the same cyanobacteria. This behaviour indicates that the position of Chl *f* molecules in FR-PSI is conserved across different culture conditions and growth stages.

The contradictory reports present in the literature on the number of Chl *f* contained in FR-PSI can be attributed to two main reasons: i) the natural variation across species in the number of Chl *f* sites present in FR-PSI (i.e., the species-specific Chl *f* site A23), and ii) the FR-PSI purification quality, methodology, and intactness. The presence of contaminants in FR-PSI samples can bias the results by either increasing or decreasing the Chl content. On one hand, the presence of FR-PSII, which present a higher Chl *f* / Chl *a* ratio (11.4%) compared to FR-PSI (8.1% considering 8 Chl *f* molecules, 7.6% when considering 7 Chl *f* per monomer), can increase the amount estimated. On the other hand, even smaller contaminations from WL photosystems can impact the estimation. The intactness of the purified complexes also influences the number of Chl *f* sites present in FR-PSI. Particularly interesting is the case of Chl *f* B30, which is lost upon the detachment of the PsaJ2 subunit, due to its formyl's group being coordinated by a far-red specific Tyr of PsaJ2 and the relatively weaker axial coordination of its Mg^{2+} by a PsaB2 ε-histidine. Among the reported Cryo-EM structures, FR-PSI from *H. hongdechloris* (116) presents the loss of PsaJ and consequently of Chl *f* B30.

The difference between Chl *b* and Chl *f* containing complexes is also on a functional level. In Chl *b* containing antenna systems, the presence and position of promiscuous sites does not impact excitation energy transfer negatively or hinder the efficiency of the complex. Chl *b* containing sites are more blueshifted compared to the bulk of other pigments and will quickly transfer excitation energy to nearby Chl *a* molecules. Conversely, in Chl *f* containing photosystems, excitation energy routes depend entirely on the positions of the long-wavelength pigments, and even though efficient uphill energy transfer from Chl *f* to

Chl *a* has been observed, Chl *f* sites will rapidly trap excitation energy, potentially resulting in its dissipation and consequent losses in efficiency.

In photosystems and antenna proteins, different functional adaptations present different "architectural" requirements for the locations and functions of the pigments, making Chl *b* containing antennae and Chl *f* containing complexes not suitable for structural and functional comparison.

4.3.7 Spectral assignment of Chl f

With an assignment for all of the 8 Chl *f* sites in FR-PSI, the next crucial step in their characterization is to assign their wavelength and therefore their function. Even though there is currently no possibility to perform site directed mutagenesis to test these hypotheses, some of the absorption peaks can be assigned based on structural considerations and comparison among the available structures of different FR-PSI complexes.

The chlorophyll in A23 is assigned here as an hydrogen-bonded Chl *f*, in agreement with earlier work (116,156) (Fig. 4.9). The conserved amino acid changes in FR-PSI around this site in *C. thermalis* and *H. hongdechloris* are present in a sub-class of FR-PSI found in approximately half of FR-species, while the other half lack these changes and have a Met instead of a Gln in PsaA2 position 365 (Fig. 4.9). In the structures of *F. thermalis* PCC 7521 and *Synechococcus sp.* PCC 7335, where PsaA2 Met^{365} is present, this chlorophyll was assigned as a Chl *a* based on sequence analysis and ESP map (129). A comparison of the 77 K absorption spectra in the literature from the two types of FR-PSI shows several differences in the 740 - 760 nm region (43,116,166) corresponding to changes in contributions of some Chl *f* molecules to the absorption spectra (i.e., excitonic coupling, reorganization energy) and the loss of a Chl *f* in the A23 site. Given these changes and the differential contribution from allophycocyanin in the different spectra, there is considerable uncertainty regarding the assignment of the absorption peak of A23, but it seems consistent with an absorption around 750 - 755 nm and an underlying absorption from A_{-1B} Chl *f* at 755 - 760 nm (Fig. 4.17, Fig. 4.20).

Other three wavelength assignments can be suggested based on the 77 K fluorescence spectra of FR-PSI. The spectra present two terminal emitters, one at ~750 nm and the other at ~800 nm, the exact wavelengths depending on the species (43,166). The two terminal emitters are expected to be separated from each other, both spatially and by an energy barrier, such as by several Chl *a* molecules. Tros et al. (190) showed that while the emitter at ~750 nm was insensitive to the presence of $P700^+$, the emitter at ~800 nm was quenched by $P700^+$. The emission at ~800 nm originates from a coupled pair of Chl *f* molecules, B37 and B38 (43,166) with a distance from P_B^+ of 26 Å. In this pair, the inter-pigment distance is below 4 Å, where short-range contributions to site energy shifts are expected to be significant (191). The quenching of the ~800 nm emitter by $P700^+$ seems likely to be protective in excess light.

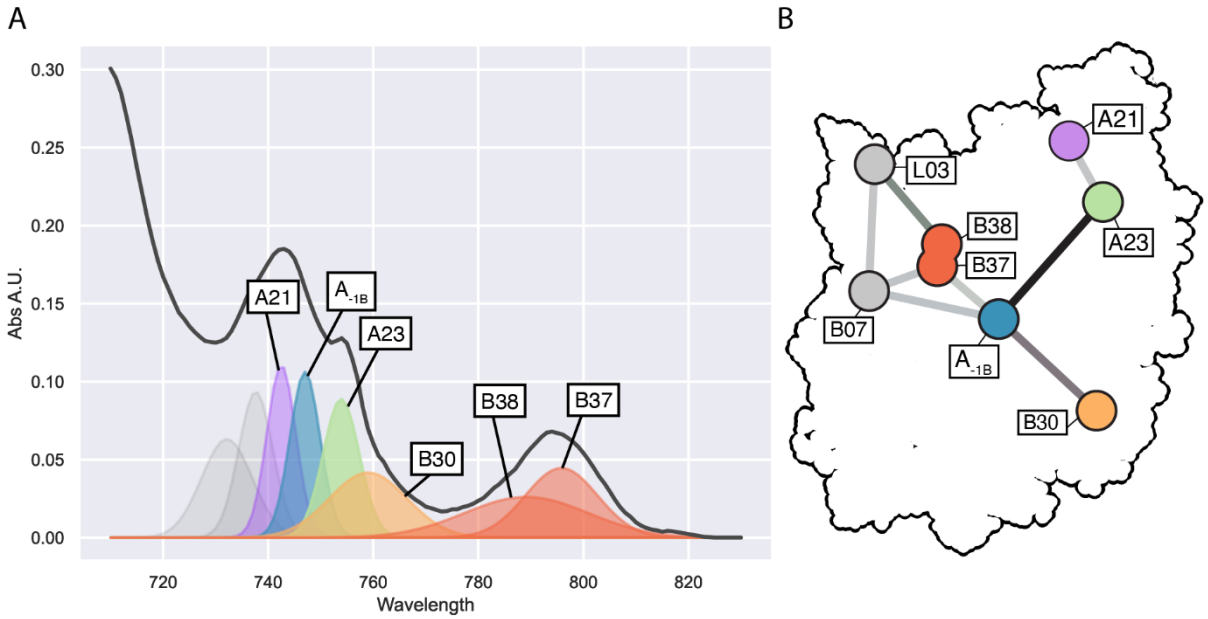


Fig. 4.20 – Wavelength assignments in FR-PSI

A) Absorption spectra at 77 K of FR-PSI from *C. thermalis*, with fitted gaussians for the 8 long-wavelength pigments. From the reddest to the bluest we assign B37, B38, B30, A23, A_{-1B}, A21, B07, L03. B) Schematic representation of the position of Chl *f* sites in a FR-PSI monomer, from the cytosolic side of the membrane.

The lack of quenching of the ~750 nm emitter likely reflects a larger distance from $P700^+$ than the 800 nm emitter. From their locations (Fig. 4.18), A21 and B30 are the most likely candidates as the emitter at ~750 nm, as already suggested (166), as they are both far from $P700^+$ (34 Å and 28 Å) and seem relatively isolated from the central core of Chl *f* molecules.

However, the wavelength assigned to B30 is 762 nm based on its contribution to the band-shifts in the difference spectrum (Fig. 4.17). Therefore, A21 appears to be the more likely of the two candidates. Given the proximity of A21 to A23, for A21 to be the terminal emitter, A23 would have to absorb at a shorter wavelength within the range assigned to it above (740-760 nm), otherwise A21 would transfer excitation energy to the lower energy pigment (A23) rather than emit. The wavelength assignment of Chl B30 and Chl A21 also fits with the structural observation of a far-red conserved change hydrogen bonding the 13^1-keto group of Chl B30 (Fig. 4.11A). This type of interaction is expected to red-shift the absorption spectra of Chl *a* molecules and can provide insight on the behaviour of Chl *f* in the same conditions. While individual Chl *f* molecules can be tentatively assigned to specific absorptions in the spectrum (A_{-1B}, 755 nm, B37/B38 pair at 800 nm, B30 at 762 nm), there is insufficient evidence to assign the others. Now they have all been localised, future work will focus on their physical and functional properties.

4.4 Conclusions

The two structures presented in this chapter provide and enrich the evidence for the positions of all the 8 expected Chl *f* sites in FR-PSI (i.e., A_{-1B}, A21, A23, B07, B30, B37, B38, and L03). Given the poorer resolution of the maps previously available and due to the challenges described in chapter 3, only the five Chl *f* that present a hydrogen bond to the C2 formyl group were correctly assigned. The other 3 sites, where the formyl oxygens are not hydrogen-bonded, i.e., A_{-1B}, B38 and L03 (Fig. 4.5, Fig. 4.13, Fig. 4.14), are more difficult to detect because of their increased negative partial charge and because they may occupy a broader range of torsion angles. This class of formyl oxygens therefore required improved resolution and the development of more thorough statistical analysis to allow them to be detected in Cryo-EM maps.

The present work also confirms that both positive and negative ESP signals can provide information on the electrostatic environment and the orientation of functional groups of cofactors and amino acid side chains, as discussed in the literature (155,162,163). In this chapter, the information obtained from a more comprehensive analysis of the ESP has provided additional mechanistic insights. When present, the hydrogen bonds to the formyl

oxygen not only stabilise the binding of the Chl *f*, but also shift the redox properties of the chlorophylls, making them easier to reduce and harder to oxidise. This would not greatly affect the antenna function of chlorophylls (other than small changes in their absorption peaks), but it would be detrimental for Chl *f* when it comes to a redox role, and particularly as the primary donor, which is the most reducing species in PSI, the low-potential photosystem. Therefore, a Chl *f* acting as a primary donor in PSI would be expected to lack an hydrogen bond to its formyl group and thus be difficult to detect by Cryo-EM. Both expectations were found to be the case for the Chl *f* in the A_{-1B} site. In fact, the environment of FR-PSI A_{-1B} Chl *f* appears to be specifically negatively charged (Fig. 4.5, Fig. 4.16), an indication of redox tuning.

The identification of Chl *f* at the L03 site in the core of the trimer and at the A_{-1B} site in the FR-PSI reaction centre represent significant advances in understanding how FR-PSI works. The L03 Chl *f* indicates exciton energy sharing between the PSI monomers in the trimer, which is crucial in low light environments, suggesting a relationship between physiological needs and oligomerization states of PSI. Moreover, the presence of a long-wavelength pigment in the A_{-1B} reaction centre site confirms photochemistry with Chl *f* as the primary donor. The previous spectroscopic studies of FR-PSI, which assumed no Chl *f* in the reaction centre, and their resulting mechanistic models of excitation energy transfer and charge separation, should now be reassessed. FR-PSI is therefore comparable to Chl *d* PSI in *Acaryochloris marina*, in that they both use near-IR photons (P740) for oxygenic photosynthesis, breaking the red photochemical energy limit. As seen with PSII (43,98), these two distinct low energy paradigms have evolved different strategies to perform the same chemistry of Chl *a* containing complexes but by using less energy. Comparisons of the two long-wavelength paradigms with each other and with WL-PSI, will advance our understanding of the mechanisms of all three. The tentative assignment of five out of eight low-temperature absorption peaks to specific sites allows a more targeted approach to site directed mutagenesis of the Chl *f* sites, with the goal of elucidating the wavelength and excitation energy transfer pathways, as well as the effect of the chemical environment among the Chl *f* sites in FR-PSI.

5 A new subunit and chlorophyll *f* assignments in CryoEM structures of FR-PSII

5.1 Introduction

Photosystem II is the first complex in the photosynthetic electron transport chain and catalyses a unique multi-electron redox reaction (192). Thanks to the energy coming from the sun in the form of photons, it can abstract electrons from water molecules, reduce plastoquinones and as a by-product, energise the atmosphere with the release of molecular oxygen.

Canonical white light PSII (WL-PSII) contains ~20 subunits and coordinates 35 Chl *a* molecules, most of which are used to harvest photons and funnel excitation energy to the reaction centre (5). In the FaRLiP cluster, 5 genes encoding long-wavelength-adapted paralogues of the main subunits in PSII are present (109). PsbA and PsbD, known as D1 and D2, are the pseudo-symmetrical central reaction centre subunits that contain (nearly) all the cofactors involved in PSII activity. PsbB and PsbC, known as CP43 and CP47, are the two main antenna proteins of the core. Finally, PsbH is a single-helix membrane-spanning subunit that plays a role in PSII synthesis, maturation, repair cycle (193), antenna size regulation, and efficient electron transfer between the quinone acceptors (194). These subunits coordinate 4 Chl *f* and a single Chl *d* (43).

Unlike FR-PSI, spectroscopic, biophysical, and structural data mostly agree on the nature and locations of the chlorophylls present in its reaction centre. Most literature considers the long-wavelength chlorophyll to be a Chl *d* located at position Chl_{D1}, where it is directly involved in charge separation (43,130,195). However, circular dichroism and magnetic circular dichroism studies allow other (i.e., the P_{D1} position) options to be considered (196). On the other hand, due to the poor intactness of the samples and the overall resolution of the maps, the structural assignments of the antenna Chl *f* have been less convincing. The first FR-PSII structure published (130) featured a PSII monomer missing most of the extrinsic subunits, including the far-red light paralogue PsbH2. In this structure, many

chlorophylls are missing and some structural features not typical of intact PSII, probably arising from isolation-induced damage, are present. More recently, the structure of a more intact FR-PSII dimer was solved (111), with extrinsic proteins present at partial occupancy together with all 35 expected chlorophyll molecules. Nonetheless, the improved resolution still allowed only a portion of the Chl *f* molecules to be detected and quantitatively assigned with methods reported in the literature (155).

As in the case of FR-PSI, the assignment of Chl *f* sites in FR-PSII is crucial to understand the routes of excitation energy transfer in this low energy paradigm of PSII. The energy-intensive redox chemistry performed by PSII comes at a price: backreaction and formation of dangerous reactive oxygen species cause the population of Photosystem II to be continuously damaged and replaced. The functional adaptations and trade-offs used by FR-PSII to perform the same photochemistry as WL-PSII, with a lower energetic initial excited state (~100 meV less), still need to be fully understood and characterised. In FR-PSII the charge separation steps need to be correctly separated in terms of standard free energy, so that Chl triplet formation and the consequent formation of singlet oxygen are minimised (14). Relating the functional adaptations to far-red light with conserved structural changes should improve our understanding of both far-red and canonical PSII function and is also potentially useful to guide biotechnological applications of this system.

To address these questions, we purified FR-PSII and determined structures of both intact and incomplete FR-PSII monomers, as well as an intact dimeric FR-PSII. These structures were resolved at 2.35 Å, 2.58 Å, and 2.17 Å, respectively. Thanks to these improvements in resolution, and the methodologies described in Chapter 3, the position of all four Chl *f* molecules could be confidently assigned, and the presence of Chl *d* at the Chl_{D1} position was confirmed. In addition, we identified a novel FR-PSII subunit encoded by a gene within the FaRLiP cluster, that is present across all intact structures. The subunit has been structurally characterised and named PsbH2' due to its extensive interactions with PsbH2. Moreover, hypotheses on the function of PsbH2' as well as the tentative assignment of the absorption wavelengths of Chl *f* sites are presented in this chapter.

5.2 Results

FR-PSII was isolated as described in the Materials and Methods section. The micrographs were collected in multiple sessions in an effort to acquire a structure of an intact dimeric FR-PSII. The initial particle stack, composed of PSII-like particles, was sorted into three *ab initio* classes that were used to perform multiple rounds of heterogeneous refinement and 3D classification. These classes led to the refinement of monomeric and dimeric structures of intact, oxygen-evolving FR-PSII (Holo-FR-PSII) and a structure of a incomplete monomeric FR-PSII (Apo-FR-PSII), lacking the manganese cluster and extrinsic subunits.

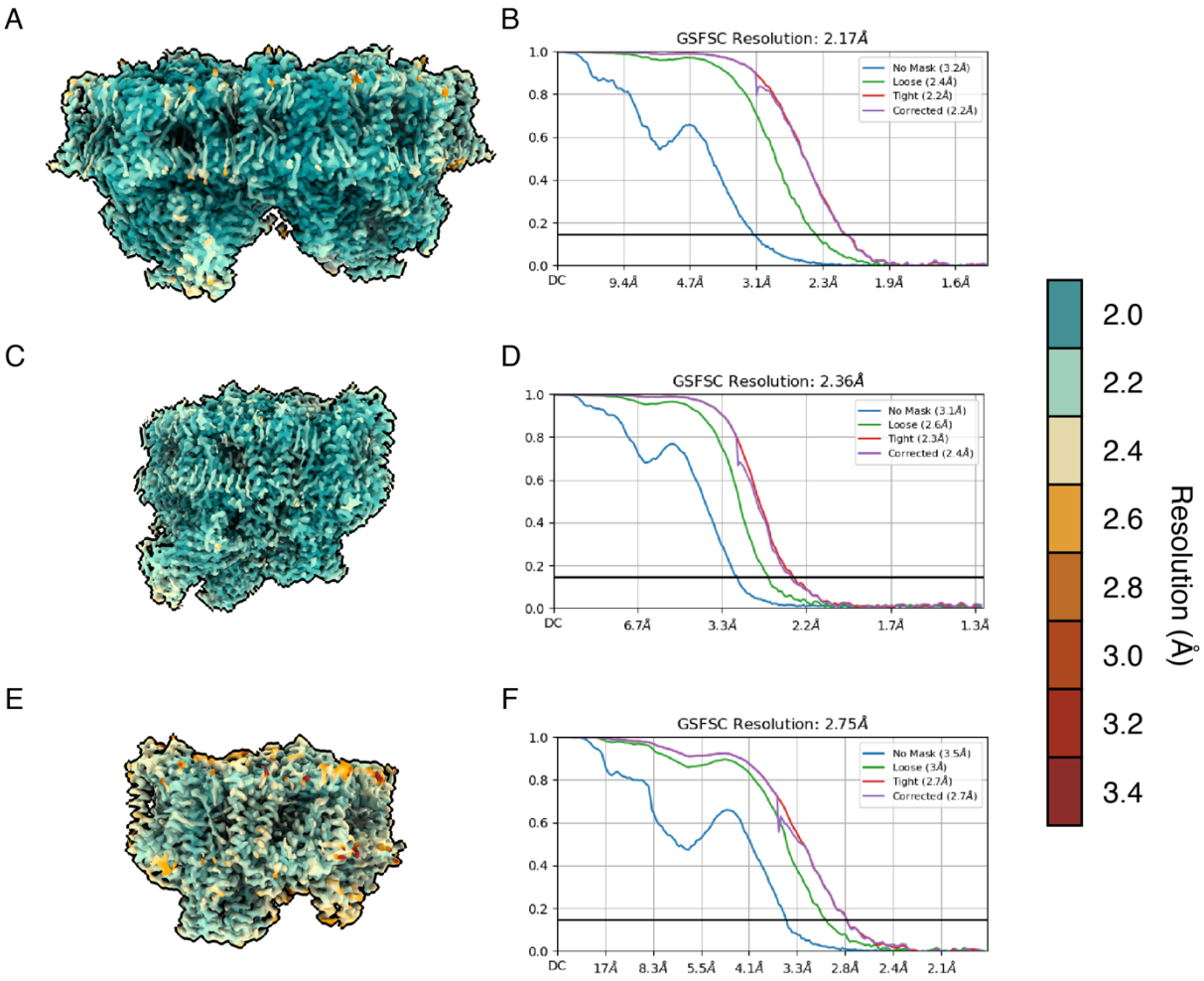


Fig. 5.1 – Local resolution of FR-PSII maps

Visual representation and GS-FSC of the structures discussed in this chapter, the colour-map is provided on the right. A) Dimeric intact FR-PSII map coloured by local resolution; B) GS-FSC of the dimeric intact FR-PSII map; C) Monomeric intact FR-PSII map coloured by local resolution; D) GS-FSC of the monomeric

intact FR-PSII map; E) Monomeric Apo-FR-PSII map coloured by local resolution; F) GS-FSC Monomeric Apo-FR-PSII map presented.

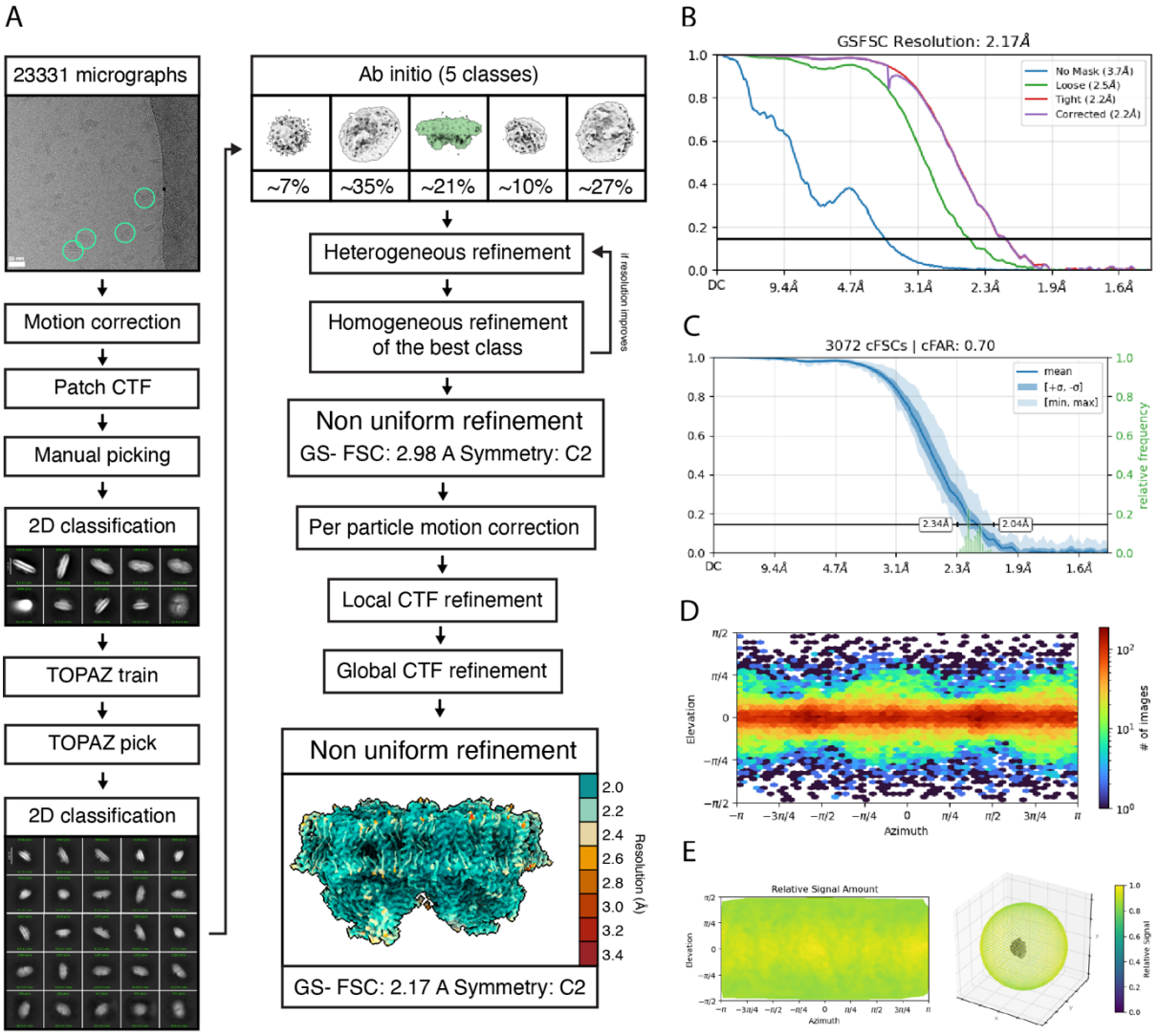


Fig. 5.2 – Processing workflow and reconstruction diagnostic plots for dimeric FR-PSII

A) processing workflow to obtain the final reconstruction of the dimeric FR-PSII complex. B) GS-FSC curve for the dimeric FR-PSII complex. C) cFSC curves for the dimeric FR-PSII complex. D) Viewing direction distribution of the particle stack. E) relative signal amount for each of the viewing directions.

Data collection

Microscope	Krios III
Camera	Falcon 4i
Magnification	155000x
Voltage (kV)	300
Electron exposure (e-/Å^2)	40
Defocus range (μm)	-0.8 to -2
Pixel size (Å)	0.723
Energy filter	Selectris (20 eV)
Exposures	23311
Image format	EER

Data processing

Box size	650 px
Initial particles (no.)	423204
Final particles (no.)	57532
Symmetry	C2
Map resolution (Å)	2.17
Map sharpening *B* factor	-40.3

Model Refinement

Refinement package	PHENIX
Initial model used	8EQM
Real/reciprocal space	Real Space
Resolution cutoff	2.20

Table 4 – Data collection and processing parameters for dimeric FR-PSII

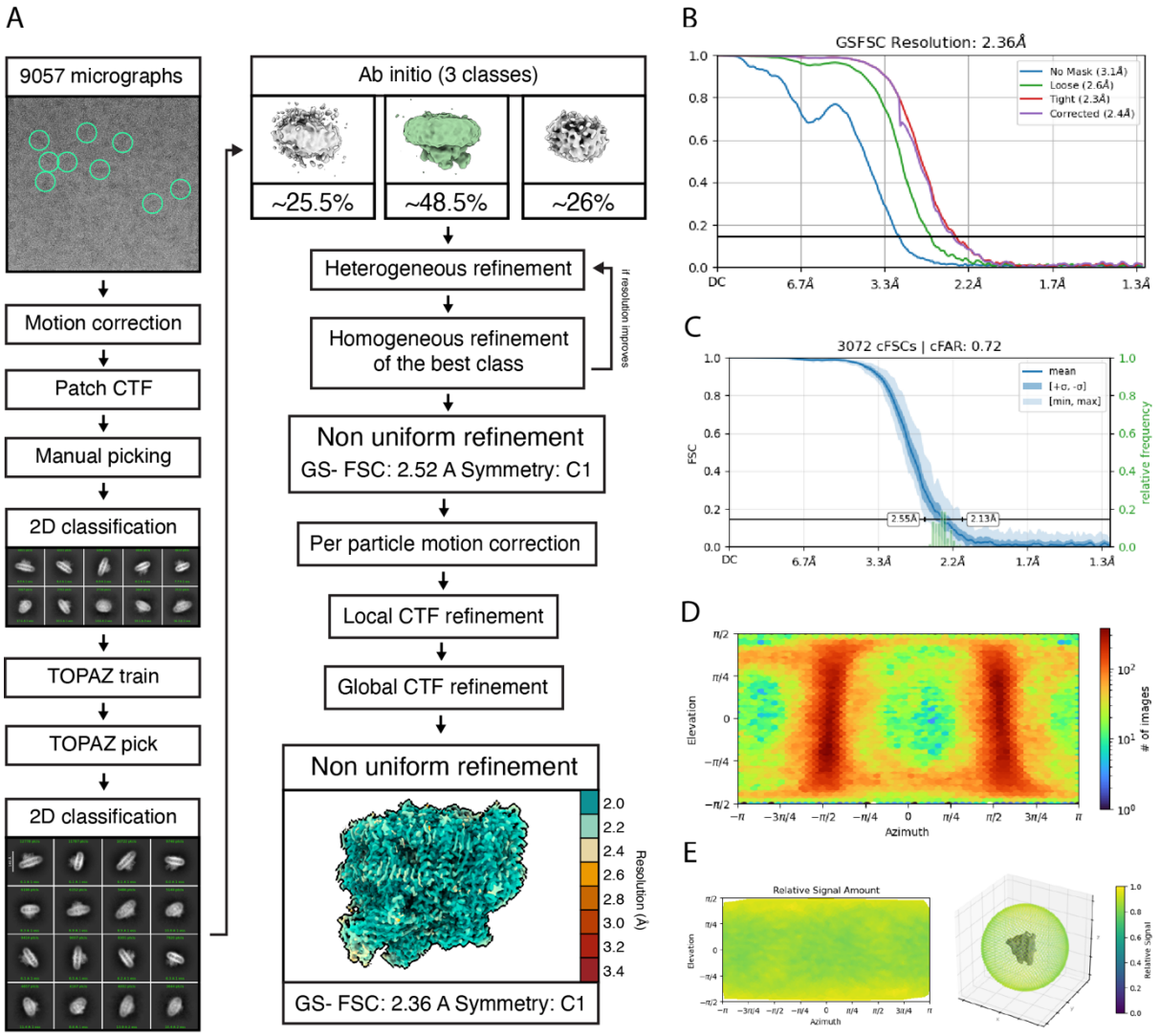


Fig. 5.3 – Processing workflow and reconstruction diagnostic plots for monomeric FR-PSII

A) processing workflow to obtain the final reconstruction of the monomeric FR-PSII complex. B) GS-FSC curve for the monomeric FR-PSII complex. C) cFSC curves for the monomeric FR-PSII complex. D) Viewing direction distribution of the particle stack. E) relative signal amount for each of the viewing directions.

Data collection

Microscope	Krios IV
Camera	K3
Magnification	155000x
Voltage (kV)	300
Electron exposure (e-/Å^2)	50
Defocus range (μm)	-0.8 to -2
Pixel size (Å)	0.651
Energy filter	N/A
Exposures	9057
Image format	EER

Data processing

Box size	512 px
Initial particles (no.)	423204
Final particles (no.)	179581
Symmetry	C1
Map resolution (Å)	2.36
Map sharpening *B* factor	-61.6

Model Refinement

Refinement package	PHENIX
Initial model used	8EQM
Real/reciprocal space	Real Space
Resolution cutoff	2.40

Table 5 – Data collection and processing parameters for monomeric FR-PSII

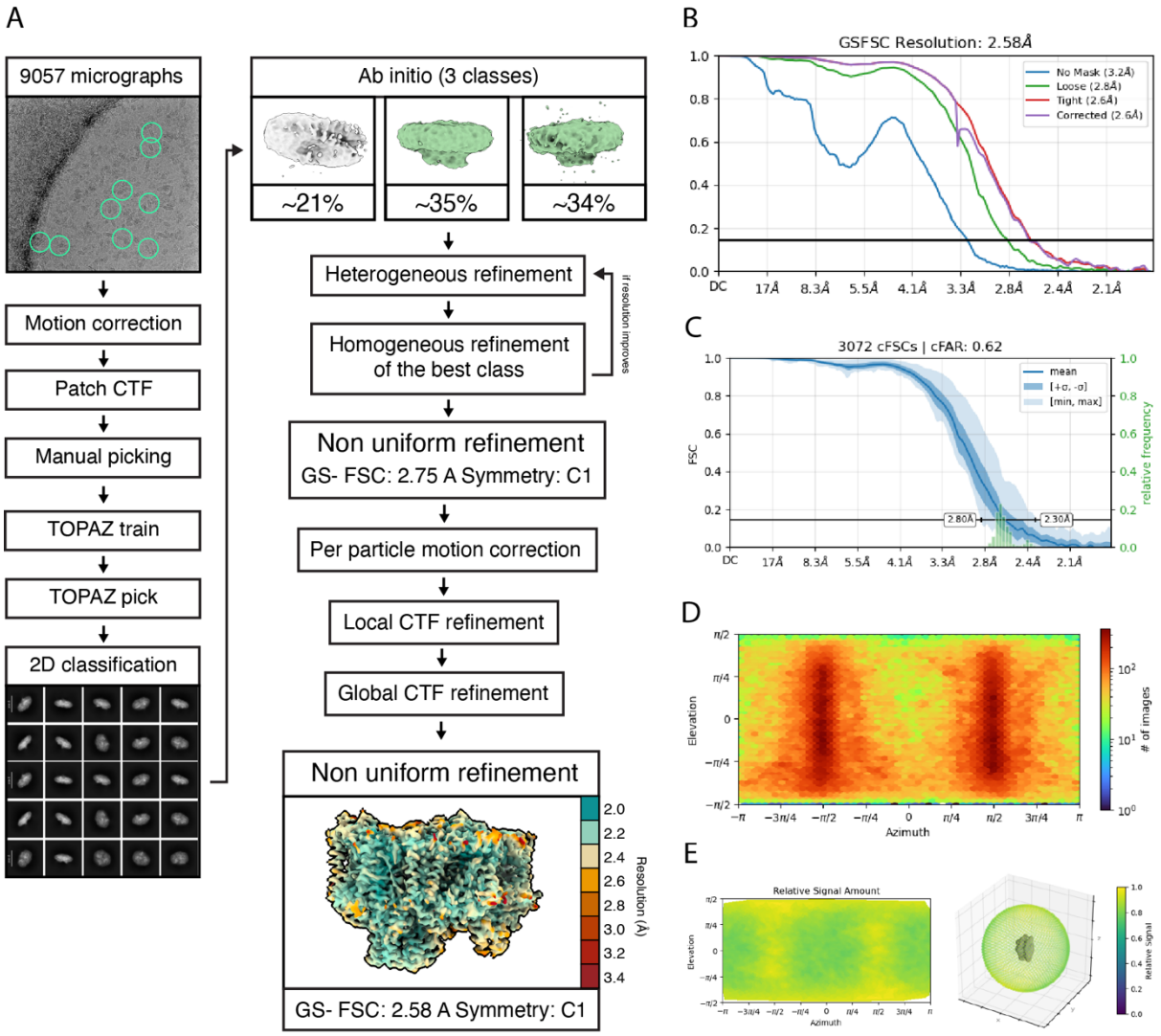


Fig. 5.4 – Processing workflow and reconstruction diagnostic plots for Apo-FR-PSII

A) processing workflow to obtain the final reconstruction of the Apo-FR-PSII complex. B) GS-FSC curve for the FR-APC complex. C) cFSC curves for the Apo-FR-PSII complex. D) Viewing direction distribution of the particle stack. E) relative signal amount for each of the viewing directions.

Data collection

Microscope	Krios III
Camera	Falcon 4i
Magnification	130000x
Voltage (kV)	300
Electron exposure (e-/Å^2)	40
Defocus range (μm)	-0.5 to -3.5
Pixel size (Å)	0.921
Energy filter	Selectris (20 eV)
Exposures	8.854
Image format	EER

Data processing

Box size	450 px
Initial particles (no.)	606396
Final particles (no.)	226959
Symmetry	C1
Map resolution (Å)	2.58
Map sharpening *B* factor	-67.5

Model Refinement

Refinement package	PHENIX
Initial model used	8EQM
Real/reciprocal space	Real Space
Resolution cutoff	2.60

Table 6 – Data collection and processing parameters for monomeric Apo-FR-PSII

The Cryo-EM ESP maps were generated at a global resolution of 2.17 Å for the dimeric Holo-FR-PSII, 2.35 Å for the monomeric Holo-FR-PSII, and at 2.58 Å for the monomeric Apo-FR-PSII (Fig. 5.1). The map of both Holo- and Apo-FR-PSII presented clear density to model 21 protein chains that coordinate the following cofactors: 2 pheophytin *a*, 10 β-carotene, the manganese/calcium cluster, two chloride ions, 2 quinones, 1 non-heme iron, 1 bicarbonate molecule, 1 heme B, 1 heme C, multiple lipids and 35 chlorophyll molecules, of which 30 Chl *a*, 4 Chl *f* and 1 Chl *d*. The apo-monomeric FR-PSII is not capable of evolving oxygen, presents an incomplete donor side, with low partial occupancy of the oxygen-evolving complex, and lacks the extrinsic subunits PsbU, PsbV and PsbO, as well as the transmembrane accessory subunits PsbJ and PsbR.

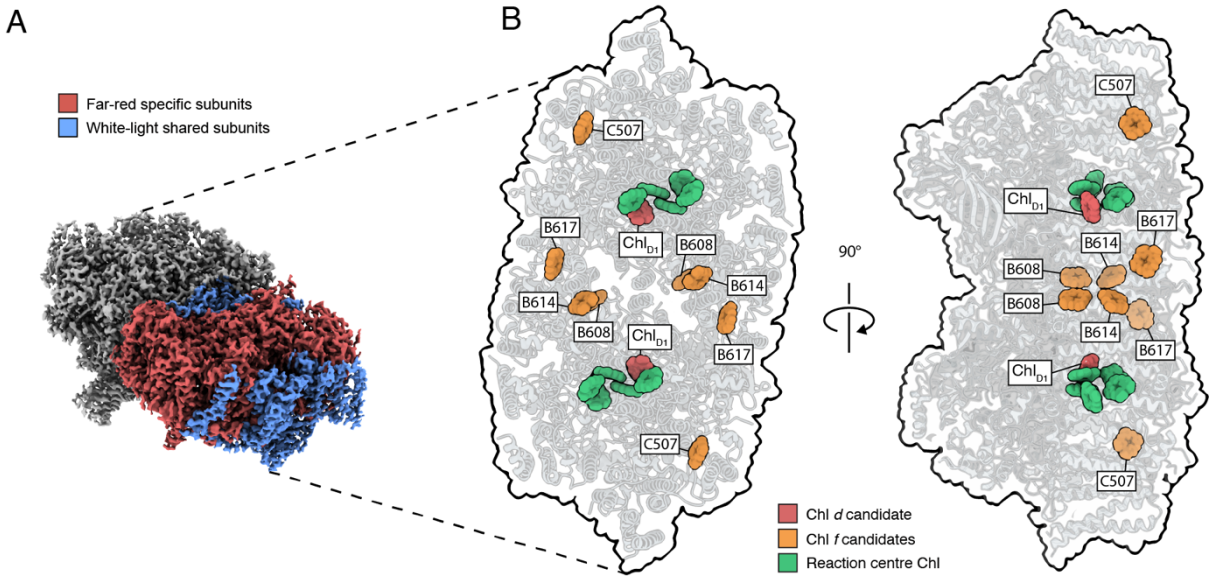


Fig. 5.5 – Positions of the Chl *f* candidates in FR-PSII

A) View of the dimeric FR-PSII map from a cytoplasmic/side view. The map presents clear ESP for all the known PSII subunits, among which the five already known far-red light specific subunits (PsbA2, PsbB2, PsbC2, PsbD2, PsbH2) and the novel far-red specific subunit (PsbH2') are represented in red, while the other fourteen subunits in common with WL-PSII are represented in blue. B) Location of the 5 candidate long-wavelength pigments in a FR-PSII dimer as viewed from the cytosolic face and membrane plane. Chl *a* of the reaction centre pigments are represented in green, the Chl *d* primary donor molecule in red and the Chl *f* molecules are represented in orange, the other 30 Chl *a* antenna molecules were omitted for clarity.

In the following section a description of the candidate Chl *f* and Chl *d* sites chemical environments is provided, as well as a list of structural features that diverge from the canonical WL-PSII (Fig. 5.5).

5.2.1 *Chl f* C507

The Chl *f* containing site at C507 has been already described in the literature (111,130,197) and presents clear phylogenetic changes that allow the incorporation of Chl *f*. As in WL-PsbC, the backbone carbonyl of PsbC2 Tyr^{264} coordinates a water molecule that stabilises the Mg^{2+} of the Chl. Moreover, while the far-red specific PsbC2 Asn^{271} (Ala in WL-PSII) provides the hydrogen bond partner to the C2 formyl moiety (~3.3 Å). The torsion angle of this residue is held in place by the presence of another conserved change, PsbC2 Gln^{275} (Met in WL-PSII), that is likely to improve the stability of the site and possibly increase the electron-withdrawing effect of PsbC2 Asn^{271} on the C2 formyl of Chl C507. The analysis of the ESP map indicates the presence of increased electrostatic potential around the C2 group, which can be attributed to the presence of a formyl oxygen (Fig. 5.6B, C).

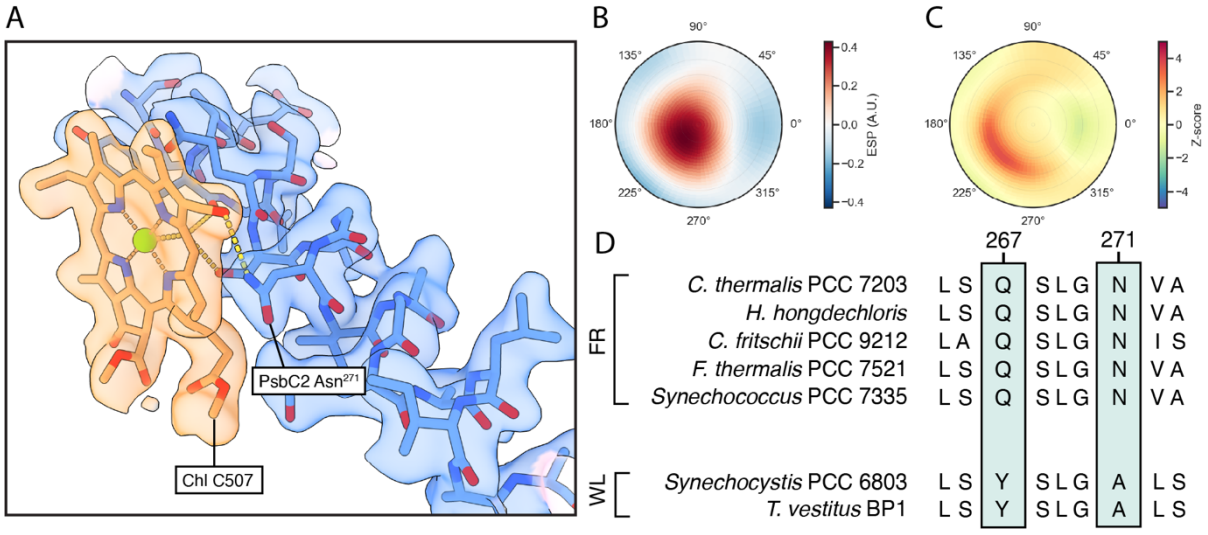


Fig. 5.6 – Chl *f* C507 site

A) ESP map and atomic model around the C507 site. The Chl density is in orange, the density of the amino acid residues in blue. Hydrogen bonds are indicated as dotted lines. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. D) Multiple sequence alignment of FR- and WL-PsbC.

Moreover, the environment of an adjacent Chl *a* is modified (197). A far-red specific interaction, originating from the far-red conserved PsbC2 Gln^{267}, provides an hydrogen bond to the 13^1-keto group of Chl *a* C505. This type of interaction is hypothesised to result in a lowering of the Chl *a* site energy, with a consequent red-shift of the absorption profile

(177). This structural feature is likely to be important for efficient energy transfer to the reaction centre.

5.2.2 Chl f B608

As described previously, phylogenetic and ESP analyses indicates that Chl B608 is a Chl *f* (111). The Mg^{2+} of this Chl is coordinated by a water molecule, as preferred by Chl *f* molecules, in a network involving the residue PsbB2 Glu^{41} and the carbonyl backbone of PsbB2 Gly^{59} as in WL-PSII (Fig. 5.7A).

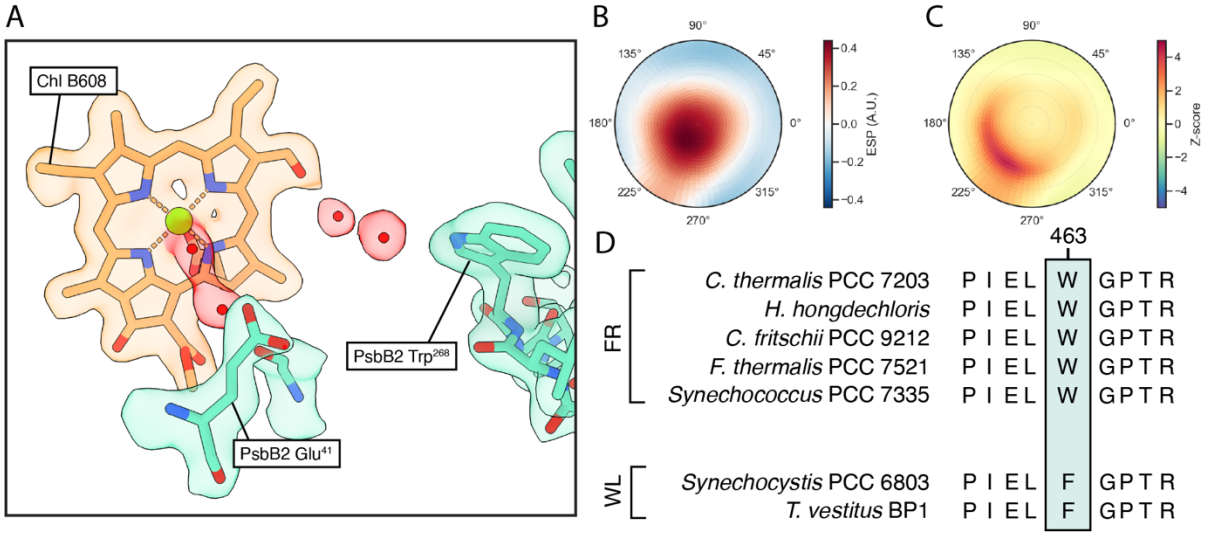


Fig. 5.7 – Chl *f* B608 site

A) ESP map and atomic model around the B608 site. The Chl density is in orange, the density of the amino acid residues in green and the density from water in red. Hydrogen bonds are indicated as dotted lines. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. D) Multiple sequence alignment of FR and WL PsbB.

The C2 formyl group of Chl *f* B608 is coordinated by a water molecule, for which clear density is present, which in turn connects it to the carbonyl backbone of PsbB2 Asn^{447} and another water molecule that hydrogen bonds the far-red specific residue PsbB2 Trp^{268}. Analysis of the ESP map around the C2 group of this site indicates the presence of a formyl oxygen (Fig. 5.7B, C). Since not all far-red sequences available have PsbB2 Trp^{268}, Chl *f* B608 has been considered to be present only in some FaRLiP-capable species (197). Nonetheless, this hypothesis seems unlikely, since the site would be stabilised anyway by the presence of the interaction with the carbonyl backbone of PsbB2 Asn^{447}.

5.2.3 *Chl f B614*

Chl B614 has also been suggested as a Chl *f* candidate from previous structural and phylogenetic data. Analysis of the ancestral sequence reconstruction of PsbC2 in FaRLiP-capable organisms, suggests this site as one of the first to have evolved, highlighting its importance in the function of FR-PSII (111,197).

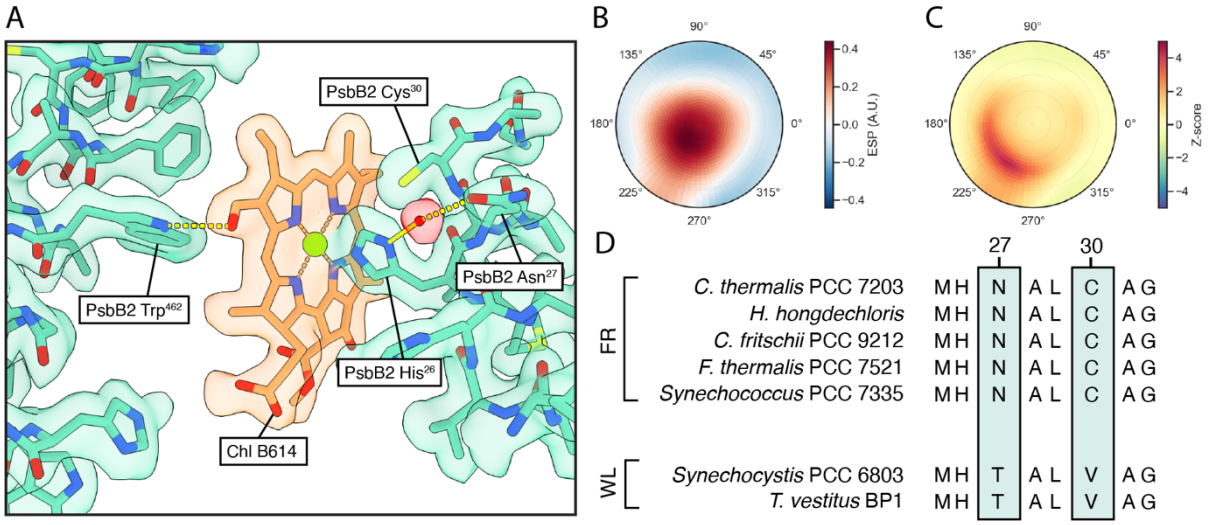


Fig. 5.8 – Chl *f* B614 site

A) ESP map and atomic model around the B614 site. The Chl density is in orange, the density of the amino acid residues in green and the density from water in red. Hydrogen bonds are indicated as dotted lines. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. D) Multiple sequence alignment of FR and WL PsbB.

In all of our maps, the C2 formyl group presents increased electrostatic potential, indicating the presence of a formyl oxygen (Fig. 5.8B, C). Moreover, some fully conserved far-red specific changes are present in the vicinity of Chl B614's C2 formyl group. A WL Phe is substituted by PsbB2 Trp^{462} which acts as an hydrogen bond donor for the C2 formyl oxygen. The Chl in this site presents a peculiarity compared to other sites, its central Mg^{2+} ion is coordinated by PsbB2 His^{26}, and not by a water molecule or polar residue as it would be expected for Chl *f* (93). Nonetheless, the ε-nitrogen of PsbB2 His^{26} is coordinated by a water molecule, which is in turn bound to a conserved change PsbB2 Asn^{27} rather than by the WL-PSII specific Thr (Fig. 5.8A, D). The presence of a polarised amino acid in this position could increase the electrostatic pull, tuning the properties of PsbB2 His^{26} to

increase its polarity (93). Finally, another far-red conserved change, the sulfhydryl group PsbB2 Cys^{30} is facing directly the Mg^{2+} and PsbB2 His^{26} of this site, potentially altering site energies or properties of the site.

5.2.4 Chl f B617

Chl B617 has an increased ESP potential around the C2 substituent and shows a clear protruding density, which is compatible with the presence of a formyl oxygen (Fig. 5.9A, B, C). Moreover, the site presents a far-red conserved change around the environment of the C2 substituent: a WL fully conserved Leu is substituted with Glu, Gln or Ser (Ser being less common) in position 107 of PsbB2, providing an hydrogen bond partner to Chl *f* in position B617 (Fig. 5.9D). This observation is strengthened by the deposited map of FR-PSII from *Synechococcus sp.* PCC 7335. In this organism, PsbB2 Ser^{107} is present, together with the density for a bridging water molecule, capable of hydrogen bonding the C2 formyl oxygen of Chl *f* B617 (111).

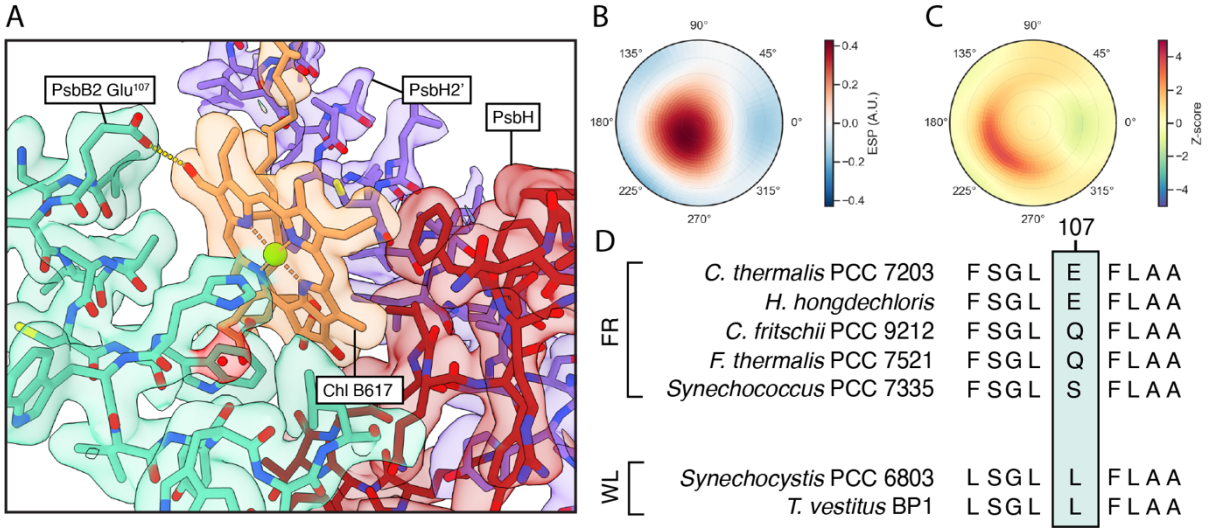


Fig. 5.9 – Chl *f* B617 site

A) ESP map and atomic model around the B617 site. The Chl density is in orange, the density of the amino acid residues from PsbB in green, the density of the amino acid residues from PsbH2 in red and the density of the amino acid residues from PsbH2' in violet and the density from water in red. Hydrogen bonds are indicated as dotted lines. B) ESP polar plot indicating the presence of an increased potential around the C2 substituent. C) Z-score polar plot indicating the presence of an increased potential around the C2 substituent. D) Multiple sequence alignment of FR and WL PsbB.

The difference in assignment with the literature regarding this site can be explained by the absence of this pigment, as well as PsbH2, in the earlier FR-PSII monomer, and with poor resolution of the stromal portion of the complex in the more recent dimeric FR-PSII structure. Chl *f* B617 is positioned at the extreme periphery of CP47, between the antenna and PsbH2. The presence of His as a Mg^{2+} ligand, which binds Chl *f* more weakly, could cause the site to be easily destabilised in absence of PsbH2. The Chl site B617 has not been proposed before as a Chl *f* containing site, and the phylogenetically conserved changes around it have never been mentioned previously. Even though the Mg^{2+} of Chl *f* B617 is bound by PsbB2 His^{114}, which is bound by a water and the backbone of PsbB2 Tyr^{117}, thanks to the improved resolution and intactness of the systems analysed here, different lines of evidence indicate the presence of Chl *f* in site B617.

The presence of a Chl *f* in this site is particularly interesting, since Chl B617 is one of the candidate for the "redmost" pigment in WL-PSII (45,198,199). Comparing the white light and far-red light site, another conserved change in PsbH is present. The white light conserved Thr that hydrogen bonds the 13^1-keto group of Chl B617 is substituted in FR-PSII with the non hydrogen-bonding Ala. This feature has been suggested in the literature to lower the site energy of Chl sites, suggesting a substantial retuning of Chl B617 in FR-PSII (43).

5.2.5 On the other proposed Chl f candidates

Among the candidates proposed in the literature as Chl *f* sites, we find Chl B611 and Chl B605 (111,197). The presence of a Chl *f* in site B611 has been proposed on the basis of a not fully conserved far-red residue, PsbB Thr^{244}. A Chl *f* in position B611 would make sense functionally, since this position is related by pseudo-C2 symmetry to the confirmed Chl *f* candidate at position C507, suggesting similar behaviour in terms of excitation energy transfer to the reaction centre (111,130). Evidence that PsbB Thr^{244} is present in the ancestral sequence of PsbB2 would imply that this site was important in FR-PSII, but this does not fit with the evidence of this site being lost throughout evolution in many species (197). Chl B605 has also been proposed as a species specific Chl *f* site, present only in a portion of FaRLiP capable organisms but the absence of conserved hydrogen bond donors

and the presence of a canonical His ligand for the Mg^{2+} for this site all favour the hypothesis that it binds a Chl *a*. In the maps presented here, no evidence of increased ESP for the putative C2 formyl groups for either of these two Chl sites is present. Moreover, the strong evidence pointing to the other 4 candidates (described above), which show completely conserved phylogenetic traits, argues against the hypothesis that FaRLiP species have species specific Chl *f* sites in FR-PSII (197) and suggests that the 4 antenna positions that evolved in FaRLiP capable strains are common to all far-red species and thus mechanistically important for maximizing excitation energy transfer and/or reducing the damage to the reaction centre.

5.2.6 A new FR-PSII specific subunit: PsbH2'

During the initial analysis and fitting of FR-PSII models into the map, the presence of a clear ESP density for an extra subunit composed of a single transmembrane helix was noticed (Fig. 5.10). A survey of the available structures to understand which known PSII subunits occupy this position yielded no results, and therefore Modelangelo (200), an automatic model builder tool, was employed to obtain a tentative initial sequence of the subunit.

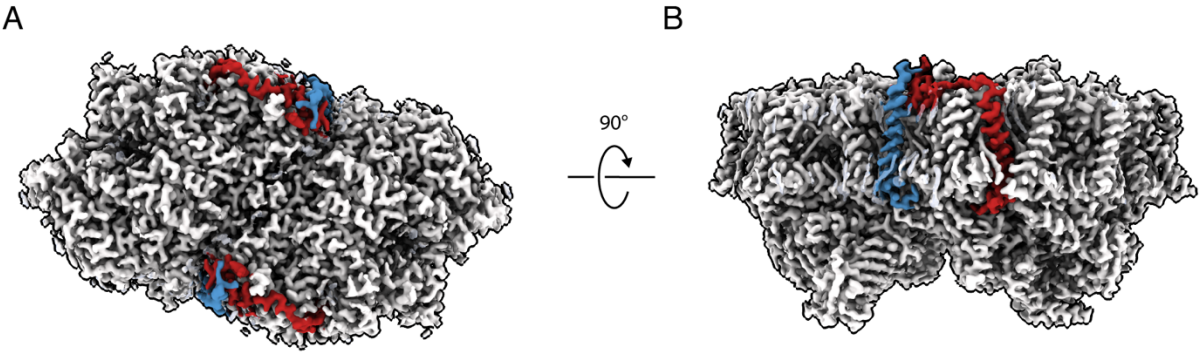


Fig. 5.10 – Position of PsbH2' in the dimeric FR-PSII ESP map

Representation of the position of the uncharacterised density in the FR-PSII map. The density of the PSII map is represented in white, the density of PsbH2 is represented in red and the density of the "mystery helix" is blue.

The sequence obtained was used to perform a BLAST search against the proteome of *C. thermalis* PCC 7203 which showed a high degree of sequence similarity with an uncharacterised hypothetical protein present in the FaRLiP cluster. The sequence

obtained fitted perfectly in the ESP density present in the map. The hypothetical protein, from here on named PsbH2', is positioned adjacently to CP47 on the opposite side of PsbH, interacting with it in the stromal region of the complex. PsbH2' is positioned near the monomer interface of a PSII dimer but does not interact with any subunit of the opposite monomer (Fig. 5.10).

5.2.7 Phylogenetic analysis of PsbH2 and PsbH2'

Sequence alignments of PsbH from a diverse set of oxygenic phototrophs, some capable of and some incapable of FaRLiP, were performed. Sequence identity and structural comparisons of PsbH2 with cyanobacterial PsbH show a high degree of sequence divergence, consistent with the Cα superposition, where the RMSD is ~0.6 Å (against an RMSD of ~0.2 Å among any two of PsbH). The multiple-sequence alignment shows that the conserved changes cluster in the N-terminal region of PsbH, which is strongly conserved among PsbH2 sequences.

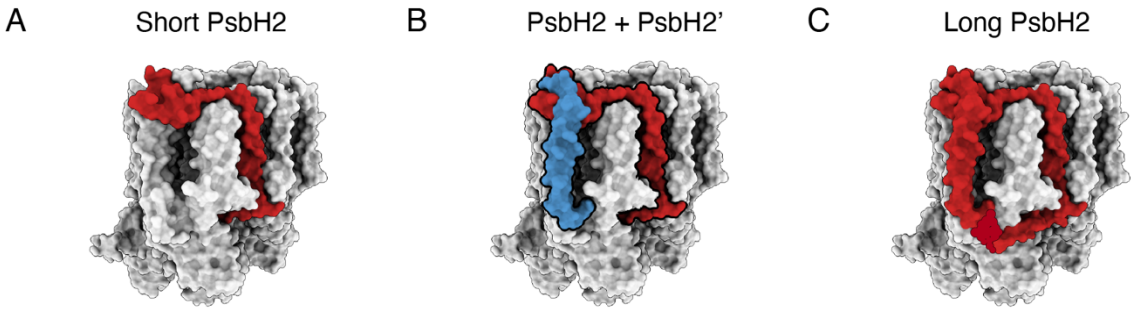


Fig. 5.11 – Phylogenetic groups of PsbH2 and PsbH2' sequences.

The figure represents the three distinct groups of PsbH2 families relative to a surface representation of FR-PSII- A) Short form of PsbH2, in these cyanobacteria the sequence for PsbH2' is absent (i.e., *Synechococcus sp.* PCC 7335) B) Short form of PsbH2 in cyanobacteria that also present the sequence for the extra far-red specific protein PsbH2'. C) Long form of PsbH2, the contains the sequence of PsbH2 fused to PsbH2' by a linker (cartoon) that joins them on the lumenal side of CP47.

Phylogenetic analyses of PsbH2 sequences reveal the presence of 3 distinct groups of FaRLiP strains. The first group presents a normal length PsbH2 and has no PsbH2' (Fig. 5.11A), the second group presents a normal length PsbH2 and PsbH2' in the FaRLiP cluster (Fig. 5.11B) and the third group presents an elongated PsbH2 sequence, roughly of the size of PsbH2 and PsbH2' combined (Fig. 5.11C). Alignments of white light and far-

red-light-adapted PsbH sequences reveal that psbH2 has a slightly longer N-terminal region (~7 aa) and a shorter C-terminal region compared with PsbH. Interestingly, the plant version of PsbH shows change in its termini more similar to those in PsbH2 than to PsbH.

5.2.8 Confirming B617 is the redmost Chl of WL PSII

To confirm whether Chl B617 is the "red-most" Chl of WL-PSII a single point mutation approach was adopted. A single-point PsbH mutant of *C. reinhardtii* was produced, in which PsbH Thr[5] was substituted with the non-hydrogen-bonding amino acid Ala. This mutant was used to acquire a low-temperature fluorescence spectra and assess potential changes in its fluorescence emission bands. The 77 K fluorescence emission spectra of the wild-type organism and the mutant show how the Thr to Ala mutant exhibits only one emission peak attributable to WL-PSII (at 685 nm), confirming the hypothesis that the hydrogen bond from PsbH Thr[5] to Chl B617 produces a shift in site energy that leads to a fluorescence emission shoulder at 693 nm to the main emission peak of WL-PSII (177,201).

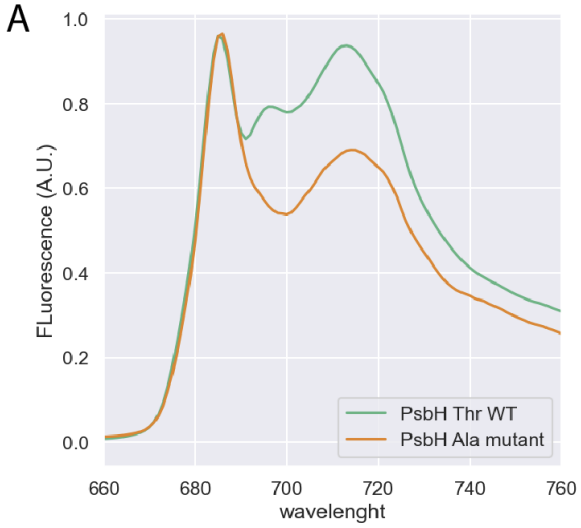


Fig. 5.12 – 77 K fluorescence emission in the PsbH Thr to Ala *C. reinhardtii* mutant

The 77 K emission spectra shows of both the mutant (orange) and the WT (green) presents all of the recognizable feature of a Chl *a* organism. The emission peak at 685 nm is originated from PSII, the broader peak at 720 nm is typical of WL-PSI, which the smaller shoulder in between them, roughly at 693 nm and present only in the WT is the emission coming from the redmost chlorophyll of WL-PSII.

5.2.9 Structural differences between broken and intact FR-PSII

The availability of structures for both Apo- and Holo-FR-PSII allows their comparison both in terms of subunit composition and RMSD. The two structure show an average RMSD of 0.327 Å; the incomplete structure lacks the extrinsic subunits PsbU, PsbV and PsbO, as well as the transmembrane accessory subunits PsbJ and PsbR (Fig. 5.13), and is not capable of evolving oxygen. The absence of known subunits involved in assembly or disassembly of PSII likely signifies that this structure represents damaged FR-PSII (202).

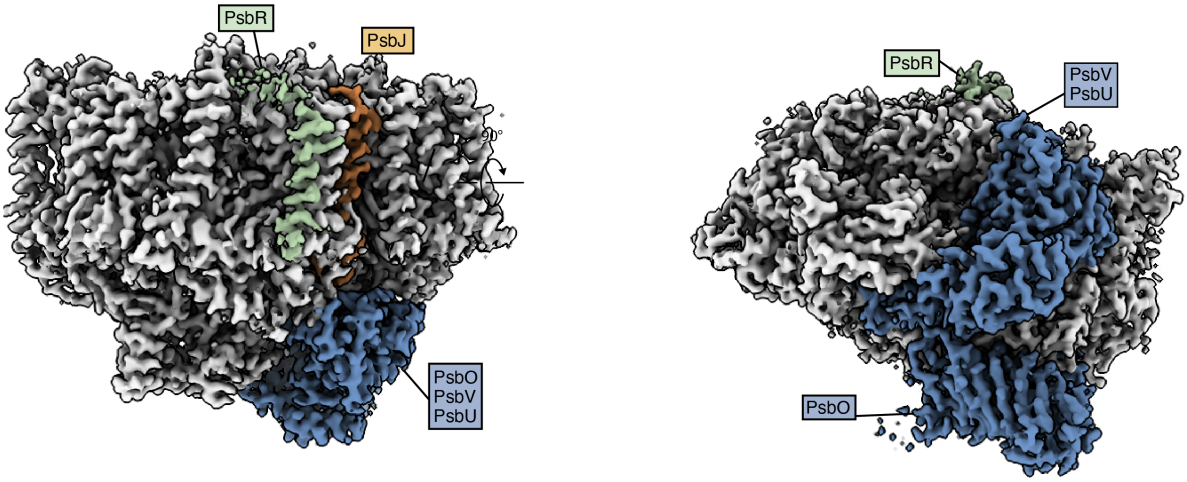


Fig. 5.13 – Comparison between Apo- and Holo-FR-PSII

Side and bottom view of FR-PSII in which the subunits that are lost in Apo-FR-PSII are coloured. PsbR is represented in green, PsbJ in orange and PsbO, PsbV and PsbU are represented in blue.

With the loss of all of the subunits that protect the oxygen-evolving complex, the donor side is damaged, with conformational changes in PsbA and PsbB. At the position of the manganese cluster an ESP density of similar size is present; while some of the amino acids that normally coordinate it present clear side chain density, other do not, suggesting that the normal coordination of the cluster is altered. This may indicate that this damaged form retains some of the atoms that constitute the cluster, or that the damaged FR-PSII structure presented might be an assembly or photoactivation intermediate. The resolution obtained does not allow to validate this hypothesis. Conversely, Apo-FR-PSII presents an intact acceptor side, preserving the Q_a site, the Fe^{2+} ion and its bicarbonate ligand, but lacking the mobile quinone Q_b. The loss of the subunits PsbR and PsbJ seems to directly influence the position of the subunits PsbE and PsbF, which together form Cyt b_{559} (Fig.

5.13B). PsbJ is normally positioned between PsbK and PsbE, interacting in its stromal portion with both. Moreover, Lys[66] PsbV interacts with the C-terminal carboxylic group of Leu[39] PsbJ, stabilizing the interaction. At the same time, PsbR interacts over the entire length of its transmembrane subunit with PsbE and shields the haem in Cyt b_{559} from the solvent. The absence of this interaction in the Apo-FR-PSII structure leads to a torsion and rotation of the PsbF helix, with effects on the position and orientation on the haem cofactor, likely influencing its redox potential. Given the extensive network of interactions between these 5 subunits, it's clear how minor disruptions could cause cascade effects that lead to the dissociation of these subunits.

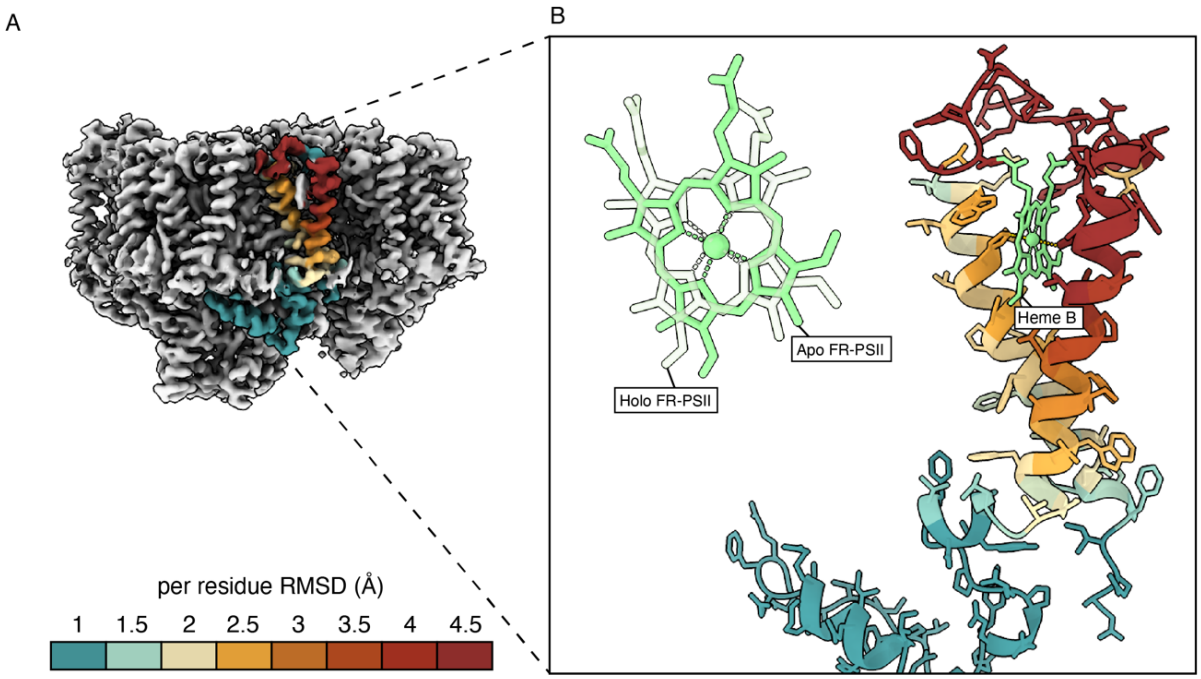


Fig. 5.14 – Alterations to Cyt b_{559} in Apo-FR-PSII

A) Model representing the per residue RMSD in the comparison between Holo- and Apo-FR-PSI. Per residue RMSD is represented as a colour scale on the map, grey areas present an RMSD lower than 1 Å. B) Chemical environment of the Heme C (in light green) cofactor of Cyt b_{559} the movement of PsaE and PsaF influence it's position and orientation.

5.3 Discussion

5.3.1 Function of the new Subunit PsbH2'

The presence of a novel subunit of PSII in the Chl *f* containing paradigm opens questions surrounding the function of PsbH both in WL-PSII and FR-PSII, and on the structural function of PsbH2'. Sequence alignments from the genomes of FaRLiP strains suggest that these organisms can be clustered in three major phylogenetic groups. One, less represented, in which only a version of PsbH2 similar in length to the canonical PsbH from white light organisms is present (Fig. 5.11A), one in which PsbH2' is present, interacts with a "WL-sized PsbH2" and wraps on the other side of CP47 (Fig. 5.11B), and one in which PsbH2 and PsbH2' are fused with a ~10 aa linker (Fig. 5.11C), suggesting a gene fusion or splitting event between the two subunits, which hairpins CP47 as in the PsbH2+PsbH2' strains.

The absence of PsbH2' from the genome of some FaRLiP organisms indicates this subunit is not essential for the normal functionality of FR-PSII. Nonetheless, its presence in most FaRLiP species indicates that this subunit might be functionally relevant in specific photo-niches. Since there is no strong environmental correlation between the ecological niches of the FaRLiP strains in the three phylogenetic groups and the function of this "extension" of PsbH cannot be inferred from structural data, a point-mutation-driven approach should be adopted to try to reveal its function. However, the position of this subunit, together with the evidence that Chl *f* occupies the B617 site, is suggestive of a potential function.

As discussed previously, Chl *f* molecules seem to have higher affinity for polar residues or water molecules as ligand for the central Mg^{2+} ion compared to histidine due to the electron-withdrawing effect of the C2 formyl oxygen (93). Since B617 presents PsbB2 His^{114} as a ligand, the Chl *f* present at this site is expected to have a lower affinity compared to the other Chl *f* sites, along with potential changes in its chemical and spectral properties (Fig. 5.9A).

This observation is strengthened comparing the system to FR-PSI, in which Chl *f* B30, also histidine-bound, is absent in the structures available that present only partial occupancy

of PsaJ2 (116), as well as the FR-PSII structure from Gisriel at al. (130) in which the loss of PsbH2 corresponds to a loss of Chl B617. In light of this evidence, it is reasonable to think that the presence of PsbH2', or a longer version of PsbH2 could be important for the stability of this site, by sterically preventing Chl *f* B617 from leaving the site due to lower affinity.

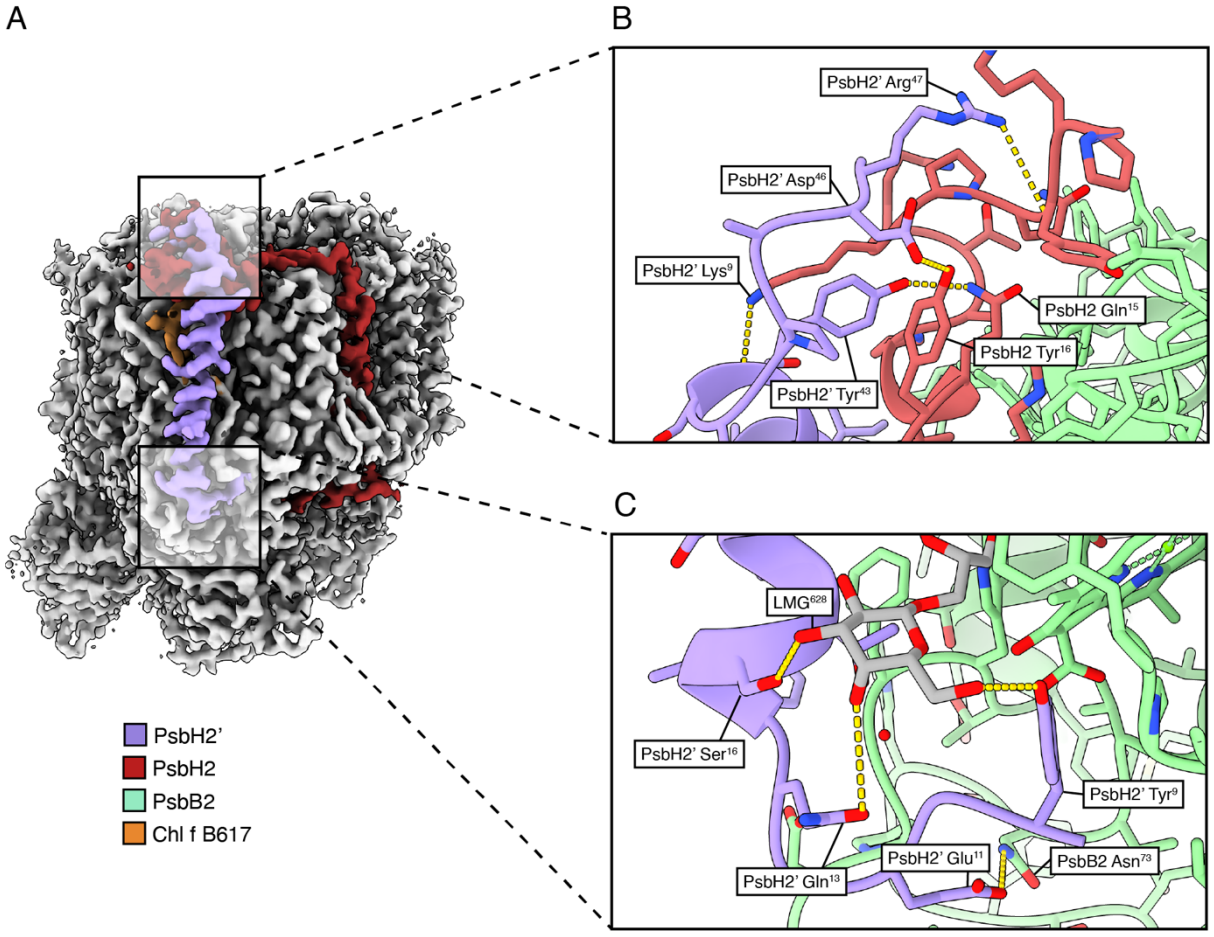


Fig. 5.15 – Potential function of PsbH2'

A) Surface representation of the position of PsbH2' in transparent purple and PsbH2 in red relative to FR-PSII in white, the Chl *f* candidate B617 in green and a nearby carotenoid in orange. The tail of B617 is involved in hydrophobic interactions with PsbH2' and is locked in position close to the carotenoid. With is present also in WL-PSII. B) Chemical environment of the stromal connection between PsbH2 and PsbH2'. C) Lumenal connection of PsbH2' with PsbB2 and LMG.

Matching far-red conserved changes are present along the sequences of PsbH2 and PsbH2' in all three phylogenetic groups, and they provide strong indication of the coevolution of the two proteins, as well as interactions that allow PsbH2' to bind only when PsbH2 is

present. The conserved changes seem to be relevant specifically in the interaction of PsbH2' with PsbH2 both in the stromal and lumenal portion of the protein (Fig. 5.15A).

On the stromal face of FR-PSII, the C-terminal portion of PsbH2' and long PsbH2 presents some fully conserved residues which create a far-red light specific interaction with the N-terminal portion of PsbH2. PsbH2' Tyr^{43} interacts with the side chain of the fully conserved residue PsbH2 Gln^{15}, while Thr^{42} interacts with a high occupancy water molecule that in turn hydrogen bonds to the backbone nitrogen of PsbH2 Gln^{15}. PsbH2' Asp^{46} hydrogen bonds another far-red conserved change in PsbH2 Tyr^{16}, which is always an Asp in WL-PSII. The fully conserved residue PsbH2 Lys^{9} (Ala in WL-PSII) reaches for the backbone carbonyl of PsbH2' Ser^{40} offering a hydrogen bond to stabilise the binding (Fig. 5.15B).

On the stromal face of FR-PSII, another set of conserved residues allows the far-red specific interaction of PsbH2, PsbH2' and a distearoyl-monogalactosyl-diglyceride (PsbB2 LMG^{628}). In most structures available of WL-PSII, this lipid is characterised by a poor ESP density that does not allow its identification. In the map presented here, there is clear ESP density for a head group, which, together with the presence of specific conserved residues in PsbH2' that stabilise it, allows its characterisation. Among the changes mentioned, PsbH2' Ser^{16} coordinates O3 of PsbB2 LMG^{628}, PsbH2' Tyr^{9} coordinates O5, while PsbH2' Gln^{13} (mostly conserved, but sometimes Ser, Asn or Glu) coordinates O4. Moreover, PsbH2' Glu^{11} is hydrogen bonding to CP47 through the far-red specific PsbB2 Asn^{73} (a Gly in WL-PSII) (Fig. 5.15C).

The presence far-red specific amino acids to coordinate a lipid may suggest that PsbB2 LMG^{628} is present only in FR-PSII in this position. The structural significance of this lipid could be related to the presence of a Chl *f* in the site B617. While the position of the phytol tail of Chl *a* B617 is identifiable in WL-PSII maps, in all maps available of FR-PSII the density of the tail is poor indicating either its absence or a higher degree of disorder. PsbB2 LMG^{628} occupies the space usually occupied by possibly disrupting B617 as a Chl *a* binding site. While there is no evidence that the mechanism of insertion of some Chl *f* is related to the length or position of their tails, it is possible that the presence of PsbB2 LMG^{628} might

lower the affinity of the site for Chl *a* and therefore increase the likelihood of the site in successfully binding a Chl *f*.

This mechanism could be particularly important given the unfavourable ligand present in this site (PsbB2 His^{114}). Other examples of altered positions for chlorophyll tails can be found in FR-PSI. Specifically, the tails of Chl B37 and B38, which form the extremely long-wavelength Chl *f* dimer, are inverted compared to WL-PSI. Nonetheless, whether the position of the tail influences or guides the binding of Chl *f* to specific sites or is just a byproduct of Chl *f* synthesis is a question that remains open.

5.3.2 Chl f B617 in comparison with WL-PSII

The assignment of the B617 site as one of the four Chl *f* sites in FR-PSII is of particular interest, since this site is considered to be the location of the "red-most" chlorophyll in WL-PSII (45,198,199). All of the core subunits of PSII show a high level of sequence conservation and consequently the complexes present characteristic features in terms of their absorption and fluorescence emission spectra. In WL-PSII, CP47 presents a red-shifted fluorescence emission shoulder at low temperature, suggesting the presence of a lower energy trap compared to the manifold of chlorophylls in the subunit. While it is confirmed that the absorption and emission feature correspond to the oscillator strength of a single Chl *a*, the location of the "red-most" Chl *a* site, as well as its function, were still a matter of debate (203,204).

A significant body of research suggests that the chlorophyll responsible for the red-shifted state is located at the interface between CP47 and PsbH (45,198,199,205). Even though this protein does not coordinate the magnesium of any Chl, structural data shows that a threonine in the stromal region of canonical PsbH (PsbH Thr^{5}) forms a hydrogen bond with the 13^{1}-keto group of Chl *a* B617. In strains with mutations affecting either the coordination of Chl *a* B617's Mg^{2+} ligand or knocking out PsbH entirely, the red-shifted spectral feature at 693nm is lost.

Nonetheless, a significant portion of the available literature still does not consider Chl B617 to be the long-wavelength emitter, proposing that major structural rearrangements

in the mutants, such as the loss of a chlorophyll or an entire subunit, could lead to long range effect on other Chl sites. While the fluorescence emission at 693 nm, as well as the hydrogen bonding threonine, are conserved in all of the canonical Chl *a* containing photosystems, all of the PsbH2 sequences available lack this hydrogen bond (73). This suggests that the site energy of this pigment is re-tuned in Chl *f* containing PSII.

For this reason, the same single amino acid change, Thr to Ala, in the terminal portion of PsbH was reproduced in *C. reinhardtii*. The mutant presents the loss of the red-shifted shoulder at 693 nm, indicating the effect of the coordination at the 13^1-keto group of Chl *a* B617 is in line with other experimental results and responsible for the reduction in site energy in Chl sites (149). One of the proposed functions for this site is that Chl *a* B617 can act as a form of photoprotection for WL-PSII when centres are closed (Q_a reduced) (45). In this safety-valve-like mechanism, when Q_a^- is present, due to changes in rates, a large part of the excitation energy of the RC returns to the CP43 and CP47 core antennae, where Chl *a* B617 can act as a low energy trap and where the triplet can be safely quenched by nearby carotenoids.

The absence of this photoprotection mechanism in Chl *f* containing FR-PSII can be rationalised by comparing the energetic landscapes of pigments in the three PSII paradigms. In WL-PSII, the site energy landscape of antenna and reaction centre is mostly flat, and excitation can be easily funnelled into a mildly red-shifted trap when excess excitation is present. On the other hand, in Chl *f* FR-PSII the manifold of higher energy pigments in the antenna is dotted with sparse, extremely red-shifted sites. In this configuration, Chl *f* B617 cannot retain this function by absorbing at ~690 nm, since excitation energy would not be directed to it to be quenched. Therefore, FR-PSII was evolutionarily forced to reconfigure, tuning the site energy of B617 and potentially adopting different photoprotection mechanisms.

This hypothesis is strengthened if *Acharyochloris marina*, where every Chl in PSII is replaced with Chl *d*, is taken as a comparison. In this organism, PsbH presents an hydrogen bonding threonine to the 13^1-keto group of Chl *d* B617, likely with similar effects on its side energy as the ones in Chl *a* PSII. In a longer wavelength system, but one in

which most pigments retain comparable site energies, the mechanism proposed by Renger and collaborators (45) is again a viable photoprotective option, strengthening the hypotheses on the function of the “red-most” chlorophyll in PSII.

5.3.3 Nature of the reaction centre long-wavelength chlorophyll: Chl d or Chl f?

In FR-PSII, the primary donor (Chl_{D1}) was initially identified as a long-wavelength chlorophyll, based on multiple biophysical lines of evidence (43). First, from an energy levels perspective, the enhancement in the intensity of the thermoluminescence emission peak, indicated a decrease in the energy gap between the charge separated and excited states of the reaction centre, consistent with a lower-energy, longer-wavelength primary donor/emitter. Second, from a spectroscopic point of view, the changes in the $P680^+$ *minus* P680 difference spectra indicated the disappearance of a blue shift assigned to Chl_{D1}’s Qy absorption at ~680 nm, together with the appearance of a marked blue shift in a pigment absorbing around 720 nm, unequivocally identifying a long-wavelength pigment in position Chl_{D1}.

From a phylogenetic point of view, the conserved amino acid changes clearly showed two possible hydrogen-bonding groups, which could potentially bind either the formyl group from Chl *f* in the C2 position or to the formyl group of Chl *d* in the C3 position of Chl_{D1} (43). Spectroscopically, the absorption peak of Chl_{D1} was assigned at 727 nm, making it the second highest energy pigment among the long-wavelength Chl molecules of FR-PSII. Consequently, since Chl *d* is generally considered to be more blueshifted than Chl *f*, and the wavelength (727 nm) of the primary donor was similar to that of Chl *d* (725 nm) in this location in *A. marina*, the working hypothesis was that the Chl_{D1} site harboured Chl *d*.

Nonetheless, given the presence of both Chl *d* and Chl *f* in FR-PSII, the exact nature of the long-wavelength pigment in the PSII reaction centre could not be addressed at the time. Previous structures of FR-PSII and more recently computational approaches have provided new insight on this, which agrees in the assignment of a Chl *d* primary donor, but without the precise hydrogen-bonding network that stabilises the reaction centre Chl *d* (Gisriel et al., 2022c, 2023c, 2024a, 2022a).

In the map presented, the improved resolution together with the cone scan analysis, allows the presence of a Chl *f* in position Chl_{D1} to be ruled out, since the substituent does not show any increased ESP around the C2 position (Fig. 5.16). Moreover, the increased ESP is clearly oriented towards PsbA3 Thr^{155} rather than towards PsbA3 Tyr^{120}, suggesting the potential hydrogen bond network that stabilises Chl *d* in this position.

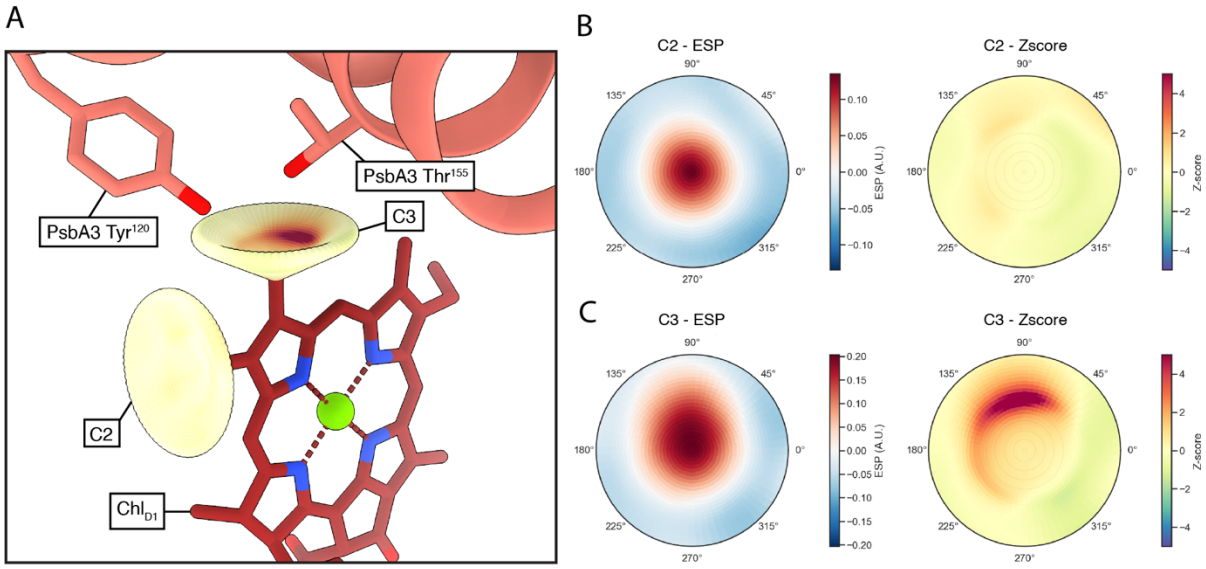


Fig. 5.16 – Cone scan analysis of the substituents of Chl_{D1}

A) Chemical environment of Chl_{D1} in FR-PSI. The PsbA3 backbone is represented in pink and Chl_{D1} is represented in red. The cone scans in the positions C2 and C3 of the ring are coloured according to their ESP Z-score. B) Cone scans of the raw ESP and Z-score of the C2 substituent of Chl_{D1} C) Cone scans of the raw ESP and Z-score of the C3 substituent of Chl_{D1}.

Moreover, the absence of conserved phylogenetic changes allows ruling out the experimental hypothesis that indicates that the long-wavelength chlorophyll in the reaction centre could be located at the P_{D2} site rather than at the Chl_{D1} site. None of the other reaction centre chlorophyll present an increased ESP at the C2 substituent, and there are no amino acids capable of stabilizing a putative formyl group at the position C3 of Chl P_{D2}. The amount of evidence for Chl *d* being present in position Chl_{D1} allows a confident assignment of the long-wavelength pigment to this position.

5.3.4 Structural changes on the donor and acceptor side of FR-PSII

On the donor side, while water splitting is expected to be essentially unchanged in FR-PSII compared with WL-PSII, two conserved changes are present around the manganese

cluster in the FR-D1 subunit. The changes, PsbA Ile^{184} to PsbA2 Phe^{185} and PsbA Met^{328} to PsbA2 Phe^{329}, insert two Phe that could interact by π-stacking and potentially shift the position of the N-terminus of FR-D1. This could potentially have effects by altering the distance between the residues that coordinate the Mn cluster or by affecting the rigidity of the complex. The resolution of the present map, together with the presence of beam-induced reduction of the cluster, does not allow accurate assessment of distances.

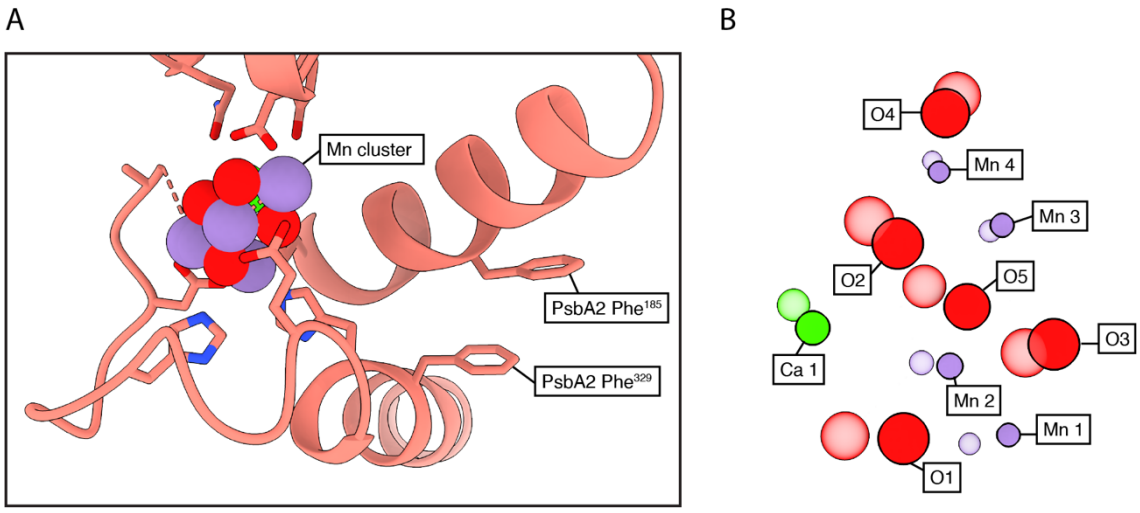


Fig. 5.17 – FR-PSII donor side conserved changes

A) Protein environment of the Mn cluster, D1 (PsbA2) is represented in pink, B) Changes in position of the atoms of the manganese cluster in the structure presented compared with 9EVX. The atoms of the FR-PSII structure are represented in full colour, while those of 9EVX is transparent.

Concerning the acceptor side of FR-PSII, recent mechanistic studies demonstrated how the redox potentials of some of the chemical species involved are altered, with the aim to maintain appropriate standard free energy gaps and avoid backreactions and the consequent damage. Despite the ~100 meV loss in energy that a Chl *d* primary donor entails, Chl *f* FR-PSII shows no increased singlet oxygen production and no increased sensitivity to high light compared with WL-PSII (98).

Specific far-red phylogenetically conserved changes are present around the acceptor side in FR-PSII, which are suggested to alter the redox potential of both the Phe/Phe^{-} couple and the Q_a/Q_a^{-} couple. In the case of Phe/Phe^{-}, the redox potential is expected to be increased by 15-30 meV by the presence of a mutation characteristic of the high light variants of PSII, a single point mutation in position 130 of PsbA, substituting a less

electronegative side chain of PsbA Gln^{130} with a Glu (206,207). The increase in potential of the Phe/Phe^{-} couple would only partially compensate for the decrease in energy of the excited state of the reaction centre Chl *d*, suggesting that the potential of the Q_a/Q_a^{-} couple has also to be altered.

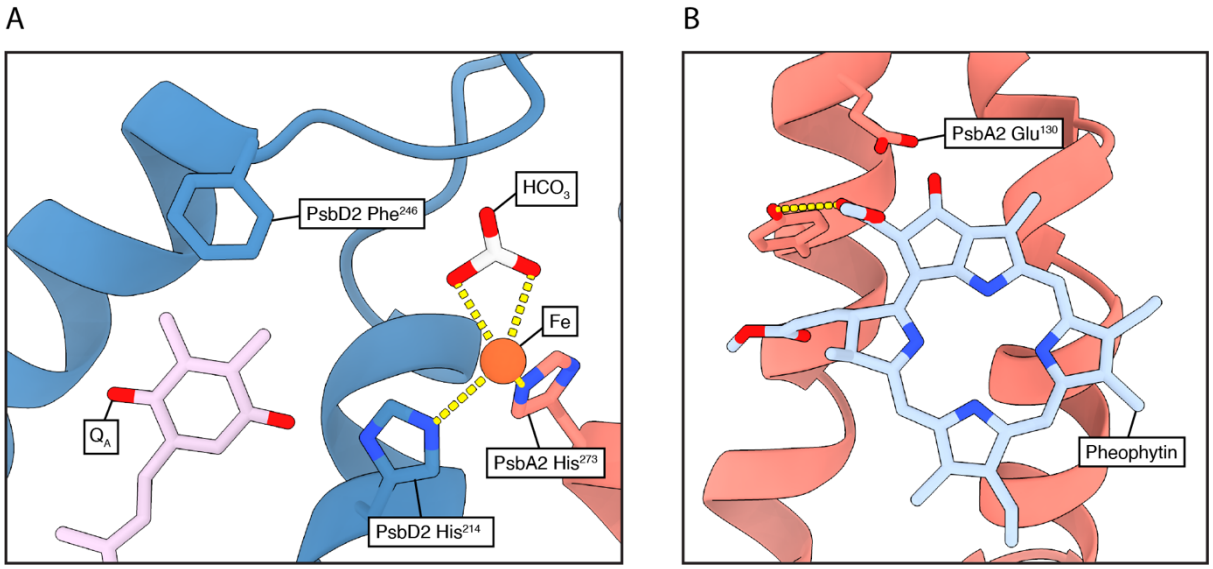


Fig. 5.18 – FR-PSII acceptor side conserved changes

A) Protein environment of Q_a, D2 (PsbD2) is represented in blue, Q_a in pink D1 (PsbA2) is presented in pink B) Protein environment of the pheophytin on the D1 side. Pheophytin is represented in light blue and D1 (PsbA2) is presented in orange.

Given the differences in the back-reaction in FR-PSII, which has a more stable charge separated state P^{+}/Q_A^{-} during S_2 (Viola et al., 2022), conserved changes in the area of the primary quinone that might be responsible for redox tuning must be identified to find potential targets for mutagenesis. Among the changes that could influence the redox potential of Q_A, a far-red single point mutation is present, PsbD Met^{246} is substituted with PsbD Phe^{246}. The presence of the phenyl ring's π system in proximity of the quinone ring could alter its potential by π-stacking.

5.3.5 Changes in the side path and ChlZ

Under conditions in which PSII is inhibited and can't transfer electrons from the oxygen-evolving complex to Q_B with high efficiency, the system incurs in the risk of generating dangerous Chl triplet states, which can convert to singlet oxygen and generate

photodamage. Under oxidoreductive stresses, such as when no electron acceptors are present in the plastoquinone pool or the manganese cluster is damaged, $P680^+$ can be re-reduced through an electron transfer event known as the side path. In these circumstances, its closest carotenoid, known as Car_{D2}, can slowly donate electrons to $P680^+$ and then be regenerated by one of two redox-active species present in PSII, Cyt b_{559} or $ChlZ_{D2}$. The biological function of the side path is not completely understood and has been hypothesised to be related to either photoprotection, by avoiding the formation of triplet states by regenerating P680 under stress condition, or the generation of a chlorophyll cation that can quench incoming excitation, limiting charge separation events altogether acting as a trap.

In FR-PSII, the side path seems to behave differently from WL-PSII. Preliminary electron paramagnetic resonance (EPR) data (unpublished) collected on Holo-FR-PSII indicate three modifications may be responsible: changes related to Car_{D2}, modifications of the P_{D2} reaction centre chlorophyll or changes to the complex redox behaviour of the Cyt b_{559}, the electron donor to the carotene, is modified.

For this reason, potential far-red conserved changes around the side path have been investigated structurally. No major changes are present around Cyt b_{559}, but two conserved changes are present around Car_{D2} and $ChlZ_{D2}$: a leucine to PsbD Met^{43} and a leucine to PsbD Phe^{49}. The role of these two modifications is not clear and cannot be inferred from the EPR data, but the presence of an electronegative amino acid such as PsbD Met^{43} located between Car_{D2} and $ChlZ_{D2}$ could influence the electron transfer rate between them, altering the behaviour of the side path. Finally, over the length of Car_{D2} PsbD Phe^{49} is present and could potentially influence the properties of the β-carotene by π-stacking.

In the Apo-FR-PSII in the protein, the absence of PsbJ and PsbR alter the position of the two α-helixes of Cyt b_{559} and consequently of the heme *b* bound to it, rotating it by ~15° and shifting it slightly away from Car_{D2}. Moreover, the absence of PsbR exposes the heme *b* to the external solvent possibly changing its redox potential.

5.3.6 *Energy landscape in FR-PSII dimers*

The five long-wavelength sites in FR-PSII allow it to have comparable yields to WL-PSII with a drastically reduced antenna size (79). In photon-limited conditions, the long-wavelength antenna sites have to be strategically positioned to increase the connectivity not only to the Chl *d* primary donor in the same monomer, but also to the antenna Chl *f* in the adjacent monomer, so to maximise the antenna size per reaction centre (Fig. 5.19).

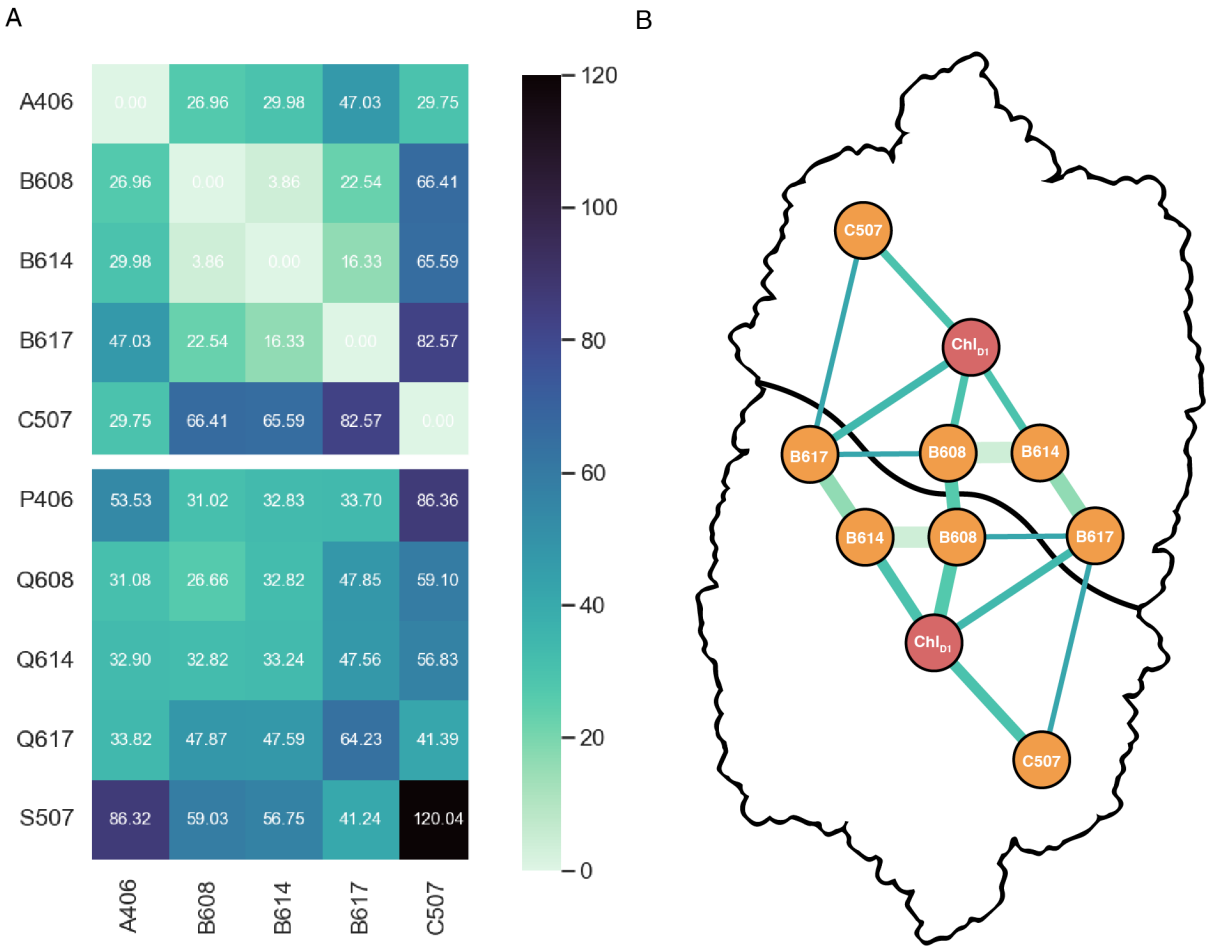


Fig. 5.19 – Long-wavelength pigments landscape of FR-PSII

A) Edge-to-edge distance matric of long-wavelength sites in a FR-PSII dimer. B) Representation of the positions of the pigments and their relative distance on a stromal view of FR-PSII, the position of the dots is similar to their structural positions but adapted to allow a better visualization of the potential excitation energy transfer pathways within the monomer and across the dimer.

The asymmetry in the long-wavelength pigments present in CP43 and CP47 (a 1 to 3 Chl *f* ratio), reflects this sharing "pigment economy" aspect of Chl *f* photosystems. In the dimeric assembly of PSII, in fact, CP43 is localised in a distal position with respect to the

dimer interface, while CP47 is one of the main components of the dimerization interface. As in FR-PSI, by placing all of the long-wavelength pigments in the core of the dimeric assembly, the system is facilitated in transferring excitation among different monomer when one of the reaction centres is closed.

Of key importance for the excitation energy transfer landscape in FR-PSII is Chl *f* B608. This Chl *f* molecule is not only the closest pigment to the reaction centre chlorophylls, but also the closest to its "mirror" site, Chl *f* B608 of the adjacent monomer, conferring it the function of excitation bridge across the centres. Chl *f* B614 is in extremely close proximity of Chl *f* B608 (<4 Å), creating an enriched core of chlorophylls and connecting it to Chl *f* B617 (Fig. 5.19A). Moreover, Chl *f* B614 and Chl *f* B617 are also close with an edge-to-edge distance of ~16 Å. These two pigments are also structurally linked by a series of far-red light conserved changes. The C2 formyl oxygen of Chl *f* B617 is hydrogen bonded to PsbB2 Glu^{107}, that, in turn, is a hydrogen bond acceptor by the amino group of PsbB2 Asn^{27}. Finally, this residue is connected via a water molecule to PsbB2 His^{26}, the axial ligand to Chl *f* B614. This long range interaction could potentially be involved in fine tuning the electrostatic environment of these two pigments and consequently their excitation energy transfer dynamics.

Among the long-wavelength pigments present in FR-PSII, Chl *f* C507 and Chl *f* B617 are the two closest pigments to the stromal surface of PSII, making them the two candidates as the injection point of phycobilisome excitation in FR-PSII. To gain insight into which long-wavelength pigment is the excitation energy inlet point from the FR-APC to FR-PSII, Mascoli and colleagues (95) calculated the hopping times between the candidate Chl *f* sites available and the reaction centre Chl *d* in FR-PSII using the Förster approximation and assigned the fastest excitation energy transfer to C507 (0.26 ns), followed by B608 (0.45 ns), B614 (0.99 ns) and finally B605 (1.49 ns). The calculation, together with time resolved fluorescence data, and the absence of a good bridging candidate Chl *f* in CP47, led to the proposition of Chl *f* C507 as the sole bridging pigment. Nonetheless, some of the assumptions made in the reported excitation energy transfer calculation are too simplistic. Specifically: i) The hopping times are calculated for FR-PSII monomers using the point dipole approximation; ii) hopping times were calculated assuming each antenna Chl *f* to

emit at 740 nm; iii) the excitation energy transfer rates of the three Chl *f* in CP47 were averaged rather than taking into account equilibration phenomena among nearby Chl *f* pigments.

For these reasons, and considering the data reported in this chapter, the current model can be challenged, offering Chl *f* B617 as an equally good bridging pigment in CP47, also considering that it is in close connection with 6 out of the 8 antenna Chl *f* in a FR-PSII dimer. Given the lack of an available structure of a FR-PSII - FR-APC supercomplex, as well as site directed mutagenesis, it is currently not possible to determine with certainty in what proportion Chl *f* B617 and Chl *f* C507 contribute to bridging excitation energy between FR-APC and FR-PSII, since they would be almost equidistant from the terminal emitter bilin of FR-APC.

5.4 Conclusions

The structures of FR-PSII presented in this chapter provide new evidence for the location of the long-wavelength pigments in the complex, confirming previous assignments available in the literature and supporting the presence of a previously not considered site. The assignment of a Chl *f* to the site B617, which was confirmed in this chapter to be responsible for the 693 nm 77 K fluorescence emission in Wl-PSII, opens questions regarding the photoprotection mechanisms of FR-PSII compared to Chl *a*- and Chl *d*-based PSII. This once again highlights the importance of analysing the three photosynthetic paradigms individually, to uncover their distinct evolutionary advantages and limitations. The two paradigms of long-wavelength photosynthesis, in particular, represent divergent evolutionary strategies for adapting to suboptimal light conditions. This finding also advances the understanding of the WL-PSII site energy landscape, providing definitive evidence for PsbH Thr^{5}, which binds the 13^{1}-keto group of Chl *a* B617, is responsible for the red-shift in absorbance of this pigment. A comparative mechanistic study of FR- and WL-PSII could reveal the function of this site and the mechanism by which excess excitation is dissipated under stress conditions.

In the maps presented in this chapter, a novel, FaRLiP cluster encoded subunit, specific to FR-PSII has been discovered and characterised structurally and phylogenetically. Possible functions have been proposed, including protection of the Chl *f* B617 site, connection to the FR-APC or prevention of the association with WL-PBS. Moreover, a rationale for potential excitation-bridging Chl *f* sites is presented, with Chl *f* B617 and Chl *f* C507 proposed as the most likely candidates.

This work provides the framework for site-directed mutagenesis with the aim of investigating the costs of Chl *f*-based photosynthesis, the site energy of putative Chl *f* sites and the function of the novel PsbH2' subunit. Now that the structures of both Chl *f* photosystems have been solved and the long-wavelength sites assigned, comparisons between the two long-wavelength paradigms of oxygenic photosynthesis (i.e., Chl *f* and Chl *d* containing photosystems) revealing evolutionary trade-offs and different adaptation strategies to far-red enriched, low light environments.

6 Cryo-EM structure of FR-APC and evolutionary adaptations to far-red light

6.1 Introduction

During the expression of the ~21 genes of the FaRLiP cluster, the photosynthetic apparatus of far-red–adapting cyanobacteria is completely reconfigured, allowing survival in unfavourable light conditions, where photons between 400 nm and 700 nm are scarce (106). To harvest and perform photochemistry with light beyond 700 nm, FaRLiP-capable cyanobacteria alter all the main light-driven components of the photosynthetic apparatus. While the far-red light specific paralogues of PSI and PSII allow a fraction of longer wavelength pigments, such as Chl *d* and Chl *f* (~8% to 10%), to be inserted into the photosystems (43,106,109), the far-red-light-adapted subunits of the phycobilisomes, the major light-harvesting antenna proteins in cyanobacteria, still rely on the same pigment for light harvesting (phycocyanobilin), and red-shift their absorption spectra by modifying the protein environment around the chromophores (179,180).

Phycobilisomes are a structurally and functionally diverse family of light-harvesting complexes that bind open-chain tetrapyrroles known as bilins to absorb photons and funnel excitation energy to the photosystems, mainly PSII (50). The remarkable variability in the size, arrangement, and absorbance profiles of these antenna complexes deeply reflects their evolutionary adaptation to diverse ecological niches (50).

Given the small number of long-wavelength chlorophylls present in far-red-adapted photosystems, a red-shifted antenna, capable of efficiently funnelling the excitation energy of far-red photons into the far-red photosystems, is necessary to increase the antenna size for wavelengths above 700 nm. Most FaRLiP gene clusters contain five phycobiliprotein paralogues (ApcB2, ApcD5, ApcD3, ApcD2, and ApcE2) that assemble with other subunits common to white light phycobilisomes (ApcF and ApcC) into a bicylindrical complex (181), which is significantly smaller than other known classes of phycobilisomes (50). However, some species possess a paralogue of ApcF, named ApcF2, that has been

suggested to replace ApcF during far-red light photoacclimation (182,183). For some species, including *C. thermalis* PCC 7203, the gene encoding ApcF2 (*apcF2*) is located outside the FaRLiP cluster.

Similar to white light phycobilisome (WL-PBS) cores, far-red allophycocyanin (FR-APC) features α- (ApcD2, ApcD3, ApcD5) and β- (ApcB2) subunits that form (αβ) heterodimers, which assemble into toroidal hexameric $(\alpha\beta)_3$ complexes (50,53,184). $(\alpha\beta)_3$ are stacked with the help of ApcC onto a specialised core-membrane linker protein, ApcE2, which also anchors the complex to the stromal side of PSII (185). The chromophores in the far-red β-subunits (ApcB2) appear to be structurally similar to their white light counterparts (58), and accordingly exhibit similar absorption properties. In contrast, the bilin pigments in the α-subunits show a more coplanar arrangement of pyrrole ring A with the rest of the tetrapyrrole, resulting in a red-shifted absorption relative to WL α-subunits, a feature also observed in heterologously expressed phycobiliproteins (179,181).

In the bicylindrical FR-APC, ApcE2 and ApcD3 have been suggested to act as terminal emitters. These two subunits lack a cysteine in proximity of the bilin site that would covalently bind the phycocyanobilin chromophore. As a result, they maintain a fully extended π-system, causing a red-shift in their absorption compared to their WL paralogues (87,186,187). However, approximately 27% of ApcD3 sequences retain the bilin-binding Cys^{78} residue, suggesting a degree of species specificity of the excitation energy transfer pathways to the photosystems. A previously determined Cryo-EM map of a disconnected portion of the FR-APC core complex from *Synechococcus* PCC 7335 contained only three of the four expected phycobiliprotein trimers (58).

In this chapter, we attempted the purification of a FR-APC - FR-PSII supercomplex and obtained a structure of an intact bicylindrical FR-APC at a resolution of 2.51 Å. A low-resolution ESP density, localised where PSII is expected to bind, is present. The bicylindrical FR-APC complex confirms the composition of the intact rod and presents an additional far-red-light-adapted beta subunit, ApcF2. The structure provides insight into which pigments are the terminal emitters, as well as the dimerisation interface between the two cylinder cores. Moreover, an analysis of the planarity of bilin pigments and their

chemical environment in relation to their absorption suggests specific evolutionary traits needed to optimise far-red light excitation energy transfer.

Finally, the possible mesoscale organisation of the photosynthetic apparatus during far-red light photoacclimation is discussed, based on functional considerations and insight provided by *in situ* Cryo-EM maps of WL-PSII associated with WL-PBS recently released on EMDB (188). The generated model also provides insight into the docking site of the FR-APC on FR-PSII's surface and helps elucidate excitation energy transfer routes from the FR-APC to the long-wavelength reaction centre of FR-PSII.

6.2 Results

The putative FR-PSII + bicylindrical FR-APC photosynthetic complex was isolated as described in the Materials and Methods section. Approximately 9000 micrographs were collected. Through multiple rounds of particle picking, 2D classification and *ab initio* modelling, a low-resolution map was generated. This map was used to assess the intactness of the complex and confirm its symmetry. The particle stack obtained was refined using multiple rounds of heterogeneous refinement and 3D classification.

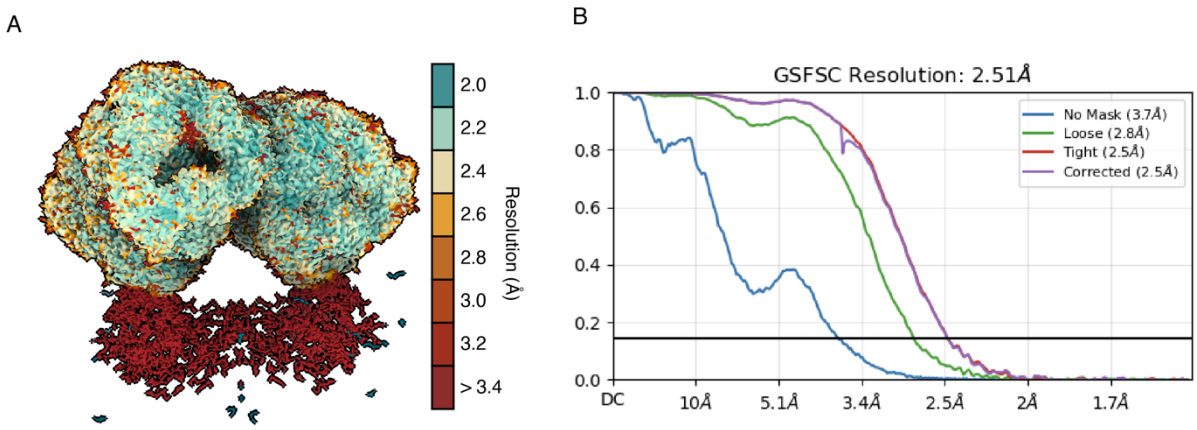


Fig. 6.1 – Local resolution of bicylindrical FR-APC map

Visual representation and GS-FSC of the structure discussed in this chapter, the colour map is provided on the right. A) Intact bicylindrical FR-APC map coloured by local resolution, a low-resolution density, probably originating from a heterogeneous photosynthetic complex population, can be noticed in red B) GS-FSC of the bicylindrical FR-APC

The Cryo-EM ESP map was generated through NU-refinement at a global resolution of 2.51 Å (Fig. 6.1Fig. 4.1B) and presents a low-resolution ESP density (> 10 Å) in proximity to the membrane linker protein ApcE2, where FR-PSII is expected to bind (Fig. 6.1A). Further local refinement on the low-resolution section of the map did not yield any improvements or clarity on the relative orientation of the photosynthetic complex to the bicylindrical FR-APC. The map presents density to fit two slightly tilted, antiparallel far-red-light-adapted allophycocyanin cores and clear ESP density to model 50 protein chains: 18 ApcD5, 22 ApcB2, 2 ApcD3, 2 ApcD2, 2 ApcE2, 2 ApcF2 and 2 ApcC subunits (Fig. 6.3A). Each chain except the two ApcC subunits coordinates a phycocyanobilin molecule for a total of 48 chromophores in each bicylindrical FR-APC complex (Fig. 6.3B).

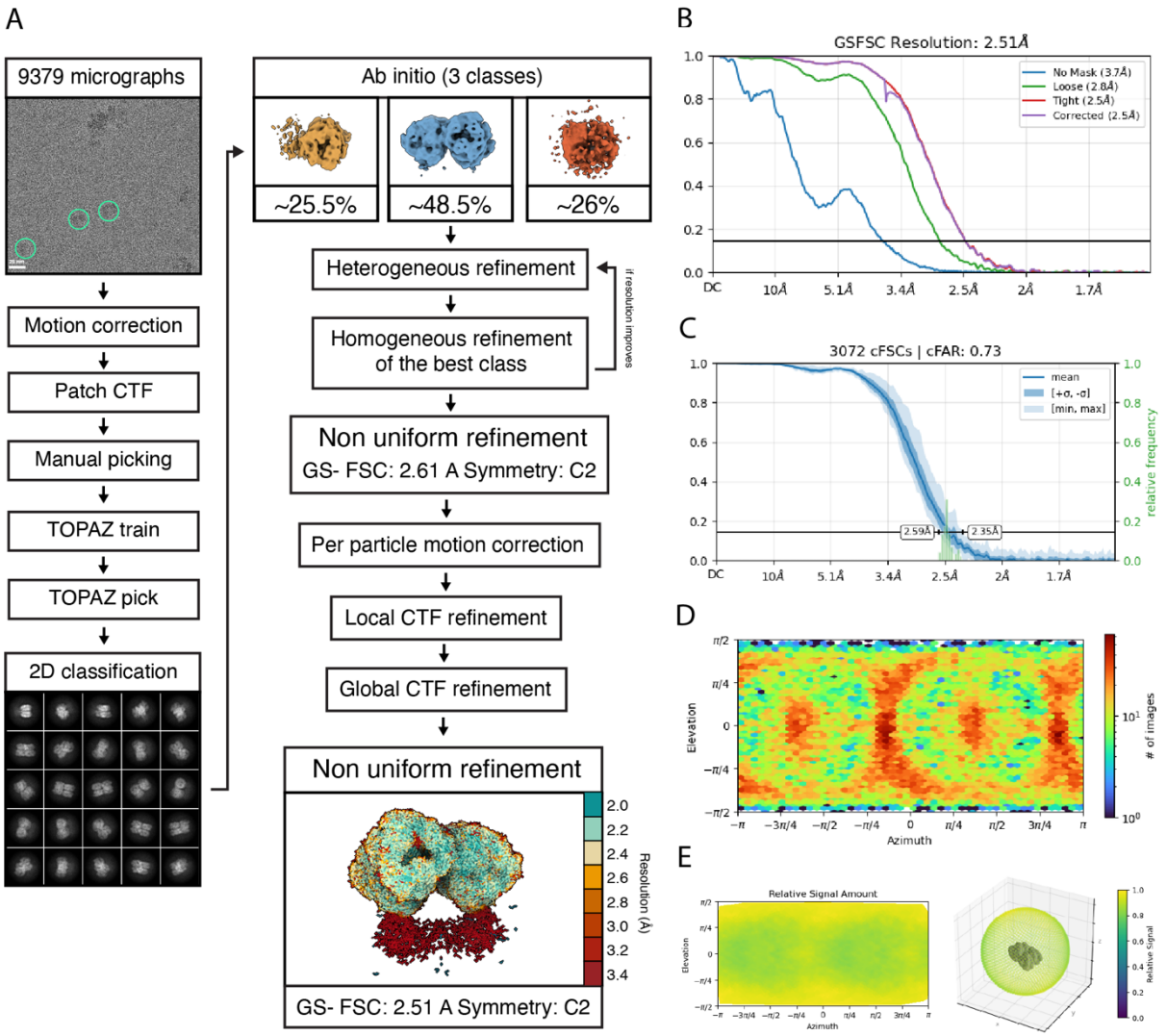


Fig. 6.2 – Processing workflow and reconstruction diagnostic plots for FR-APC (9IR1)

A) processing workflow to obtain the final reconstruction of the FR-APC complex. B) GS-FSC curve for the FR-APC complex. C) cFSC curves for the FR-APC complex. D) Viewing direction distribution of the particle stack. E) relative signal amount for each of the viewing directions.

Data collection

Microscope	Krios III
Camera	Falcon 4i
Magnification	155000x
Voltage (kV)	300
Electron exposure (e-/Å^2)	40
Defocus range (μm)	-0.8 to -2
Pixel size (Å)	0.723
Energy filter	Selectris (20 eV)
Exposures	9397
Image format	EER

Data processing

Box size	700 px
Initial particles (no.)	423204
Final particles (no.)	36300
Symmetry	C2
Map resolution (Å)	2.51
Map sharpening *B* factor	-41.9

Model Refinement

Refinement package	PHENIX
Initial model used	8UHE
Real/reciprocal space	Real Space
Resolution cutoff	2.60

Model Validation

MolProbity score	1.29
ClashScore	5.35
Bond length R.M.S.D. (Å)	0.002
Bond angles (°)	0.555
Poor rotamers	0.24%
Favored rotamers	93.48%
Ramachandran outliers	0.05%
Ramachandran favored	98.67%

Table 7 – Data collection and processing parameters for FR-APC complex (9I1R)

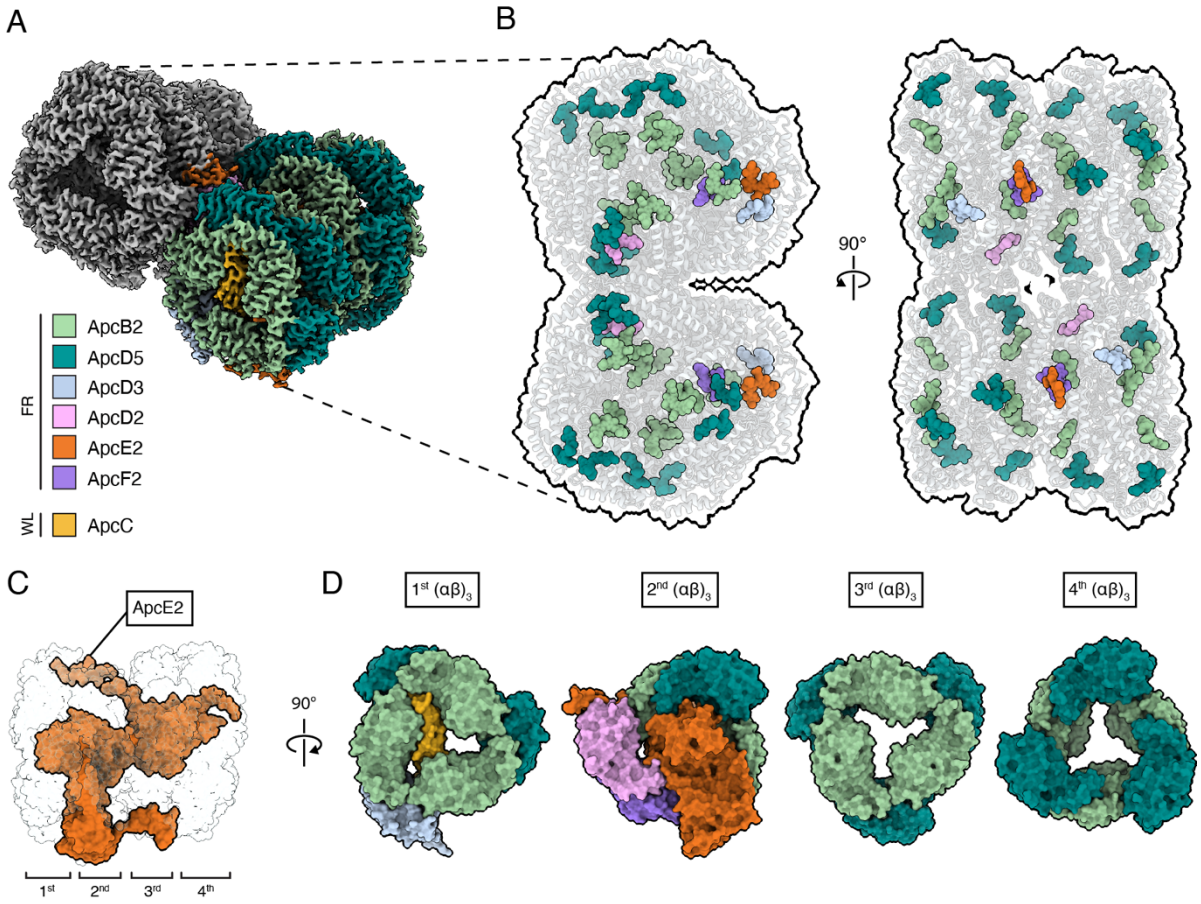


Fig. 6.3 – Positions of the bilin pigments in the FR-APC

A) View of the dimeric FR-APC map from a tilted perspective from the cytoplasmic side. The map presents clear ESP for all the known FR-APC subunits, that are coloured according to the colour key at the bottom left of the panel. B) Location of the phycocyanobilin pigments in a bicylindrical FR-APC complex as viewed from the membrane plane and from the stromal side of the membrane. The pigments are also coloured according to the colour key in panel A. C) Side view of a single FR-APC cylinder with ApcE2 coloured in orange and every other subunit transparent D) Separate representation of the four $(\alpha\beta)_3$ trimers, also coloured according to the colour key in panel A. It is important to note that the ApcE2 subunit represented together with the second trimer.

Each of the cylinders contains four $(\alpha\beta)_3$ trimers, arranged in two face-to-face $(\alpha\beta)_6$ hexamers. The proximal hexamer (relative to the membrane) contains the far-red-light-specific subunits ApcB2, ApcD2, ApcD3, ApcD5, ApcE2, ApcF2, and the white light-shared subunits ApcC (Fig. 6.3D). As suggested previously (58), the second hexamer is assembled onto the second REP domain of ApcE2 (Fig. 6.3C) and contains only 6 ApcD5 and 6 ApcB2 subunits, but contrary to expectations is missing the distal ApcC subunit. The absence of the distal ApcC seems not to be related to loss of intactness during the isolation procedure since it would be sterically prevented from detaching by the presence of 3

ApcB2 subunits. Even among the more common subunits in the structure (i.e., ApcB2 and ApcD5) there are minor conformational variations in the position of the pigment since the binding pocket of the bilin is partially formed by the core subunits (ApcE2 or ApcC). The fine tuning of the chemical environment of each bilin binding pocket has consequences on the absorption spectrum of each of the chromophores independently. This could tune specific bilin sites to modulate the site energy landscape of the pigments so that excitation can be efficiently funnelled to the terminal emitter or to suit specific functions. The absence of the distal ApcC might be physiologically relevant to tune the distal pigments in the last $(\alpha\beta)_3$ trimer.

In the following section, a brief spectroscopic characterisation of the sample is provided. Moreover, the dimerisation interface of the bicylindrical complex is described and the bilin sites are analysed based on their chemical environment and planarity.

6.2.1 Absorption and low-temperature fluorescence spectroscopy

The absorption spectra of the FR-APC complex show the same features previously reported in the literature, with two main absorption peaks at 650 nm and 710 nm, corresponding to the β-subunits and α-subunits respectively (87,179,181). Moreover, the absorption spectrum presents, although with low intensity, a peak around 440 nm, which indicates the presence of chlorophylls and photosynthetic complexes in the sample (Fig. 6.4A). Further analysis with low-temperature fluorescence emission spectroscopy indicates that when chlorophylls are excited at 440 nm, there is a clear emission peak at 750nm, with a shoulder at 730 nm, which corresponds to the fluorescence emission of FR-PSII, with some of the excitation being transferred back to the terminal emitter of the FR-APC. When exciting the sample with 550-600 nm light, where the absorbance of the FR-APC dominates, the fluorescence emission arises mainly from the bicylindrical FR-APC at 730nm, with a smaller emission shoulder at 750 nm (Fig. 6.4B). These data indicate that the sample consists mostly of disconnected bicylindrical FR-APC, but that a portion of the complex is connected to FR-PSII.

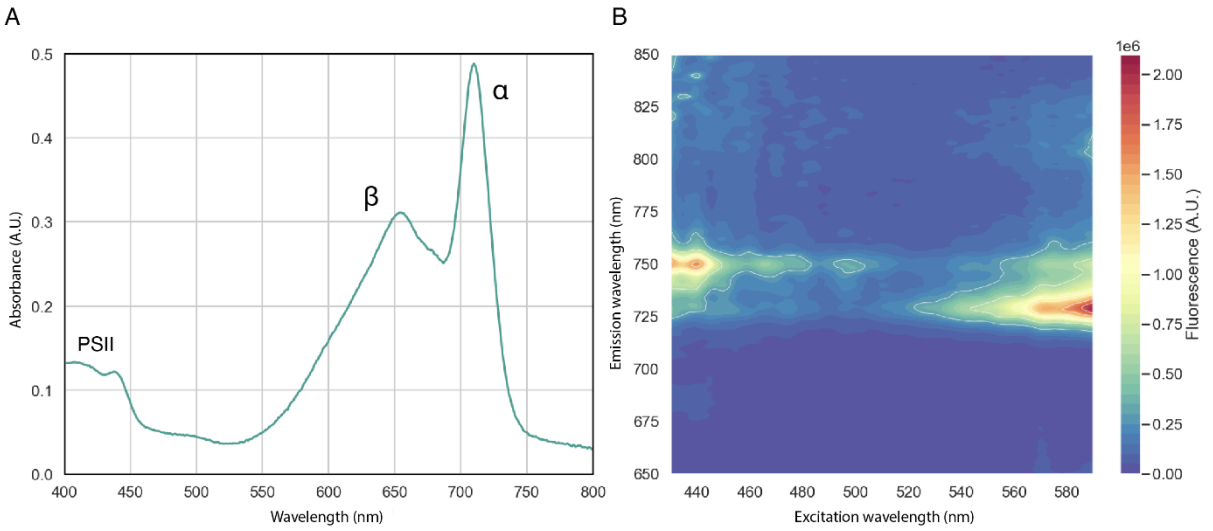


Fig. 6.4 – Absorption and fluorescence spectra of the bicylindrical FR-APC complex

A) Room temperature absorption spectrum of the bicylindrical FR-APC. The green band highlights the portion of the spectrum related to the Q_x absorption band of Chl molecules, while the blue band a portion of the bilin absorption spectra. B) 2D low-temperature fluorescence (77 K) of the bicylindrical FR-APC, excitation wavelength is represented on the X axis and emission wavelength on the Y axis, the colour gradient to interpret the intensity of the fluorescence is represented on the right.

6.2.2 Planarity of bilin sites

The planarity and chemical environment of the bilin pigments strongly influence their absorption spectra (46,189). For this reason, to gain insight into the structural origin of the red-shifted absorbance of some of the subunits of the FR-APC complex (Fig. 6.4A), as well as potential differences in the bilin sites across the different rings, the angles between ring A and ring B of each bilin molecule and those between ring C and ring D were measured and plotted (Fig. 6.5). The clusters obtained can be rationalised when considered together with the phylogenetically conserved changes around the bilin binding sites in FR paralogues of the α and β subunits. The angles between rings B and C do not vary significantly and are therefore not included in the analysis.

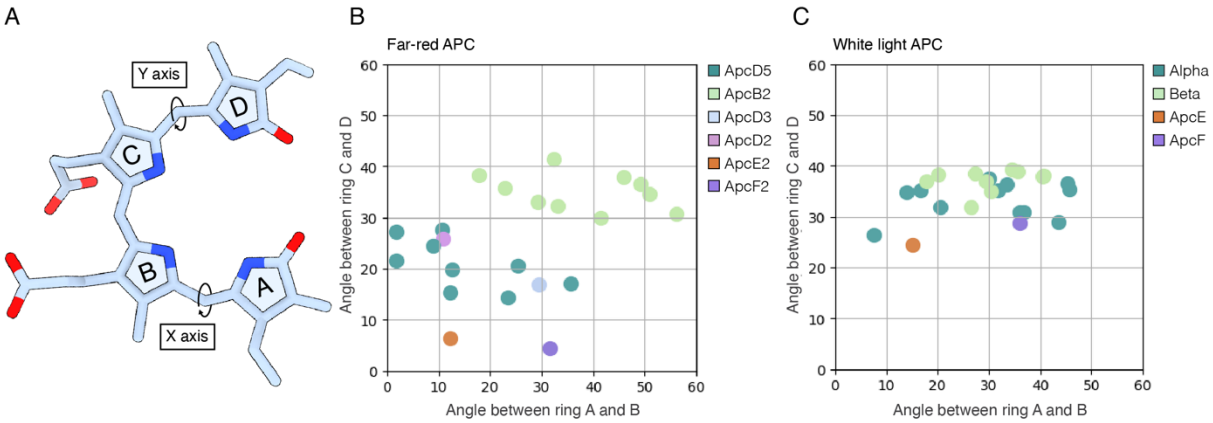


Fig. 6.5 – Analysis of the planarity of bilin pigments in FR-APC and WL-APC

A) Representation of an unbound planar phycocyanobilin pigment with the measured torsion angles indicated. B) Variation of bilin planarity in FR-APC. The x-axis reports the degrees off plane of ring A with respect to ring B, while the y-axis reports the degrees off plane of ring D with respect to ring C. C) Variation of the planarity of the bilin pigments in WL-APC from *Synechocystis sp*. PCC 6803 (PDB ID: 7SC7). The x-axis reports the degrees off plane of ring A with respect to ring B, while the y-axis reports the degrees off plane of ring D with respect to ring C.

The bilins contained in the α- and β-subunits of WL-APC (190) present very similar conformational landscapes. In these subunits, ring D is generally between ~30° and ~40° off plane, while ring A shows a larger spread ranging from ~20° to ~50° off plane (Fig. 6.5B). The bilins contained in FR-APC present a more diverse arrangement; the β-subunits (ApcB2) maintain similar torsion angles in ring D compared to their WL counterparts, while the α-subunits (i.e., ApcD2, ApcD3 and ApcD5) are far more planar, in contrast to WL subunits (Fig. 6.5C). This structural difference is reflected in the absorption spectra of the FR-APC complex (Fig. 6.4A), where the two main absorption peaks can be attributed to the less planar β-subunits (~655 nm peak) and to the more planar α-subunits (~710 nm peak) (53,179). The only exception is the bilin pigment in the β-subunit ApcF2, which has an almost planar ring D compared to the ApcB2 subunits.

Different behaviour of α- and β-subunits implies that the spectral range covered by the FR-APC evolved to be broader than the WL-PBS complex even when the latter is structurally much more complex. Sacrificing physical absorption cross-section, both in terms of size and spectral coverage of the FR-APC pigments, indicates a need for Chl *f*-based photosynthesis to optimise space in the membrane to be able to increase the number of photosystems present. Additionally, the steeper gradient in absorbances of the bilin

pigments in FR-APC might also be important for achieving a faster and more efficient excitation energy transfer to FR-PSII and consequently avoid photodamage by keeping the rate of excitation energy funnelled into FR-PSII constant in low light conditions.

The bilin pigment in ApcE2 is almost completely planar, with both A and D rings being less than 10° off plane. Conversely, ApcD3 presents more off-plane rings, suggesting that the absorption spectra of this subunit should be blueshifted compared to that of ApcE2. This suggests that ApcE2 probably remains the primary terminal emitter of the FR-APC complex, even in the absence of the Cys that normally covalently binds the bilin in ApcD3.

Furthermore, it is interesting to note that the bilin in ApcE2 is likely approaching the red limit of spectral tuning achievable by phycocyanobilin, as its near-planar conformation leaves little room for further red-shifting. This observation provides an important point of reflection regarding the spectral boundaries of soluble PBS-type antenna systems. It may also help explain why organisms such as *A. marina* have evolved to abandon PBS as their primary antenna and instead rely on more red-shifted, Chl *d*-containing membrane proteins.

6.2.3 Changes in FR-APC subunits lead to changes in the absorbance spectra

Previous structural studies reported the loss of cysteine linkage in ApcE2 (Fig. 6.6B) and ApcD3 (Fig. 6.7D) (58). In comparison, the FR-APC complex from *C. thermalis* PCC 7203 also presents the β-subunit ApcF2, which also lacks the cysteine linkage. A comparison of the chemical environments of the bilins in the far-red-specific subunits ApcE2 and ApcF2 with their white light counterparts ApcE and ApcF provides insight into the features that tune their site energy. The bilin present in the far-red specific membrane linker protein ApcE2 retains the same ZZZ,ssa conformation as in WL-ApcE (Fig. 6.6). Nonetheless, the loss of the cysteine linkage in ApcE2 extends the π-electron system of the pigment, allowing the bilin rings to adopt a more coplanar conformation (Fig. 6.6A, B, C). This structural change contributes significantly to the red-shifted absorption of the pigment (53,179).

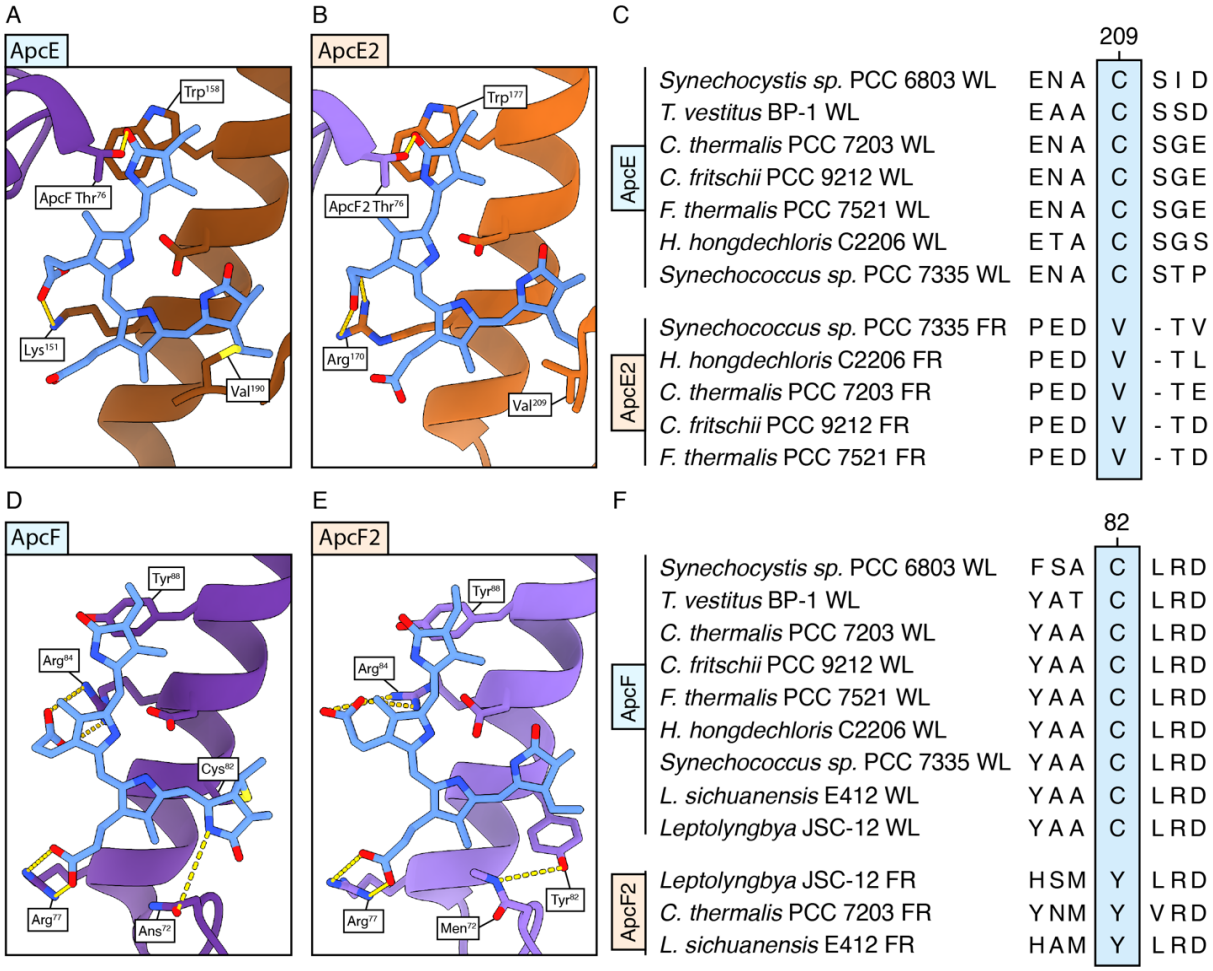


Fig. 6.6 – Protein environment of the ApcE2 and ApcF2 subunits

A) Chemical environment of the bilin pigment in the ApcE protein (7SC7). B) Chemical environment of the bilin pigment in the ApcE2 protein. C) Sequence alignment between selected ApcE and ApcE2 sequences. The light blue box highlights the discussed conserved loss of Cys between WL and FR paralogues. D) Chemical environment of the bilin pigment in the ApcF protein (7SC7). E) Chemical environment of the bilin pigment in the ApcF2 protein. F) Sequence alignment between selected ApcF and ApcF2 sequences. The phycocyanobilin pigments are represented in blue, ApcE in brown, ApcE2 in orange, ApcF in purple, ApcF2 in violet, and hydrogen bonds in yellow. The light blue boxes in the sequence alignments highlight the conserved loss of Cys in the FR paralogues.

In ApcF2, the cysteine responsible for covalent attachment of the bilin is also absent compared to its white light counterpart, resulting in a change in the configuration of the bilin pigment. While in ApcF the bilin is in ZZZ,asa, in ApcF2 the pigment is in ZZZ,ssa conformation (i.e., as in ApcE and ApcE2) (Fig. 6.6). The flip of ring A in ApcF2 leads to the loss of the hydrogen bond between the methylated Asn^{71} and the ring A nitrogen. These changes contribute to the red-shift from 651 nm in ApcF to 675 nm in ApcF2 in heterologously expressed proteins (183,191).

The presence of two distinct non-covalently bonded ZZZ,ssa bilins provides an opportunity to compare their protein environments and suggest potential changes that further red-shift the ApcE2 phycocyanobilin to suit its role as the putative terminal emitter. The chemical environments of the two chromophores differ in the following features. Firstly, ApcE2 features Trp^{177} which sterically interacts with ring D, holding it in a more planar position and potentially π-stacking with it. In an equivalent position, ApcF2 presents Tyr^{88}, which does not interact with the chromophore. Secondly, in ApcE2 the keto group of ring D is hydrogen-bonded by the conserved ApcF/ApcF2 Thr^{76}. Finally, the interaction partners of the propionate moieties are different. In ApcF2, Arg^{77} and Arg^{84} interact with both propionates of the bilin chromophore, whereas only the propionate of ring C interacts with ApcE2 Arg^{120}. This could tune the electrostatic environment of the chromophores, influencing their absorption properties.

To identify the amino acid residues responsible for red-shifting the absorption spectrum of α-subunits, far-red protein sequences were compared to their white light paralogues. Although ApcD5, ApcD3, and ApcD2 are paralogous to ApcD1, they were compared to ApcA instead, as all occupy the same structural position in the APC cores (45,53,182). The phycocyanobilins in ApcD5, ApcD3 and ApcD2, when compared to that in the white light expressed ApcA, have their rings A and D shifted to a more planar conformation with ring B and ring C. The bilin-binding site lies adjacent to a protein loop that differs between FR and WL paralogues (Fig. 6.7E). A loss of this loop (residues AYG in ApcA) in ApcD5 results in a change in packing that influences the position of ring C and the hydrogen bond network of ring A (Fig. 6.7B, E). The deletion displaces Asn^{71} from the nitrogen of ring A, removing the hydrogen bond. In turn, the conserved far-red change in ApcD5, ApcD3, and ApcD2 Thr^{65} (Ile in ApcA) provides a hydrogen bond to the keto group of ring A. Additionally, the FR-specific ApcD5, ApcD3, ApcD2 Ile^{118} (Thr in ApcA) also contributes to the shift of ring A by steric hindrance with its bulkier side chain.

The shift of ring D between ApcA and ApcD5, ApcD3 and ApcD2 could be attributed to the Tyr^{84} to Trp^{84} conserved change (Fig. 6.7), resulting in a 0 to ~30° distortion of ring D relative to ring C (Fig. 6.7, Fig. 6.5B). The position of ring D in ApcD5, ApcD3 and ApcD2 is also influenced by a conserved Thr^{75} to Cys^{75} change in ApcB2. The bulky sulfur side

chain of ApcB2 Cys[75] shifts ring D of the bilin in ApcD3, making it more coplanar with ring C. Finally, the binding of the propionate moiety of ring B also differs among the subunits. In ApcA, the propionate is not directly bonded and is in proximity to an electron-rich residue (Met[77]). In ApcD5 and ApcD3, the propionate forms an ionic bond with Lys[74], while in ApcD2 it forms a hydrogen bond with Tyr[74]. These distinct bonding interactions alter the electrostatic environment surrounding the bilin chromophores and likely contribute to the red-shifted absorption properties observed in the far-red α-subunits (180).

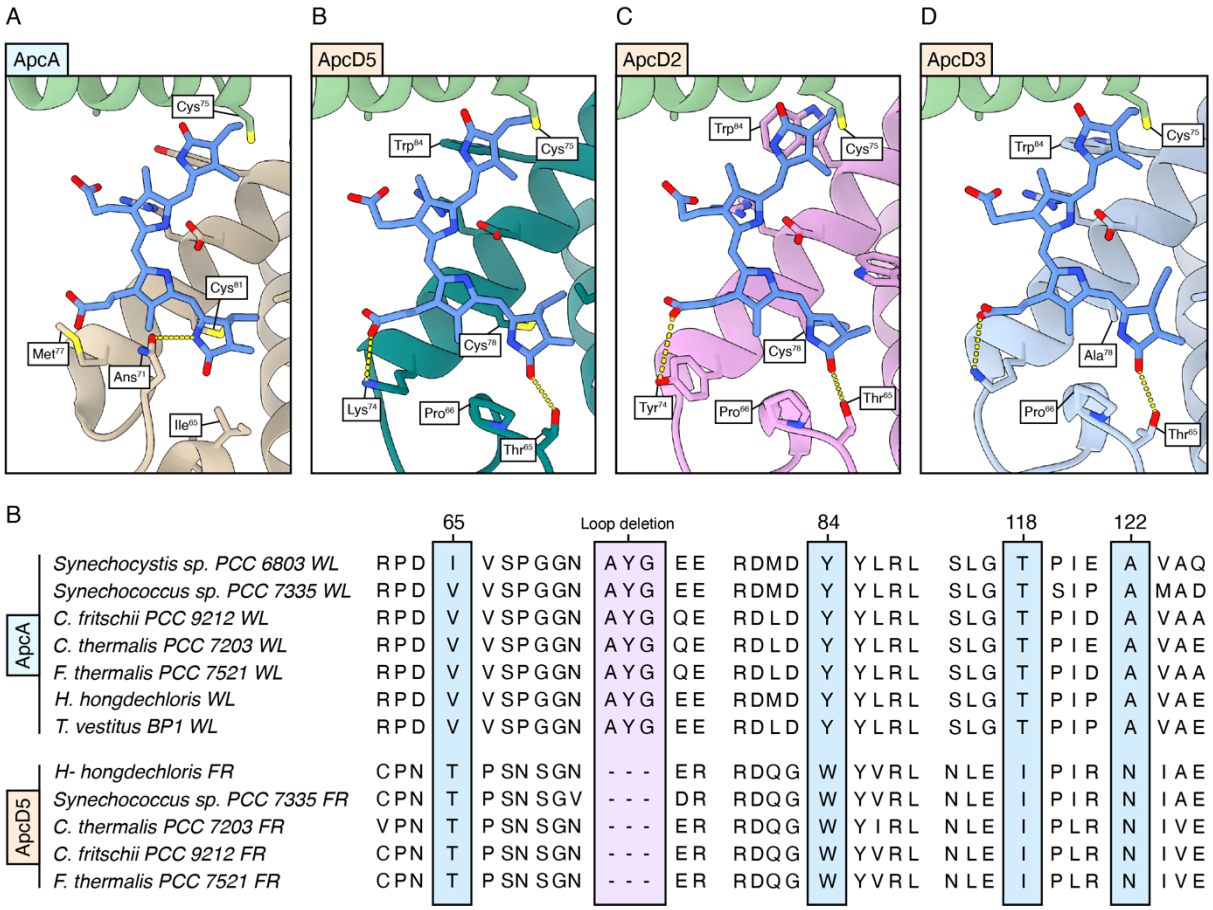


Fig. 6.7 – Changes in the bilin sites of ApcD2, ApcD3 and ApcD5

Comparison of the chemical environment of the bilin pigment in the ApcA protein (A) against the ApcD5 (B), ApcD2 (C), and ApcD3 (D) proteins. The bilin pigments are always represented in blue. E) Sequence alignment between ApcA and ApcD5, the light blue box highlights conserved changes in the sequence between white light and far-red paralogues and the pink box highlights the loop deletion discussed in the text.

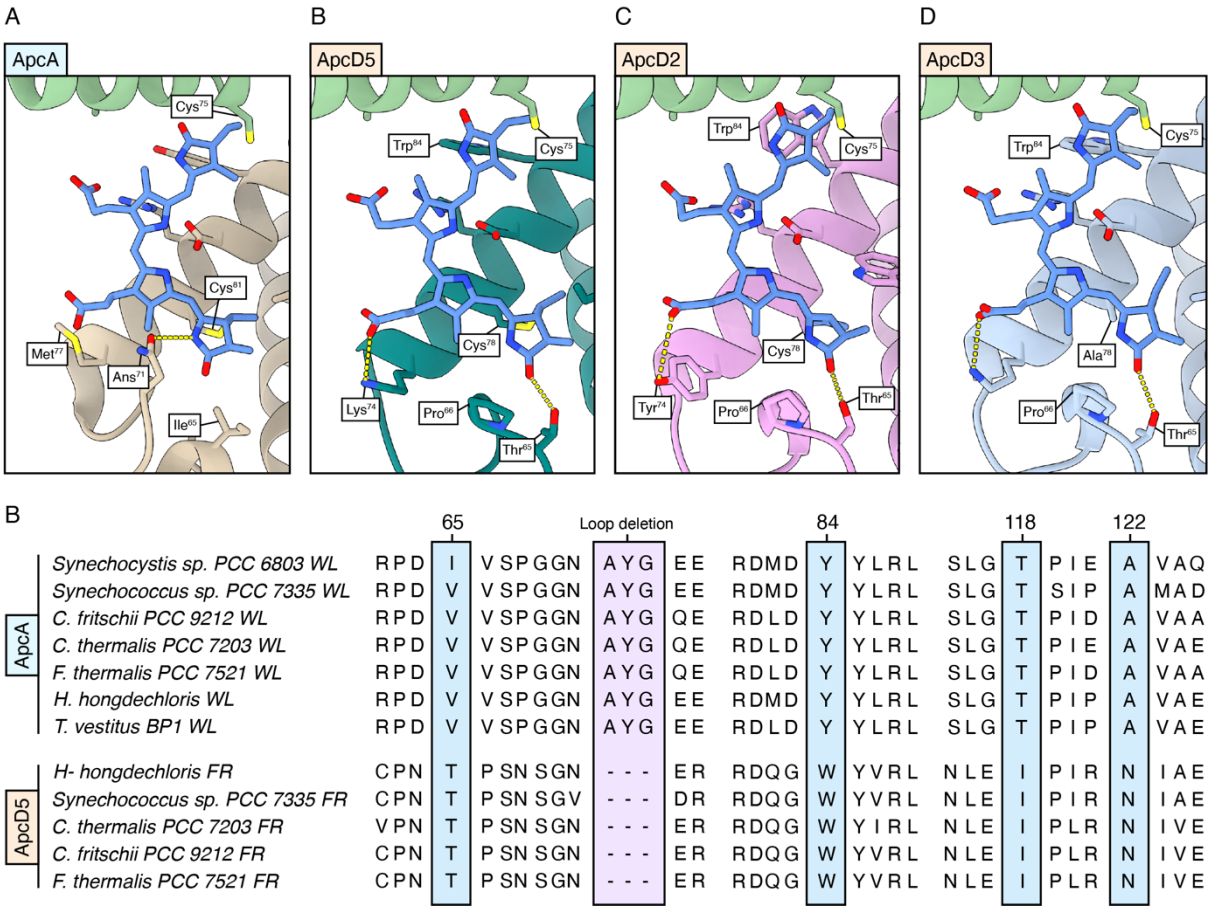


Fig. 6.7Some more changes differentiate the three ApcD paralogues present in FR-APC. ApcD3, one of the previously proposed emitters (53,181), Cys^{75} is replaced by Ala in most available sequences (58). In ApcD2, the far-red conserved residue Trp^{84} is in a different rotameric conformation. While Trp^{84} is oriented perpendicular to ring D in ApcD3 and ApcD5, sterically constraining its position, it lies parallel to ring D in ApcD2. As in ApcE and ApcE2, Trp^{84} may engage in π-stacking with ring D, influencing the electronic distribution of the bilin pigment and consequently its absorption (180). No major structural changes are observed in the bilin chemical environment of the far-red-adapted β-subunit ApcB2. Consistent with its unaltered bilin environment, the absorption spectrum of ApcB2 remains highly similar to that of its white light-expressed paralogue, ApcB (179,181).

6.2.4 *Absence of the distal ApcC*

As pointed out in the structural overview of the complex, the distal ApcC subunit, which is present between the third and fourth rings of the complex in WL-APC, is absent in FR-APC (Fig. 6.9). The dimensions of the ApcC subunit would not easily allow its detachment during purification, since it is held in position by the presence of the fourth ring on one side and the ApcE2 on the other. For ApcC to be released, the entire fourth ring would also need to be absent, as in the case of the previously published structures (58). Moreover, phylogenetic and structural analyses reveal poor sequence conservation in the C-terminus of ApcE2 with some of the residues with bulkier side chains creating steric hindrance that would prevent ApcC from inserting in the distal site when compared with the white light paralogue ApcE (Fig. 6.8). These findings support the conclusion that the distal ApcC in FR-APC is not missing due to purification damage but that it is instead a genuine feature of the FR-APC complex.

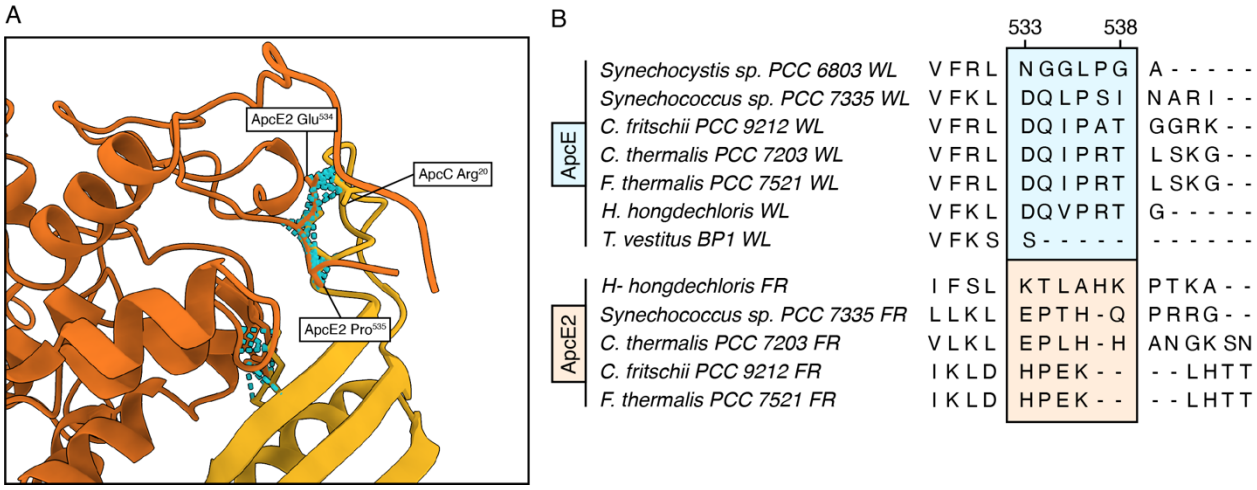


Fig. 6.8 – Clashes between distal ApcC subunit and ApcE2

A) Structural analysis of the potential interaction between ApcC and ApcE2 if ApcC was in the same position compared to WL-PBS. ApcE2 is represented in orange and ApcC is represented in yellow, in blue are represented the molecular clashes between the two subunits. B) Phylogenetic analysis of the ApcE and ApcE2 FR and WL.

For this reason, it is important to analyse the differential effects of the presence and absence of ApcC on the bilin binding site of adjacent subunits. By comparing the two relevant ApcB2 subunits, one in the first ring, which interacts with ApcC both in WL-APC and FR-APC (Fig. 6.9A), and one in the fourth ring, which interacts with ApcC in WL-

APC but not in FR-APC (Fig. 6.9), it is clear that ApcC influences the conformation of bilin pigments.

The presence of the hydrogen bonding residue ApcC Gln^{40}, together with the steric effect of ApcC Phe^{37}, ensures that the orientation of ring D is fixed and in a more planar conformation when compared to the ApcB2 bilin in the absence of ApcC (Fig. 6.9), likely affecting the absorbance spectrum. The absence of the distal ApcC may tune the pigments in the distal fourth $(\alpha\beta)_3$ trimer, creating an energy gradient that facilitates excitation energy transfer to the terminal emitter ApcE2.

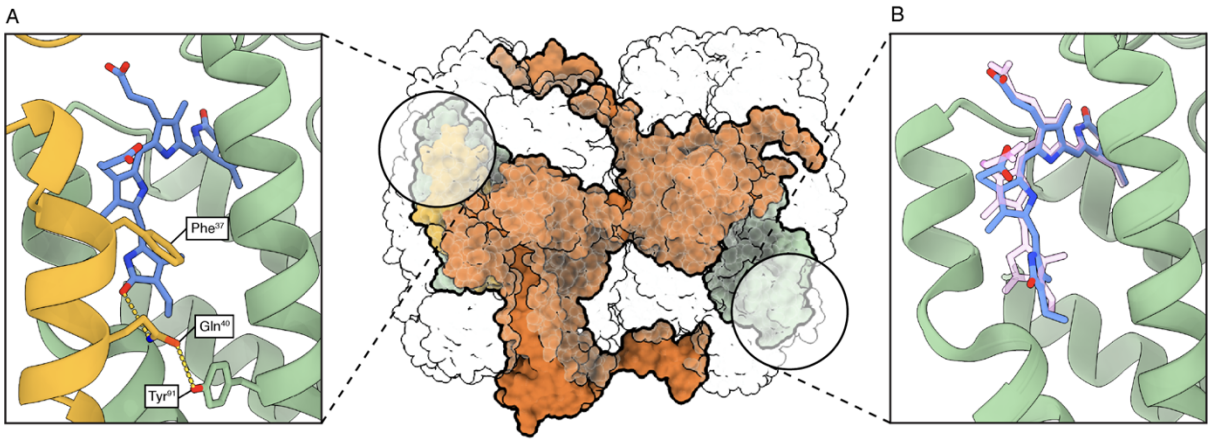


Fig. 6.9 – Differences in the ApcB2 bilin site in the presence or absence of ApcC

In the centre, one cylinder with ApcE2 is represented in orange, ApcC is represented in yellow and two ApcB2 subunits in green. A) Proximal ApcB2 bilin binding site in the presence of ApcC. ApcB2 is represented in green, ApcC is represented in yellow, and phycocyanobilin is represented in blue. B) Distal ApcB2 bilin binding site in the absence of ApcC. ApcB2 is represented in green while phycocyanobilin is represented in blue. The position of the proximal bilin site compared is represented in transparent pink.

6.2.5 Dimerization interface between the core cylinders

The dimer interface of the two cylinders is consistent with that predicted from the previous structural data of the monomeric first three rings of the complex (Gisriel et al., 2024), with ApcD2 being one of the main contributors, interacting extensively with an ApcB2 and an ApcD5 subunit of the third ring, as well as with an ApcD5 subunit of the fourth ring in the neighbouring cylinder (Fig. 6.10B). Nonetheless, contributions to the structural integrity of the bicylindrical complex are not limited to this subunit. The C-terminal tail of the

membrane linker protein, ApcE2, interacts extensively with an ApcD5 subunit in the third ring of the opposing cylinder (Fig. 6.10A). Moreover, the ApcF2 subunit may interact with the ApcF2 in the opposing cylinder. When the two cylinders are brought in proximity by the dimerization, the two ApcF2 subunits are in position for long polar side chains to interact and provide stability to the complex. The high degree of flexibility of these side chains, together with the poorer ESP density at the periphery of the map, does not allow these interactions to be observed.

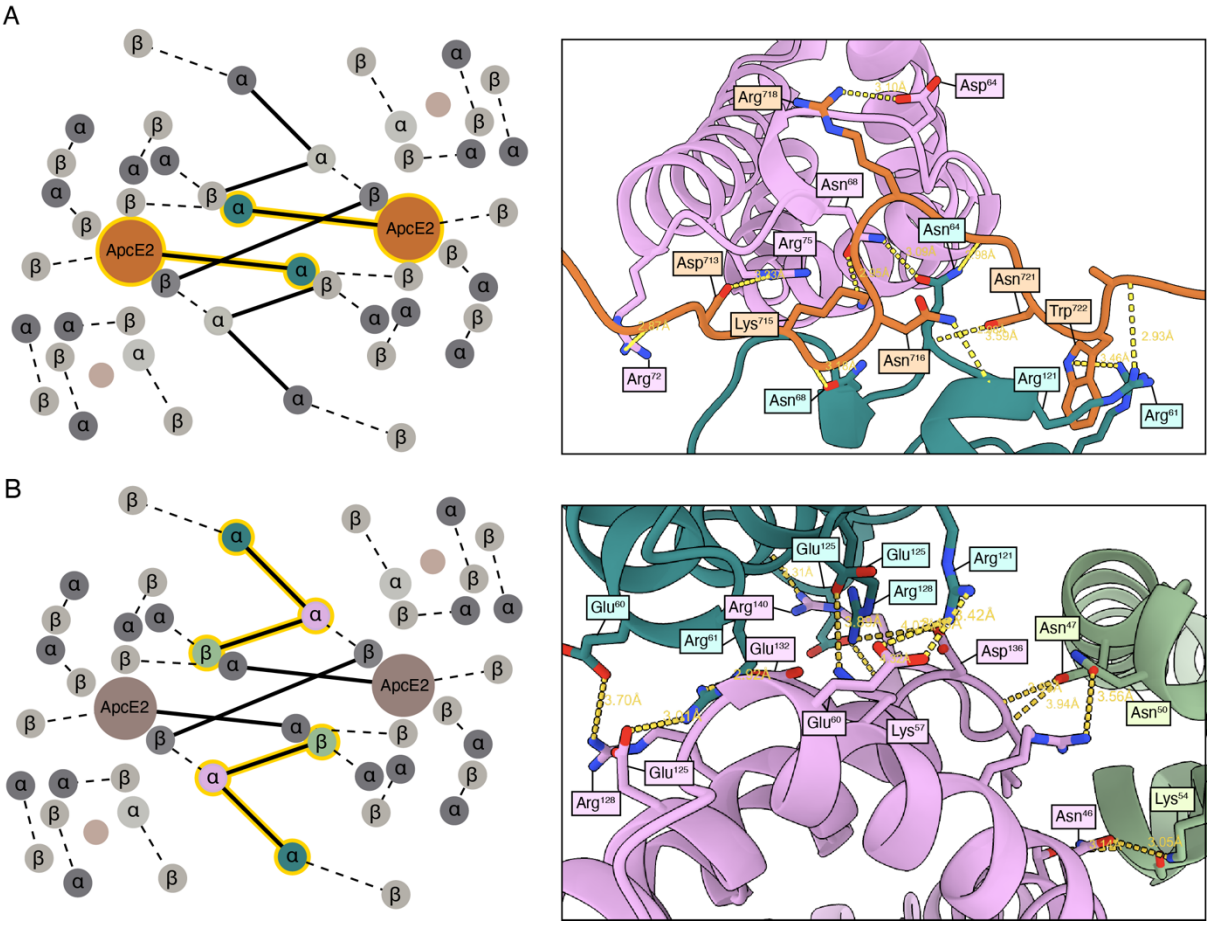


Fig. 6.10 – Interfaces stabilizing the bicylindrical FR-APC

A) The interaction of ApcE2 (in orange), with an ApcD5 (in teal) subunit of the third ring of the opposing cylinder is analysed on the panel on the right. B) ApcD2 (in pink) interactions with an ApcB2 (green) of the third ring and an ApcD5 (in teal) subunit of the fourth ring of the opposing cylinder are analysed in the panel on the right. On the left of panel A and B the network of the interaction between monomers of the bicylindrical FR-APC is represented. Subunits are represented as circles of dimension proportional to their length in amino acids. Interaction of α and β subunits to form (αβ) monomers are represented as dashed lines, while inter-cylinder interaction are represented by black solid lines.

6.3 Discussion

6.3.1 Array arrangement of PSII

During far-red light photoacclimation, the photosynthetic apparatus undergoes a series of profound changes, which alter the protein and pigment compositions of antenna proteins and reaction centres. Under these conditions, the structural arrangement of photosystems and APC is likely to be altered and to reflect the need for these cyanobacteria to maximise photosynthetic productivity with the limited number of photons available.

To gain insight into the distances between bilin pigments in the bicylindrical FR-APC and Chl *f* molecules present in FR-PSII, the structural models of FR-APC and FR-PSII were fitted to the recently released *in situ* Cryo-EM single particle analysis structure of the PBS-PSII supercomplex from cyanobacterial *Arthrospira platensis* (61) (Fig. 6.11A). In the model obtained, the C2 symmetry axes of FR-PSII and of the bicylindrical FR-APC are not parallel, but at an angle of ~20°, suggesting that the two cylinders of the FR-APC are likely to have two independent docking sites on FR-PSII dimers (Fig. 6.11B). This is confirmed when comparing the two opposite connection sites of the FR-APC cylinders to the array of FR-PSII complexes. On one side, ApcE2 is positioned directly above CP43, while on the other, ApcE2 is wedged between two adjacent FR-PSII dimers. The resolution of the *A. platensis* PBS-PSII supercomplex does not allow the identification of specific residues that stabilise these interactions, and different species may have different interaction interfaces. For this reason, the model serves as an approximate indication of distances rather than an actual representation of the array structures in far-red photoacclimated organisms.

The presence of two distinct docking sites for the ApcE2 subunits on the opposing sides of the FR-PSII array has important implications both in terms of excitation energy transfer dynamics from the bilin terminal emitter to the Chl *f* molecules in FR-PSII, as well as physiological and experimental repercussions.

In terms of physiology, this indicates that the PBS complexes, both in white light and far-red organisms, are capable of connecting to photosystems only when they are arranged in an ordered array. In such an arrangement, it is possible that intra-PBS excitation energy

transfer is relevant, allowing excitation energy to be shared over the entire array of photosystems. Moreover, it's possible that the formation of the arrays themselves is induced by the presence of ApcE, which stabilises the side-by-side interaction of adjacent PSII dimers.

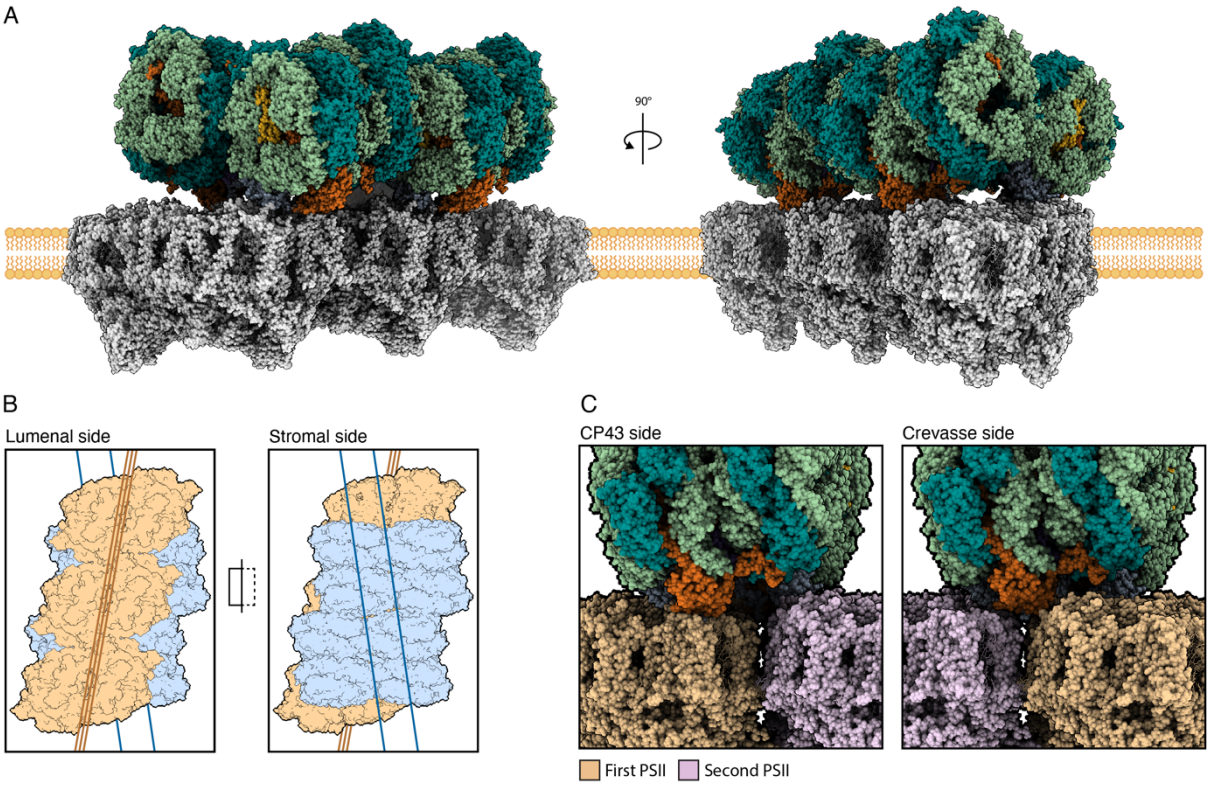


Fig. 6.11 – Mesoscale arrangement of FR-PSII and FR-APC in thylakoids

A) representation an array of three FR-PSII dimers connected to two bicylindrical FR-APC. B) Top and bottom view of the FR-APC + FR-PSII array complex with their respective C2 symmetry axes. FR-PSII is represented in orange, while FR-APC in blue. C) Close up on the connection of FR-APC to FR-PSII on both sides of the FR-PSII array. On the left the interaction with CP43 can be seen and on the right ApcE2 is instead between two FR-PSII dimers. FR-APC is coloured with the same colour scheme used throughout the chapter. The two adjacent FR-PSII dimers are represented in orange and pink.

From an experimental point of view, this behaviour effectively explains the lack of solved structures of PSII+PBS in Cryo-EM single particle analysis, since membrane solubilization treatments would inevitably destroy the supercomplexes and consequently cause the disconnection of the PBS to the photosystems. This highlights the importance of following new strategies to characterise these complexes, with *in situ* CryoET being perhaps the most elegant approach. An alternative possibility is to obtain a Cryo-EM SPA structure of these

complexes while taking particular care during the membrane solubilization steps. To isolate intact arrays of FR-PSII and FR-APC, milder detergents would be required. The use of digitonin has proven capable of maintaining supercomplexes intact; therefore, it would be a good starting point in developing protocols for the isolation of cyanobacterial PSII+PBS complexes.

6.3.2 Terminal emitter of FR-APC and the injection Chl f in FR-PSII

In the model generated, three FR-PSII dimers and two bicylindrical FR-APC are present, providing insight into the relative position of the FR-APC terminal emitters to the Chl *f* molecules present in the FR-PSII dimers, as well as the distance of Chl *f* pigment between adjacent FR-PSII dimers. The edge-to-edge distance between the two putative terminal emitters (i.e., the bilin pigments in subunits ApcD3 and ApcE2) and the candidate Chl *f* sites (see Chapter 5) present in two adjacent FR-PSII dimers has been calculated (Fig. 6.12A).

On both sides, the bilin pigment in ApcD3 is always further away from Chl *f* molecules than the bilin in ApcE2. On the CP43 side, it is ~40 Å away compared to the ~31 Å of ApcE2, while on the "crevasse" side it is ~36 Å compared with the ~27 Å of ApcE2. Nonetheless, the distance of ApcD3 from ApcE2 is only ~22 Å, allowing efficient funnelling from ApcD3 to ApcE2. These observations indicate that the candidate terminal emitter bilin in ApcD3 is unlikely to have a major contribution to excitation energy transfer to FR-PSII since it is always further away from Chl *f* compared to ApcE2.

This is in accordance with the phylogenetic analysis discussed previously, which shows that only a proportion of FaRLiP species lack the cysteine that covalently binds the bilin pigment in ApcD3 (Gisriel et al., 2024b). The absence of this cysteine is considered to be crucial for a pronounced red-shift in the bilin absorption spectra, which is needed to efficiently transfer excitation energy to long-wavelength pigments such as Chl *f* molecules. For this reason, it is possible that, as in the WL version of the APC cylinders, there is only one terminal emitter, the bilin bound to ApcE2. Nonetheless, it is possible that different species present competing paths for excitation energy transfer to FR-PSII, involving the other subunits such as ApcF2 and ApcD3.

The distances of the FR-APC terminal emitter to Chl *f* molecules in this complex, reveal two different scenarios when comparing the ApcE2 positioned on top of CP43 and the one wedged between the two adjacent FR-PSII dimers (Fig. 6.12B). On the CP43 side, the closest long-wavelength pigment is the Chl *f* C507 at a distance of ~31 Å (Fig. 6.12B). The Chl *f* C507 site has been confirmed structurally in different organisms and is indicated as an early FR-PSII Chl *f* site based on ancestral sequence reconstruction (197). Moreover, this site is currently considered as the injection site for excitation energy from the FR-APC to FR-PSII (114).

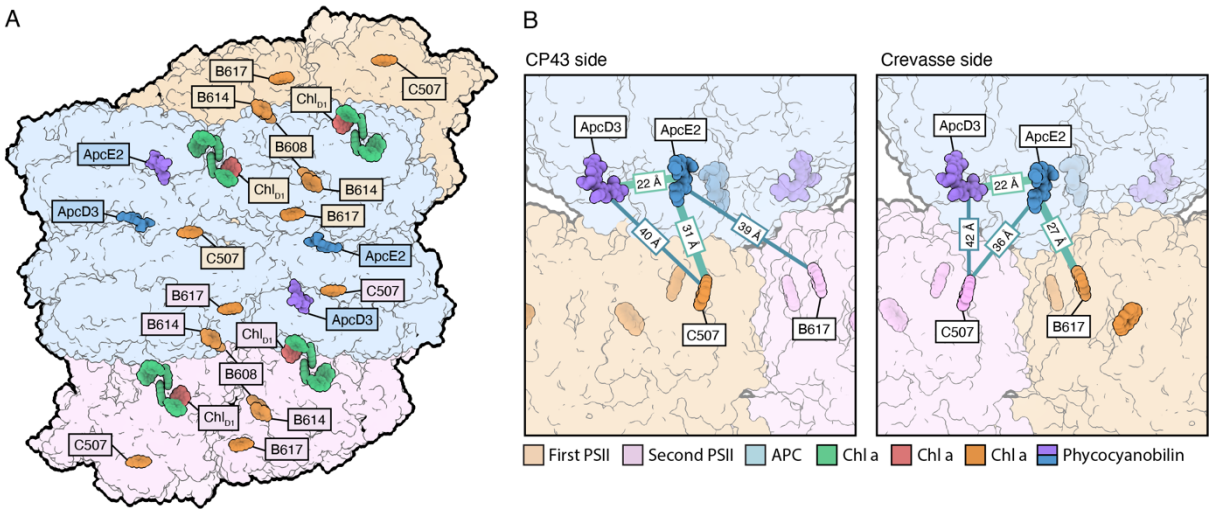


Fig. 6.12 – Emitter-injector distances in a PSII dimer array assembled with FR-APC

A) Lumenal view of two adjacent FR-PSII dimers (in orange and pink) mounted by a bicylindrical FR-APC (in light blue). Reaction centre Chl *a* are coloured in green, Chl *f* sites are coloured in orange, Chl *d* Chl$_{D1}$ is coloured in red and the putative bilin terminal emitters are coloured in purple and blue. B) Close up on parallel to the membrane plane of the two sides of the complex. Chl *f* sites are of the same colour of their respective dimer. On the left panel, the ApcE2 subunit on top of CP43 is represented, while on the right panel the ApcE2 subunit that is between two dimeric FR-PSII is represented.

Conversely, on the side where ApcE2 is between two adjacent FR-PSII, the closest long-wavelength pigment is Chl *f* B617 at a distance of ~27 Å compared to the ~36 Å of Chl *f* C507 (Fig. 6.12B). This observation strengthens the arguments presented in the previous chapter that identify this site as containing a Chl *f*. The presence of the new subunit PsbH2' might be related to the presence of an important secondary excitation energy injection point into FR-PSII. In light of this, the potential function of the novel subunit PsbH2' might be to prevent the formation of WL-PBS + FR-PSII complexes, or possibly to correct

the orientation of the bicylindrical APC with respect to FR-PSII to improve connectivity between the two.

The structural data presented here is in agreement with the time-resolved fluorescence data collected on FR-APC + FR-PSII complexes (114), which indicate that trapping times of FR-APC + FR-PSII complexes are faster when the bilin pigments are excited compared to when Chl *f* pigments are directly excited. However, the presence of the ApcE2 terminal emitter in close proximity to Chl *f* C507 on one side and to Chl *f* B617 on the other, does not allow to discern which of the two pathways is more efficient in transferring excitation energy to Chl_{D1}. On one side, Chl *f* C507 is considered to be the Chl *f* with the fastest transfer rate to Chl_{D1} in FR-PSII (114). On the other, the proximity of Chl *f* B617 and the network of Chl *f* pigments to Chl_{D1} needs to be considered. A site directed mutagenesis campaign targeting Chl *f* C507 and B617 would allow determination of whether there is a preferential route of excitation energy to the FR-PSII reaction centre from the FR-APC and, if so, which one it is.

6.3.3 PsbH2' and PsbH2 as a potential stabilizing structure of FR-APC

In the FR-PSII + FR-APC complex, the ApcE2 and ApcD3 subunits, located between adjacent FR-PSII dimers, are in close proximity to the far-red-specific subunits PsbH2 and PsbH2'. The N-terminal residues of PsbH2 and the C-terminal residues of PsbH2' interact with each other; however, several polar residues remain unengaged in specific hydrogen bonds. Among these conserved residues, ApcD3 Arg^{170} is positioned near PsbH2' Asp46, while PsbH2' Thr^{42} and Tyr^{38} are exposed and do not participate in any discernible interactions in the resolved structures. Similarly, PsbH2 Asn^{20} is in proximity to ApcE2 Arg^{162}, and PsbH2 Ser^{23} is situated near the backbone of ApcE2. Additionally, PsbH2 Glu24 and Lys^{27} are in close proximity to ApcE2 Ser^{80} and Asn^{73}. Regarding the interaction with the adjacent FR-PSII dimer, where the PsbC2 subunit is positioned near ApcD3 and ApcE2, PsbC2 Lys^{251} is in close proximity to ApcD3 Arg^{175}, with several other polar residues exposed. While additional interactions are likely present, the level of uncertainty in the structural fitting precludes their precise identification.

Molecular docking analyses, coupled with a comparative investigation of phylogenetically conserved substitutions in far-red light-associated subunits relative to their white-light-adapted counterparts, could provide further insights into the binding interface between the WL-PBS complex and arrays of PSII dimers, as well as contribute to elucidating the structural determinants of FR-APC binding specificity.

The novel subunit PsbH2', whose function remains uncharacterised, may contribute to stabilizing the interaction between FR-PSII and FR-APC or, alternatively, may prevent the association with WL-PBS through steric hindrance. However, the current data remain preliminary, and experimental validation through knockout mutants and structural approaches (e.g., cryoET) will be necessary to elucidate its functional role.

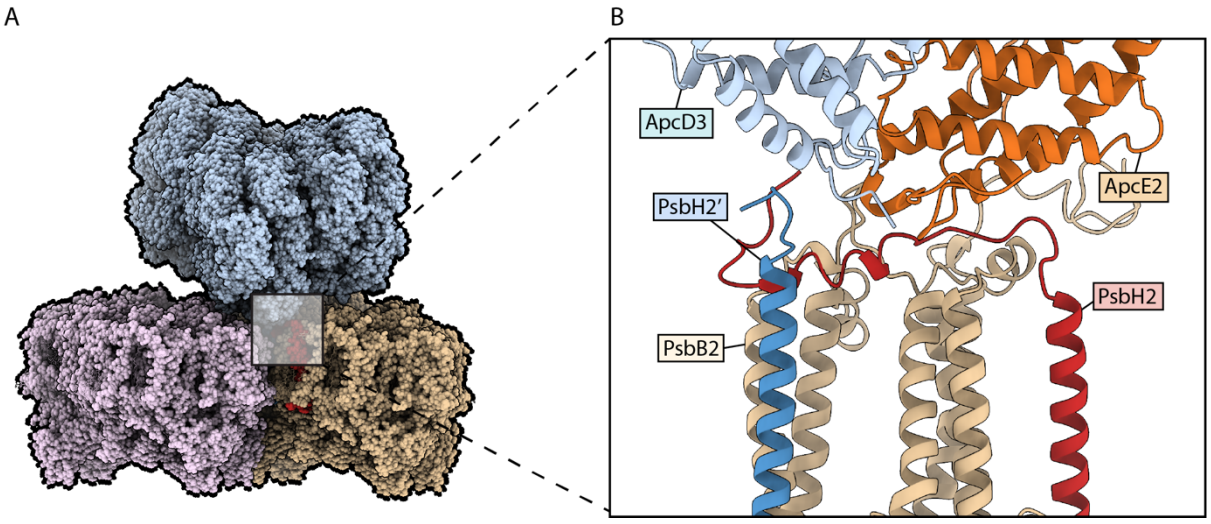


Fig. 6.13 – Potential interactions between PsbH2 and PsbH2' with ApcD3 and ApcE2

A) Side view of the FR-PSII + FR-APC model, the two FR-PSII dimers are represented in pink and yellow, while the FR-APC is represented in light blue. B) Close up of the model fitting as ribbons showing the close proximity of ApcD3 and ApcE2 with PsbH2 and PsbH2'

6.4 Conclusions

The first intact structure of the bicylindrical FR-APC presented in this chapter provides key structural data to understand excitation energy transfer from the far-red light specific antenna to the far-red-adapted reaction centres. This is crucial to confirm and strengthen the Chl *f* candidates presented in the previous chapter and to understand which Chl sites act as primary acceptors both in far-red-adapted and white light organisms. By considering the arrangement of photosynthetic complexes and FR-APC in the thylakoid membrane, the location of Chl *f* sites in FR-PSII can be rationalised and understood, with important repercussions for our understanding of the limitations that these organisms experience in unfavourable light conditions. Future engineering of Chl *f* in crop plants should take into account the organization of photosynthetic membranes in higher plants and lean towards the specific insertion of long-wavelength absorbing pigments in key sites rather than favouring their random insertion.

The analysis performed on the planarity of bilin pigments, together with the in-depth analysis of phylogenetically conserved changes in different subunits of WL-PBS and FR-APC, provides a clear indication of which amino acids are used to constrain the chromophores in the FR-APC complex and consequently how their absorption spectra are tuned. The fact that ApcB2 retains a spectrum similar to the white light β-subunit while ApcD5 is more red-shifted allows FR-APC to cover a broader spectral range compared to WL-APC. The more complex WL-PBS structures allow a larger antenna cross-section but are limited in the wavelengths they can absorb. On the other hand, Chl *f*-containing organisms seem to prioritise spectral coverage as well as reducing the intermembrane space to pack more membranes and photosystems.

The evolutionary drive towards having a high number of reaction centres with smaller antennae, compared with larger far-red antennae with fewer reaction centres, might indicate that there could be functional limitations that prevent bilin pigments from absorbing wavelengths longer than ~710 nm. Consequently, these organisms may have to rely on Chl *f* sites in the photosystems themselves to access those photons.

Consistent with the previous chapters, the work presented here allows for even more complex and higher-scale computational modelling of the excitation energy transfer in FR- and WL-adapted membranes, providing ways to analyse the benefits and drawbacks of extending the photosynthetic red energy limit as well as changing the organization of the thylakoid membranes. The structure of FR-APC completes the triad of light-harvesting and light-driven complexes present in the FaRLiP cluster, allowing a comprehensive view of the far-red light photoacclimated photosynthetic apparatus in *C. thermalis* PCC 7203.

Cryo-EM single particle analysis of photosynthetic complexes has, in recent years, provided a number of new structures from many different photosynthetic organisms. The vast amount of data that can be obtained from a high-resolution structure needs to be integrated into a broader view of the structural organization of photosynthetic membranes. Different growth conditions can lead to different membrane organizations that can be rationalised by leveraging Cryo-tomography data as well as *in situ* single particle analysis structures.

7 Towards understanding RfpA dynamics

7.1 Introduction

A debated aspect regarding the improvement of productivity in photosynthetic bioprocesses and crop plants is the possibility of expanding the absorption range of photosystems and antennae to limit the amount of self-shading in leaves and photobioreactors (90). Far-red light photoacclimation occurs in natural niches that present some features in common with those found in high-density cell cultures grown in photobioreactors. In both cases, cyanobacteria are in a condition in which photons in the absorbance bands of Chl *a* are scarce and therefore in a situation where photosynthesis is light-limited. For this reason, far-red light photoacclimation presents an important form of adaptation that can be studied to assess the advantages and disadvantages of engineering longer wavelength antenna and reaction centres in Chl *a* containing organisms in photobioreactors and, more probably, in the lower leaves of the canopies of crop plants.

To understand the biological costs of engineering Chl *f* in Chl *a*-only photosynthetic organisms, it is crucial to study not only the architecture of long-wavelength pigments in photosystems and antenna complexes expressed during this type of photoacclimation, but also to understand its triggers, speed and the light conditions in which it results in a net positive impact on growth.

The extensive remodelling of the photosynthetic apparatus that cyanobacteria undergo during FaRLiP is caused by the expression of a cluster composed of ~21 genes, which encode for the putative Chl *f* synthase gene (*PsbA4*) (107,108), alternative subunits of phycobilisomes and photosystems, and the three constitutionally expressed proteins (i.e., RfpA, RfpB and RfpC) that are needed to sense and transduce changes in light quality and control the activation of the cluster itself. The ratio of white light to far-red light is sensed through a knotless phytochrome, called RfpA (Regulator of far-red photosynthesis A) and transduced by a DNA-binding response regulator (RfpB) and a cheY-like gene (RfpC)

(109). Knocking out any of the three proteins results in the loss of FaRLiP and therefore survival in light with wavelengths longer than 700 nm (109).

Phytochromes are a widespread superfamily of receptors present in all photosynthetic organisms, as well as non-photosynthetic bacteria and fungi (219). These receptors are capable of sensing both changes in quantity and quality (i.e., wavelength distribution) of the incoming photons, as well as temperature (220–222).

Phytochromes are capable of sensing light through a covalently bound bilin pigment, which is capable of switching reversibly between two stable conformers, referred to as P_r (red light absorbing) and P_{fr} (far-red light absorbing) (223,224). As discussed in Chapter 6, the absorption spectra of bilin pigments are strongly influenced by their planarity and chemical environment. When the P_r state is excited by red light or the P_{fr} is excited by far-red light, the ring D of the bilin undergoes an isomerization event that alters its spatial arrangement, allowing it to distinguish between the two spectral ranges. The change in geometry of the pigment results in conformational changes in the secondary and tertiary structure of the phytochrome itself, activating or inhibiting its function accordingly.

When exposed to light, a dynamic photoequilibrium is established between the two forms, with the proportion of molecules in each form being determined by the relative photon fluxes in the two spectral ranges. The active form of RfpA is the P_r state; this state, activates its terminal histidine kinase domain, allowing the transduction of the signal to the other response elements of the FaRLiP cluster (i.e., RfpB and RfpC) and the expression of the far-red-light-adapted subunits. According to the KEGG database (225), phylogenetic analyses (109), and domain architecture, RfpA belongs to a knotless phytochrome group known as Type IV. The overall structure of RfpA is composed of four domains: an N-terminal photosensory GAF domain, which binds the phycocyanobilin chromophore; the PHY domain, exclusively found in phytochromes, which connects the photosensory and effector portions of the phytochrome and is probably involved in dimerization; the PAS domain, which is related to signal transduction in various proteins; and a histidine kinase (HK) domain, which is speculated to be phosphorylating RfpC and initiating the signal cascade that leads to the activation of the cluster.

From a biotechnological point of view, an interesting aspect of Chl *f* containing organisms, when compared to Chl *d* containing organisms, is that Chl *f* containing photosystems seem to show no increased sensitivity to photodamage, with backreaction routes of FR-PSII similar to WL-PSI despite having nearly equal amounts of energy available to Chl *d* PSII present in *A. marina* (98). Moreover, the facultative nature of far-red light photoacclimation allows this response to be triggered in cells subject to specific light conditions. Changes in the absorbance of Chl *f* containing organisms vary from species to species, but these acclimation processes enhance light harvesting at wavelengths between 700–750 nm, accounting for a potential ~19% increase in available photons, which could allow room for improvement in the theoretical maximum efficiency of photosynthesis (90). Nonetheless, these are merely theoretical estimates, and improvements in productivity by altering the photosynthetic apparatus remain speculative (86).

Even though FaRLiP capable cyanobacteria present extremely slow growth rates, engineering the facultative expression of Chl *f* containing photosystems and red-shifted antenna has been proposed as a possible biotechnological strategy to improve the yields of lower leaves in the canopy of crop plants. Moreover, cyanobacteria capable of performing this long-wavelength photosynthetic adaptation have been proposed as ideal candidates to harvest longer wavelength photons in the inner chambers of dual-layer photobioreactors, while exchanging nutrients (e.g., reduced nitrogen species and reducing equivalents) with an external chamber through semipermeable membranes. In particular, the versatility of FaRLiP would allow for a range of operational conditions in such a dual-chamber photobioreactor. However, to design effective biotechnological applications of far-red photosynthesis, the long-wavelength-adapted photosynthetic complexes and the already known functional penalties that come with this improved harvesting capability (98) need to be fully characterised. Moreover, a better understanding of the acclimation process itself is also needed to enable the tuning of its triggers, speed and overall effects on the photosynthetic apparatus to ultimately be evaluated in agricultural and biotechnological processes.

In this chapter, RfpA, the phytochrome responsible for the activation of the FaRLiP cluster, has been heterologously expressed in *E. coli* and photochemically characterised.

Moreover, a mathematical model describing the dynamic response of the phytochrome to different light spectra and irradiance conditions has been developed. To collect growth data and Chl *f* synthesis rates, a dual-light growth chamber has been built. The growth chamber allows experimentation with different ratios of red to far-red light and therefore the full characterization of the model. The population dynamics that regulate the switching are interesting both from the physio-ecological and synthetic biology points of view. The spectroscopic and structural characterization of RfpA would help our understanding of the circumstances that trigger the FaRLiP, elucidating the ratios of white light to far-red light needed for the transition to take place, the rates of conversion between both states, and the influence of temperature on the equilibrium. It would thereby provide the basis to determine physiological and growth properties of Chl *f* containing organisms.

7.2 Results

7.2.1 Photobioreactor setup

Essential to the investigation of the FaRLiP response and to the synthesis of Chl *f*, is an experimental setup that allows for precise control of different light sources. For this reason, a growth chamber prototype was built and equipped with two different spectrally calibrated sets of LEDs (Fig. 7.1).

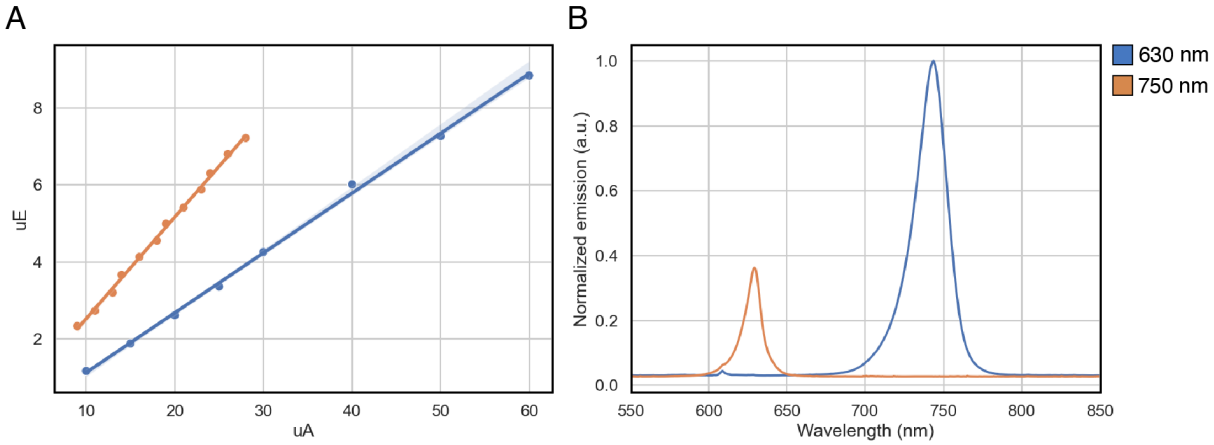


Fig. 7.1 – LEDs emission profiles and current response

Key characteristics of different LED types. A) Linear relationship between the LED's current input and its light output, demonstrating the operational range for each LED. B) Normalised emission spectra, where each color represents a specific LED type as indicated by the figure's legend.

Both the LED custom bars are enclosed within a blackout box, which completely shields them from external light (Fig. 7.2). The blackout box can contain up to two 2L growth bottles, that can be stirred and bubbled with wet air or air/CO_2 mixtures. The internal temperature is regulated by a fan.

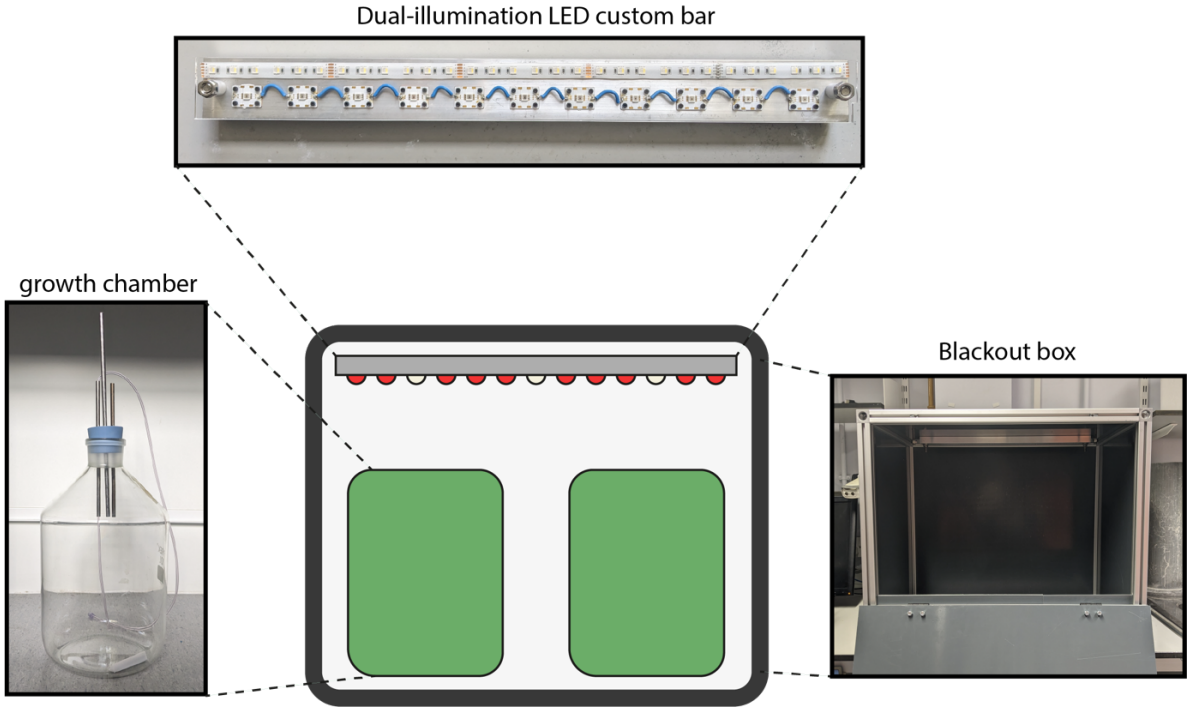


Fig. 7.2 – Dual illumination photobioreactor setup

Schematic illustrating the custom-built dual-light photobioreactor system. This setup, featuring two spectrally calibrated LED sets (top panel) housed within a blackout box (right panel), enables precise, time-dependent control of photon fluxes. It's designed to investigate responses like FaRLiP by simulating various light conditions, including field solar irradiance data.

Simultaneous control of the photon fluxes from both sets of LEDs, in a time dependent manner, allows the simulation of physiologically meaningful irradiance spectra that vary in time. Moreover, field datasets of solar irradiance can be recorded using of a portable spectrophotometer (Solar-Rad, Stellarnet Inc.) to evaluate the expected induction of FaRLiP and Chl *f* synthesis under realistic conditions and then simulated in the growth chamber.

7.2.2 Heterologous expression of RfpA

Since RfpA is naturally produced by *C. thermalis* in extremely low concentrations and would be difficult to isolate natively, the histidine-tagged GAF domain (N-RfpA) and the histidine-tagged full-length protein (F-RfpA) have been heterologously expressed and purified from *E. coli* (KRX strain) using two plasmids: the first expressing the apo-N-RfpA (Fig. 7.3B) or apo-F-RfpA (Fig. 7.3C) with resistance to ampicillin, and the second expressing the enzymes necessary for the conversion from heme to phycocyanobilin (Fig.

7.3A), which contains a resistance gene to chloramphenicol. The plasmids used in this line of research were kindly provided by Dr. Matthew Pope.

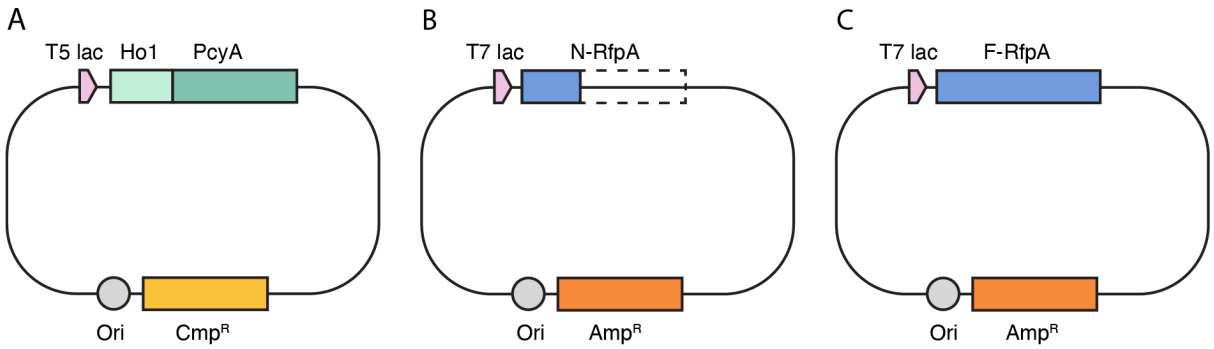


Fig. 7.3 – Plasmids for the expression of Holo-RfpA

A) Plasmid containing Ho1 (heme oxygenase) and PcyA (biliverdin reductase) with resistance to chloramphenicol (Cmp^R) B) Plasmid containing N-RfpA in blue and the resistance gene to ampicillin (Amp^R) C) Plasmid containing F-RfpA in blue and the resistance gene to ampicillin (Amp^R)

The synthesis enzymes used are Heme Oxygenase 1 (Ho1, gene *sll1184*), which breaks the tetrapyrrolic ring of Heme *b* and form biliverdin IXα, and a PCB:ferredoxin oxidoreductase (PcyA, gene *slr0116*), which catalyses the reduction of biliverdin IXα to phycocyanobilin (Fig. 7.4) (226).

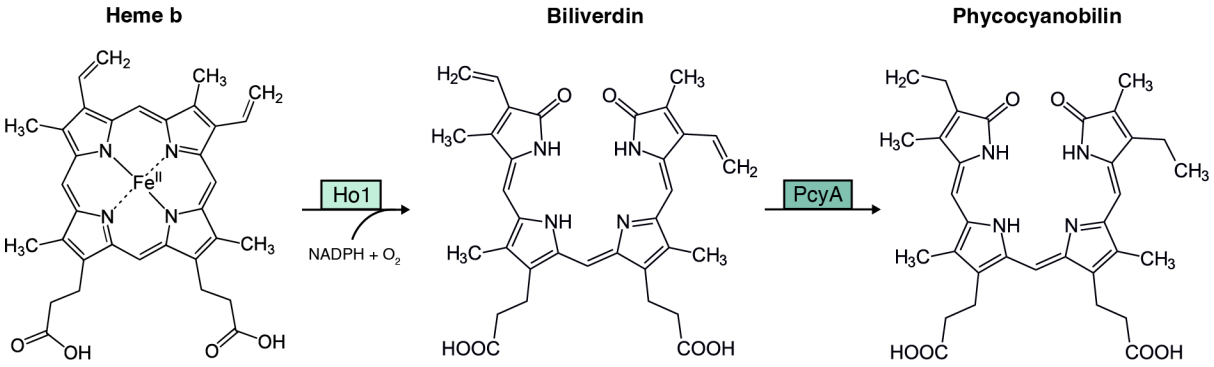


Fig. 7.4 – Synthesis of phycocyanobilin starting from Heme *b*

Schematic enzymological pathway of the heme to PCB synthesis.

After induction (see materials and methods), the pellets presented a light blue hue, indicative of the successful conversion of Heme *b* to PCB.

7.2.3 His-tag purification of RfpA

After cell breakage (see Materials and Methods) the membrane debris and unbroken cells were pelleted by ultracentrifugation. The supernatant of the N-RfpA expressing cells presented a clear blue hue, indicating that the protein was in the soluble fraction, while that of the F-RfpA expressing cells was just slightly tinted, with the pellet retaining most of the pigment.

This suggests the formation of inclusion bodies containing most of the expressed RfpA. The supernatant was incubated with Ni^{2+} resin as described in Materials and Methods. The resin was then washed twice with buffer containing 30 mM imidazole. Then, the protein of interest was eluted with buffer containing 50 mM imidazole three times. The initial flowthrough and the eluted fractions were then loaded on an acrylamide gel for SDS-PAGE (Fig. 7.5).

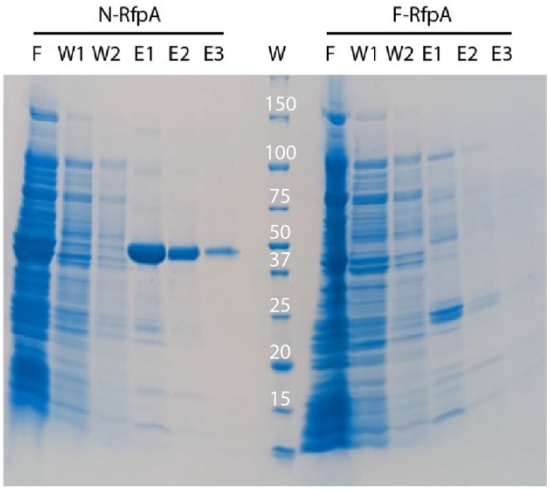


Fig. 7.5 – SDS-PAGE gel of the expression of N-RfpA and F-RfpA

Lanes are labelled according to the following scheme: flowthrough (F), first wash with 30 mM imidazole (W1), second wash with 30 mM imidazole (W2), first elution with 150 mM imidazole (E1), second elution with 150 mM imidazole (E2), third elution with 150 mM imidazole (E3).

The left half of the gel shows the results of the purification of the N-RfpA from the supernatant. In the elution fractions, the gel presents a clear band at the molecular weight of 49 kDa, the expected molecular weight of the GAF domain expressed by the N-RfpA construct. On the right half of the gel, most bands present in the elutions are non-specific binding to the Ni^{2+} resin of contaminants of cleaved forms of the full length product, a

band can be seen around 103 kDa (the expected weight for the full length monomeric form of RfpA), but the yield is low and the band could be due to a contaminant of the same weight.

7.2.4 Absorbance spectra and photoswitching of the GAF domain

The absorbance spectra of the isolated GAF domain were measured to verify the incorporation of the phycocyanobilin pigment. After leaving the protein solution in far-red light or red light red light illumination for prolonged time, the holo-GAF photo converts completely reproducing the spectra available in the literature (109). The P_r state presents an absorption maximum around 635 nm with a shoulder at 600 nm, while the P_{fr} state presents an absorption maximum at 705 nm with a shoulder at around 650 nm (Fig. 7.6A). The presence of a shoulder in the absorption spectra of phytochromes is in line with experimental spectra available in the literature (223,226).

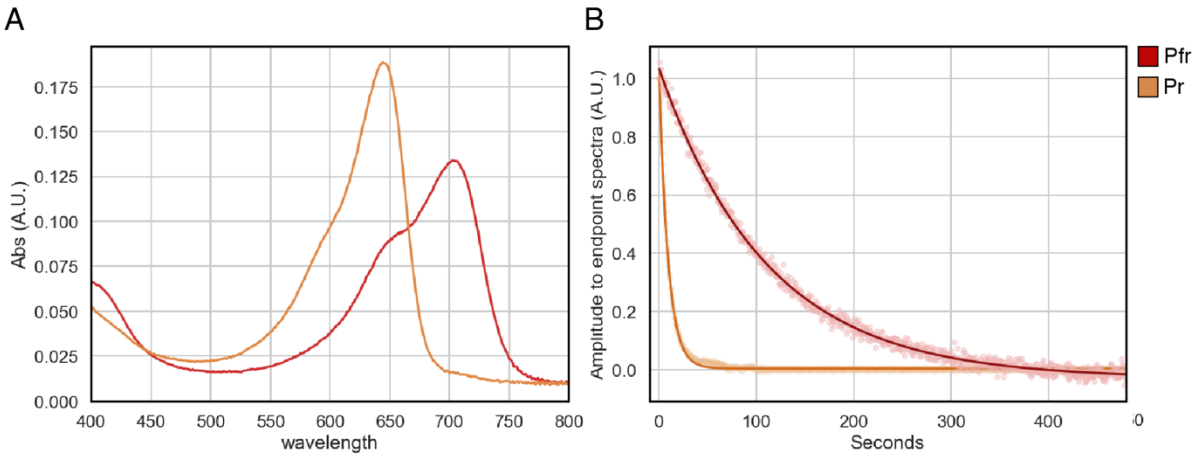


Fig. 7.6 – Absorbance spectra of the P_r and P_{fr} states and photoconversion rates

A) Absorption spectra of the P_r (in orange) and P_{fr} (in red) states. B) Photoconversion of the P_r (in orange) and P_{fr} (in red) states. On the y axis the population of the state of interest based on the ratio between the maximum intensity of the absorption peak and the one at the time measured, from 1 (no conversion) to 0 (full conversion).

To measure the conversion rates of the P_r and P_{fr} states, N-RfpA photoconversion was induced using 630 nm (for the P_r state) and 750 nm (for the P_{fr} state) LEDs emitting the same photon flux at the absorption peak of the P_r and P_{fr} state, while measuring the absorbance spectra every 0.5 s. Rates of conversion could be calculated starting from the exponential decay of the amplitude between the 655 nm P_r peak and the 705 nm P_{fr} peak in

difference spectra with the last acquired data point (Fig. 7.6B). The estimated ratio between the rate of decay of the P_r state and the P_{fr} state is 0.082. This estimate is compatible with exploratory *in vivo* data indicating that the transition is unaffected when a low level of white light (less than 10%) is supplemented during growth.

7.2.5 Attempts to determine the Cryo-EM structure of full length RfpA

The characterization of phytochromes through crystallization is very complex, even though high-resolution crystal and time-resolved crystal structures of different phytochromes were solved, the terminal effector HK domain is always missing due to self-catalytic cleavage during the crystallization process (227–229). For this reason, Cryo-EM SPA could constitute a better methodology to obtain structural data on how the signal is transduced across the domains of RfpA. To date, due to their size, just a handful of phytochrome structures have been determined at medium and low resolution using Cryo-EM (230–232). Purifying RfpA and solving its structure in the P_r and P_{fr} states would advance our knowledge of its structure function relationship and provide insight into the functional and structural variety of light sensors present in photosynthetic organisms.

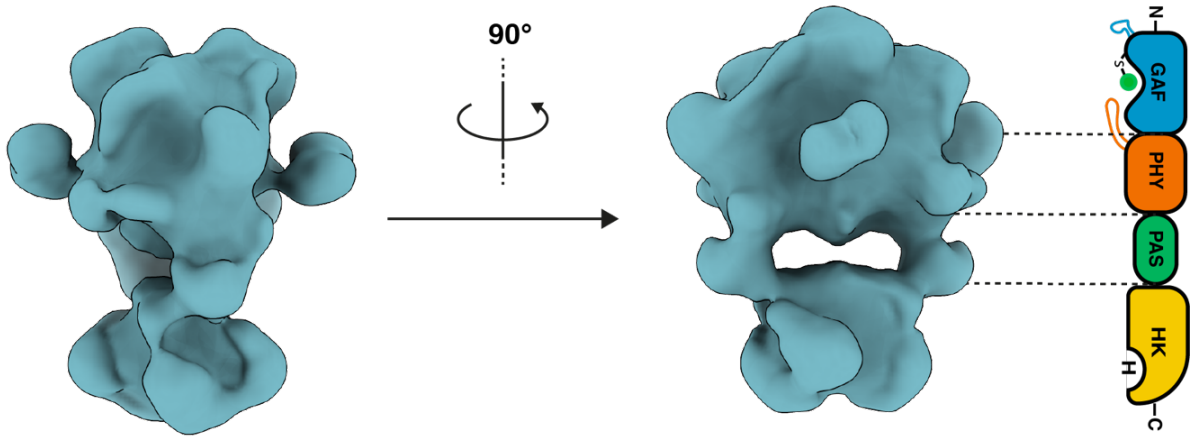


Fig. 7.7 – Low-resolution model of RfpA

Low-resolution *ab initio* model of dimeric RfpA presented in blue. On the right a schematic representation of the domains associated to the ESP density is present.

For this reason, the structure determination of full-length RfpA was attempted. RfpA was exposed to visible light for 10 minutes and then spotted on a grid and frozen. Grids presented good quality ice, with particles of compatible size in multiple orientation. The

data collection yielded 2D classes that resemble the ones found in the literature for solved phytochromes, and a low-resolution non-uniform refinement volume compatible with a dimeric head-to-head conformation of RfpA (Fig. 7.7).

RfpA dimers are ~200 kDa, with multiple flexible regions and two different conformations. The determination of the structure of RfpA requires a more careful sample preparation, to decrease sample heterogeneity. In addition, a larger dataset and a smaller pixel size would be necessary to achieve higher-resolution and retrieve the conformational landscape of the two forms of RfpA.

7.2.6 Mathematical modelling of the influence of RfpA on FaRLiP

To gain insight on how light quality and quantity influences far-red light photoacclimation a time dependent ordinary differential equation (ODE) model was developed, which captures the photoswitching population dynamics of RfpA, as well as Chl f synthesis and culture growth in response to a different light spectra.

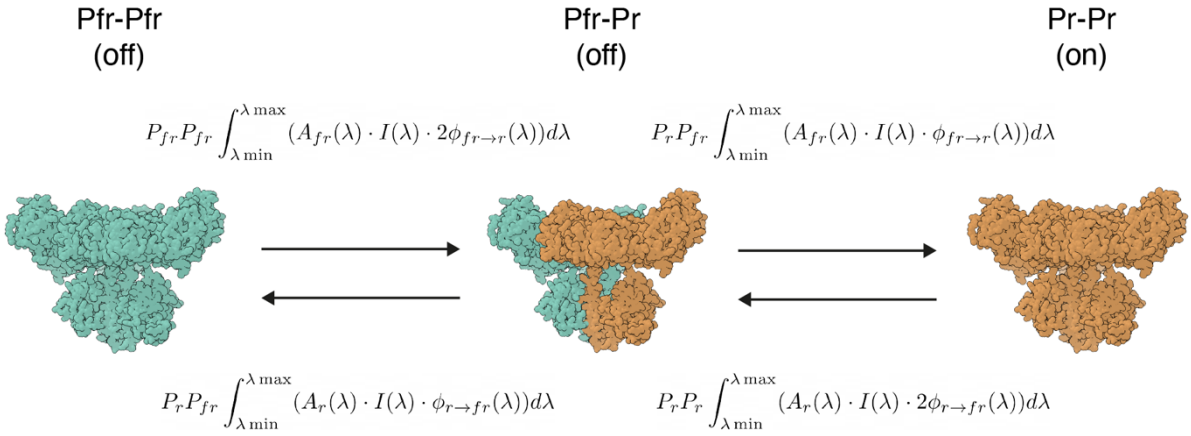


Fig. 7.8 – Illustration of the three-state RfpA model

Graphical representation of the three states described in the RfpA model. Each state is represented as the RfpA phytochrome coloured in blue when in the P_{fr} and in orange when in the P_r state.

Each photoswitching event refers to a single chromophore contained in each of the monomers that constitute dimeric species. Preliminary data collected for the GAF domain showed no thermal reversion at room temperature from the P_r state back to the P_{fr} state. Therefore, in a wavelength-dependent manner, the chromophore photoswitching rates depend linearly on the absorbance (A), light intensity (I) and quantum yield of conversion

(ϕ) only. The monomers switch independently from one another, leading to complex population dynamics that depend on the variations of the light spectrum over time. To quantify how a RfpA triggers the activation of the FaRLiP cluster, and consequently the synthesis of Chl *f*, the photoacclimation dynamics for the FaRLiP cluster gene expression were assumed to follow a switch-like activation (sigmoid term) in terms of the biologically active state P_rP_r. The set of differential and algebraic equations comprising the dynamic model are summarised below:

$$\frac{dP_{fr}P_{fr}}{dt} = -P_{fr}P_{fr}\int_{\lambda_{min}}^{\lambda_{max}}\left(A_{fr}(\lambda)\cdot I(\lambda)\cdot 2\phi_{fr\to r}(\lambda)\right)d\lambda + P_rP_{fr}\int_{\lambda_{min}}^{\lambda_{max}}\left(A_r(\lambda)\cdot I(\lambda)\cdot \phi_{r\to fr}(\lambda)\right)d\lambda \tag{1}$$

$$\frac{dP_rP_r}{dt} = -P_rP_r\int_{\lambda_{min}}^{\lambda_{max}}\left(A_r(\lambda)\cdot I(\lambda)\cdot 2\phi_{r\to fr}(\lambda)\right)d\lambda + P_rP_{fr}\int_{\lambda_{min}}^{\lambda_{max}}\left(A_{fr}(\lambda)\cdot I(\lambda)\cdot \phi_{fr\to r}(\lambda)\right)d\lambda \tag{2}$$

$$\begin{aligned}\frac{dP_rP_{fr}}{dt} = &-P_rP_{fr}\int_{\lambda_{min}}^{\lambda_{max}}\left(A_{fr}(\lambda)\cdot I(\lambda)\cdot \phi_{fr\to r}(\lambda)\right)d\lambda \\ &- P_rP_{fr}\int_{\lambda_{min}}^{\lambda_{max}}\left(A_r(\lambda)\cdot I(\lambda)\cdot \phi_{r\to fr}(\lambda)\right)d\lambda \\ &+ P_rP_r\int_{\lambda_{min}}^{\lambda_{max}}\left(A_r(\lambda)\cdot I(\lambda)\cdot 2\phi_{r\to fr}(\lambda)\right)d\lambda \\ &+ P_{fr}P_{fr}\int_{\lambda_{min}}^{\lambda_{max}}\left(A_{fr}(\lambda)\cdot I(\lambda)\cdot 2\phi_{fr\to r}(\lambda)\right)d\lambda\end{aligned} \tag{3}$$

Moreover, the total population of RfpA is constant, so that the P_rP_r state, P_rP_{fr} state, and the $P_{fr}P_{fr}$ state add up to 1 :

$$P_rP_r + P_rP_{fr} + P_{fr}P_{fr} = 1 \tag{4}$$

P_rP_r is the biologically active state in which both monomers have photo-switched upon excitation by a far-red light photon. P_rP_{fr} is a biologically inactive state in which only one of the two monomers has been activated by a far-red light photon. $P_{fr}P_{fr}$ is a biologically inactive state in which both monomers can be activated after photoswitching with far-red light photons. $A_{fr}(\lambda)$ and $A_r(\lambda)$ are the wavelength dependent absorbance of the P_{fr} and P_r states, respectively. $I(\lambda)$ is the wavelength-dependent irradiance expressed in $\mu E \cdot s^{-1} \cdot m^{-2}$. $\phi_{fr\to r}$ and $\phi_{r\to fr}$ are the wavelength-dependent quantum yields of conversion between both states.

The rate of Chl *f* synthesis was described by a sigmoid, in which the threshold T is the amount of P_rP_r at which the rate is 50% of the maximal rate, and the power n controls the slope of the sigmoid during the switch, approaching a switch-like behaviour for higher values of n.

$$\frac{dChl\ f}{dt} = k_{synth}\frac{P_rP_r^{\,n}}{P_rP_r^{\,n} + T^n} - k_{deg}Chl\ f \qquad (5)$$

The absolute amounts of Chl *a* and Chl *f* are never considered in the model; instead, only the ratio between Chl *a* and Chl *f* is used. Furthermore, since Chl *a* is a widely abundant substrate for the synthesis of Chl *f*, the synthesis rate k_{synth} of Chl *f* does not depend on the concentration of Chl *a*.

7.3 Discussion

7.3.1 Towards model calibration

The two simulations presented below (Fig. 7.9) illustrate the model's predictions when the phytochrome is initially in one of its fully converted forms, either $P_rP_r(0)=1$ or $P_{fr}P_{fr}(0)=1$, and is then exposed to different LEDs to trigger the forward or reverse photoacclimation process. The RfpA population presents a behaviour similar to that obtained *in vitro*, where illumination with an LED emitting light at a wavelength corresponding to an absorbance peak of RfpA induces complete conversion of the resulting absorption spectrum (Fig. 7.9).

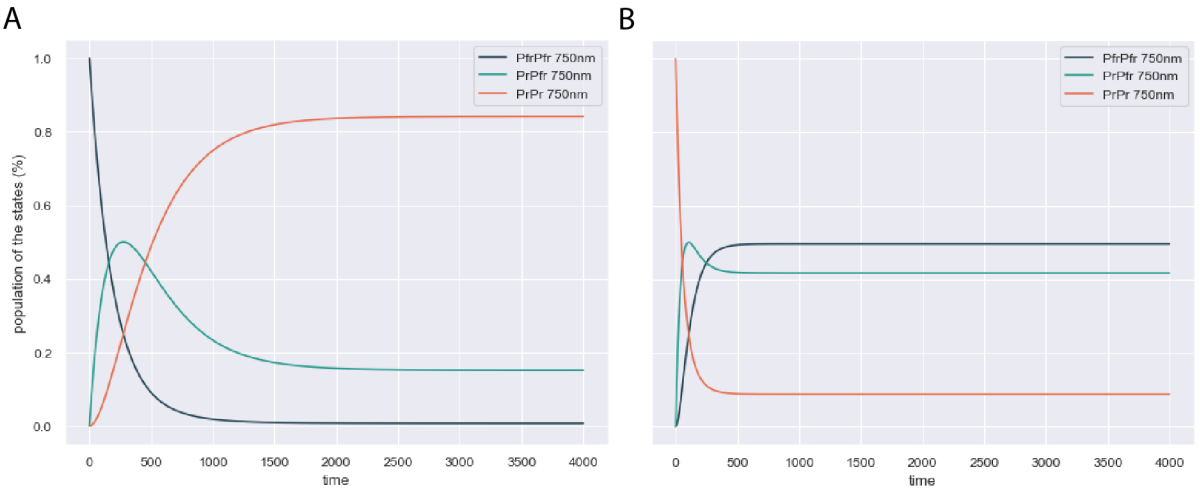


Fig. 7.9 – Predicted state populations from the RfpA model

A) Simulation obtained by setting the initial population of RfpA to be completely converted in the $P_{fr}P_{fr}$ state, and $I(\lambda)$ to be that of the 750 nm LED and $\phi_{fr\rightarrow r}$ and $2\phi_{r\rightarrow fr}$ to be unit quantum yield of conversion. B) Simulation obtained by setting the initial population of RfpA to be completely converted in the P_rP_r state, and $I(\lambda)$ to be that of the 630 nm LED and $\phi_{fr\rightarrow r}$ and $2\phi_{r\rightarrow fr}$ to be unit quantum yield of conversion.

In Fig. 7.9A, the population of the $P_{fr}P_{fr}$ state decreases exponentially while the P_rP_r population increases. Concurrently, the mixed state $P_{fr}P_r$ appears and its population eventually settles to around 20%. In Fig. 7.9B, the population of the P_rP_r decreases sharply to round 10%, while the populations of the P_rP_r and mixed state increase to around 50% and 40% respectively. This simulated behaviour aligns with experimental measurements that show a slower photoconversion rate from the state P_r to the P_{fr} state compared to the reverse (P_{fr} to P_r) reaction.

Nonetheless, the prediction of the model only qualitatively reflects the experimental observations, since most of the parameters of the model have been set arbitrarily due to the lack of calibration data. In future work, comparing the predictions of the RfpA model with experimental data on the isolated protein will be crucial for calibrating the quantum yields of conversion for the two forms. Dedicated wavelength-dependent experiments will furthermore provide experimental data on the two quantum yields of conversion.

After calibrating the parameters relative to RfpA's population dynamics, it should be possible to exploit the model to design specific experiments and further refine the values of n (steepness of activation) and T (threshold population of the active P_rP_r). By conducting growth experiments and measuring the relative amounts of Chl *f*, it should be possible to obtain the threshold population of the active phytochrome required to induce the of the FaRLiP cluster expression. Subsequently, investigating whether higher amounts of far-red light influence the rate of Chl production could provide information on the activation of Chl f synthesis.

After calibrating all the parameters of the RfpA model, it could be used to conduct a model-based analysis and investigate long-standing biological questions and common problems in the laboratory growth conditions of FaRLiP strains. One of the main problems involved in the study of far-red light photoacclimation is the scarcity of material, which is directly related to the low growth rate of these organisms under "pure" far-red light. The model could provide information about the optimal (time-dependent) light spectrum for maintaining the phytochrome in P_r state (i.e., maintaining the FaRLiP cluster activated) while maximizing growth at the same time.

Ultimately, the dual-light photobioreactor system (Section 7.2.1) will enable us to conduct growth experiments to test these optimised strategies *in vivo* and iteratively refine the RfpA model itself. Testing the effects of very short pulses as well as low background levels of white photons will provide information on the speed of the transition in more physiological conditions and on the mechanisms involved in Chl *f* synthesis and the remodelling of the thylakoid membrane. Moreover, new optimised light protocols will be

implemented into the current growth strategies to improve the preparation and isolation of photosystems.

Establishing the photochemical properties of RfpA is a first step towards designing *in vivo* growth experiments aimed at elucidating the effect of mixed light conditions and their temporal variation in the ecological niches that host FaRLiP species, and ultimately to help establish improved growth conditions for FaRLiP species. These ecological niches are thought to provide a less stable light environment compared to those that host the Chl *d* cyanobacterium *A. marina*. Bacteria and cyanobacteria are known to migrate rapidly in mats, and weather conditions also contribute to alter the amount of light available to Chl *f* cyanobacteria continuously (233,234). This distinction clarifies the main difference between these two long-wavelength evolutionary adaptations. While Chl d cyanobacteria can sacrifice resistance to photodamage, Chl *f* containing cyanobacteria have to maintain it while sacrificing nominal antenna size. For this reason, FaRLiP is intrinsically a "survival" adaptation, and it is plausible that have evolved to carefully evaluate the available light in quantity, quality and temporal availability. Moreover, aspects of the initial triggers of the transition could be elucidated. Super-rogue PSII (Chl *f* synthase) is probably powered by white light photons since there would be no Chl *f* to begin with, but it is possible for Chl *f* to be incorporated Chl *f* once its production is set and perform photochemistry on less energetic photons.

7.3.2 Further expression trial of full length RfpA

To completely characterise the RfpA type IV family of phytochromes, further photochemical and biochemical experiments are needed (i.e., determining the wavelength dependent quantum yield of conversion between the two states), and structural information on the complete holoprotein would be crucial to understand the molecular rearrangements that drive the transduction of the signal from light sensing to activity.

To obtain a more homogeneous protein sample, RfpA expression and purification need to be optimised. To this aim, expression in different *E. coli* strains resistant to the expression of toxic proteins will be tested (i.e., ArcticExpress, C43). Moreover, the expression vector of RfpA could be modified by decreasing the strength of the promoter to avoid

overexpression or by changing the inducer to stagger and delay the expression of the backbone in relation to the formation of PCB. Shorter expression windows, lower temperatures and precursor supplementation in the media could also be tested to improve solubility and limit inclusions and aggregation. If these simple approaches were to be unsuccessful, the host could be changed to a model cyanobacterium. *Synechocystis sp.* PCC 6803 would be an ideal candidate for the heterologous expression of RfpA, since PCB is naturally produced in large quantities ensuring a higher chromophore occupancy.

Conformational and structural heterogeneity of the sample hindered a satisfactory reconstruction map. SDS-PAGE, native PAGE, and spectroscopic measurements reveal that in the expression of F-RfpA most of the protein is captured into inclusion bodies, probably due to two main reasons: on one hand, the C-terminal histidine kinase (HK) domain is certainly toxic to *E. coli*, leading to cleavage by endogenous proteases and capture into inclusion bodies; on the other hand, since the PCB production pathway was engineered in *E. coli* starting from heme, the occupancy of the chromophore could be far lower than 100%, producing a misfolded protein population that could act as nucleation for aggregation events.

7.4 Conclusions

In this chapter, a protocol for the heterologous expression of RfpA, the phytochrome that triggers the activation of the FaRLiP cluster, has been presented. The purification of the intact protein still presents issues related to the occupancy of the bilin chromophore and to the poor solubility of the isolated protein. Nonetheless, further optimization of the purification protocol should allow to produce enough intact sample for the determination of the structure of the intact RfpA phytochrome.

This structure could provide insight into the unusual subunit arrangement of the Type IV knotless phytochromes and into the biological signal transduction following photoisomerization. By using the mathematical model of the RfpA photoconversion, it should be possible, in principle, to maximise the rates of conversion between both states, and consequently to solve the conformational rearrangements that follow the photoisomerization of the chromophore.

To finalise the mathematical modelling that includes both clusters triggering and growth of the organisms in far-red light, more experiments are needed to validate the model structure and calibrate its parameters. Once the model has been appropriately parametrised and calibrated, it could be used to answer fundamental biological questions surrounding the transition and the potential biotechnological use of FaRLiP.

The model could provide insight on Chl *f* synthesis, clarifying whether small amounts of white light can improve the rates of synthesis of Chl *f*. The synthesis mechanism of Chl *f* has not been elucidates, and it is not known whether this synthesis requires photons in the visible range or once the production has been kickstarted the enzyme can also function with far-red photons. By appropriately using the model to design growth experiments it should be possible to investigate this question and also to optimise growth conditions for the laboratory use of far-red-light-adapted cyanobacteria.

Concerning the biotechnological applications of Chl *f*-based photosynthesis, the mathematical model could start guiding the development process for dual-layer photobioreactors that use far-red-light-adapted organisms in the inner layer. Not only can

the model aid the creation of an optimal photobioreactor design, but it can also suggest operating conditions that could maximise growth rates in photobioreactors.

8 Conclusions and Outlook

Since the discovery of cyanobacteria capable of performing high efficiency photochemistry with photons beyond what was considered the “red limit of photosynthesis” (43), the field of photosynthesis research has been presented with significant opportunities to deepen our understanding the energy limits of photosynthesis. Far-red photosynthesis also potentially provides new avenues of research aimed at improving the productivity of photosynthetic organisms (90). From the initial isolation (91) to the first biophysical (43) and then structural studies (112), substantial progress has been made in our understanding of these new paradigms of photosynthesis, yet new questions have emerged at every stage. The combination of existing knowledge, unresolved questions, and the growing scientific community in this area has provided an ideal context for addressing fundamental challenges related to the physical limits of photosynthesis, and to apply state-of-the-art techniques in structural biology. The present thesis contributes to the field in several ways.

In Chapter 3, two methodologies were developed to classify chlorophyll substituents in Cryo-EM ESP maps: the improved cone scan and the sphere scan. The improved cone scan method builds on earlier frameworks by incorporating statistical comparisons of the ESP of chlorophyll substituents, allowing robust assignments. Conversely, the sphere scan method offers a more automated approach, combining PCA and supervised classification. While it is effective for certain substituents like hydrogen-bonded formyl groups, it lacks the interpretability of the cone scan method. Together, these complementary techniques offer a toolkit for advancing the accurate identification of Chl substituents. These methods have been used, together with other lines of evidence, to assign Chl *f* sites in the structures of FR-PSI and FR-PSII presented in the following chapters (Chapter 4 and 5).

In Chapter 4, two distinct FR-PSI structures are presented. These structures allowed the assignment of all eight expected Chl *f* sites, including one directly involved in photochemistry (A_{-1B}). Chl *f* A_{-1B} presents a novel sulfhydryl-π-formyl motif, which is expected to tune the redox potential of this site to facilitate photochemistry with Chl *f* as a primary donor. This finding challenges the prior structural work and mechanistic models

that considered Chl *f* to be only an antenna pigment in FR-PSI. Among the sites assigned, Chl *f* L03 provides insight into excitation energy transfer dynamics in FR-PSI trimers, by acting as a bridge for excitation energy to be transferred to adjacent FR-PSI monomers. In this chapter, tentative site-energy assignments for the proposed Chl *f* location are discussed, establishing the basis for targeted site-directed mutagenesis approaches to explore excitation energy transfer pathways and site-specific effects.

In Chapter 5, structures of FR-PSII, both in the intact and non-oxygen-evolving states, revealed the location of the 4 Chl *f* and the Chl *d* pigments, confirming some of the candidates present in the literature and revealing a new Chl *f* site in position B617. This site is of interest also for WL-PSII, since it has been demonstrated in this work to be responsible for its characteristic 693 nm low-temperature fluorescence emission shoulder. In the maps present, a novel subunit encoded in the FaRLiP cluster was presented. The helix has been named PsbH2' due to its extensive interactions with the far-red light specific subunit PsbH2. The phylogenetic diversity of this subunit is discussed together with its function, which is hypothesised to be related to the protection of specific Chl *f* sites or to the binding of FR-APC complexes.

Chapter 6 concludes the structural investigations of this thesis and presents the first intact structure of the bicylindrical FR-APC complex from *C. thermalis*. Phylogenetic analyses reveal amino acid changes that fine-tune the spectral properties of bilin pigments, allowing them to absorb in a broader spectral coverage compared to its visible light absorbing counterparts. The adaptive trade-offs in far-red-light-adapted organisms seem to indicate that spectral coverage and membrane compaction is prioritised over larger antenna cross-sections. Moreover, experimental pitfalls in the isolation of supercomplexes containing both phycobilisomes and photosystems are discussed, laying the groundwork for future work aimed at solving the FR-PSII – FR-APC supercomplex.

In Chapter 7, RfpA, the phytochrome that controls far-red light photoacclimation, has been heterologously expressed and characterised, providing an initial assessment of the condition that trigger this photoacclimation. Moreover, a mathematical model that describes its behaviour under different light conditions is presented.

Overall, this work provides a comprehensive exploration of the structural and mechanistic adaptations that enable far-red light photoacclimation. The availability of structures for all of the far-red-light-adapted light driven complexes, FR-APC, FR-PSI, and FR-PSII, provides a new set of information to model excitation energy transfer on the mesoscale, and advances our understanding of the specific evolutionary adaptations in the far-red-light-adapted photosynthetic apparatus, contributing to the body of knowledge needed to lay the foundation for potential future applications in bioengineering and agriculture.

The prospect of introducing Chl *f* into the lower leaves of the canopy in crop plants seems remote, and its practical benefits are highly debatable, particularly considering the slow growth observed in far-red light adapted organisms, which likely points to a survival strategy. Nonetheless, this once again reminds us of the important of basic research, not only in driving technological improvement, but crucially, also in providing insights on which ideas hold genuine technological promise.

9 Bibliography

1. Nickelsen J, Rengstl B. Photosystem II Assembly: From Cyanobacteria to Plants. *Annual Review of Plant Biology*. 2013;64(1): 609–635. https://doi.org/10.1146/annurev-arplant-050312-120124.

2. van Grondelle R, Dekker JP, Gillbro T, Sundstrom V. Energy transfer and trapping in photosynthesis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 1994;1187(1): 1–65. https://doi.org/10.1016/0005-2728(94)90166-X.

3. Niyogi KK, Truong TB. Evolution of flexible non-photochemical quenching mechanisms that regulate light harvesting in oxygenic photosynthesis. *Current Opinion in Plant Biology*. 2013;16(3): 307–314. https://doi.org/10.1016/j.pbi.2013.03.011.

4. Prywes N, Phillips NR, Tuck OT, Valentin-Alvarado LE, Savage DF. Rubisco Function, Evolution, and Engineering. *Annual Review of Biochemistry*. 2023;92(1): 385–410. https://doi.org/10.1146/annurev-biochem-040320-101244.

5. Umena Y, Kawakami K, Shen JR, Kamiya N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature*. 2011;473(7345): 55–60. https://doi.org/10.1038/nature09913.

6. Kawakami K, Umena Y, Kamiya N, Shen JR. Structure of the catalytic, inorganic core of oxygen-evolving photosystem II at 1.9 Å resolution. *Journal of Photochemistry and Photobiology B: Biology*. 2011;104(1–2): 9–18. https://doi.org/10.1016/j.jphotobiol.2011.03.017.

7. Lubitz W, Chrysina M, Cox N. Water oxidation in photosystem II. *Photosynthesis Research*. 2019;142(1): 105–125. https://doi.org/10.1007/s11120-019-00648-3.

8. Lu Y. Identification and Roles of Photosystem II Assembly, Stability, and Repair Factors in Arabidopsis. *Frontiers in Plant Science*. 2016;7(FEB2016). https://doi.org/10.3389/fpls.2016.00168.

9. Svensson B, Vass I, Styring S. Sequence Analysis of the D1 and D2 Reaction Center Proteins of Photosystem II. *Zeitschrift für Naturforschung C*. 1991;46(9–10): 765–776. https://doi.org/10.1515/znc-1991-9-1008.

10. Oliver T, Sánchez-Baracaldo P, Larkum AW, Rutherford AW, Cardona T. Time-resolved comparative molecular evolution of oxygenic photosynthesis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2021;1862(6): 148400. https://doi.org/https://doi.org/10.1016/j.bbabio.2021.148400.

11. Cardona T, Sedoud A, Cox N, Rutherford AW. Charge separation in Photosystem II: A comparative and evolutionary overview. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2012;1817(1): 26–43. https://doi.org/https://doi.org/10.1016/j.bbabio.2011.07.012.

12. Diner BA, Rappaport F. Structure, dynamics, and energetics of the primary photochemistry of Photosystem II of oxygenic photosynthesis. *Annual Review of Plant Biology*. 2002;53(Volume 53, 2002): 551–580. https://doi.org/10.1146/annurev.arplant.53.100301.135238.

13. Dekker JP, Van Grondelle R. Primary charge separation in Photosystem II. *Photosynthesis Research*. 2000;63(3): 195–208. https://doi.org/10.1023/A:1006468024245.

14. Rutherford AW, Osyczka A, Rappaport F. Back-reactions, short-circuits, leaks and other energy wasteful reactions in biological electron transfer: Redox tuning to survive life in O2. *FEBS Letters*. 2012;586(5): 603–616. https://doi.org/10.1016/j.febslet.2011.12.039.

15. Rappaport F, Guergova-Kuras M, Nixon PJ, Diner BA, Lavergne J. Kinetics and Pathways of Charge Recombination in Photosystem II†. *Biochemistry*. 2002;41(26): 8518–8527. https://doi.org/10.1021/BI025725P.

16. de Causmaecker S, Douglass JS, Fantuzzi A, Nitschke W, Rutherford AW. Energetics of the exchangeable quinone, QB, in Photosystem II. *Proceedings of the*

National Academy of Sciences. 2019;116(39): 19458–19463. https://doi.org/10.1073/pnas.1910675116.

17. Saito K, Rutherford AW, Ishikita H. Mechanism of proton-coupled quinone reduction in Photosystem II. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(3): 954–959. https://doi.org/10.1073/pnas.1212957110.

18. Shen JR. The Structure of Photosystem II and the Mechanism of Water Oxidation in Photosynthesis. *Annual Review of Plant Biology*. 2015;66(1): 23–48. https://doi.org/10.1146/annurev-arplant-050312-120129.

19. Bhowmick A, Hussein R, Bogacz I, Simon PS, Ibrahim M, Chatterjee R, et al. Structural evidence for intermediates during O2 formation in photosystem II. *Nature*. 2023;617(7961): 629–636. https://doi.org/10.1038/s41586-023-06038-z.

20. Greife P, Schönborn M, Capone M, Assunção R, Narzi D, Guidoni L, et al. The electron–proton bottleneck of photosynthetic oxygen evolution. *Nature*. 2023;617(7961): 623–628. https://doi.org/10.1038/s41586-023-06008-5.

21. Kok B, Forbush B, McGloin M. Cooperation of charges in photosynthetic O_2 evolution – A linear four step mechanism. *Photochemistry and Photobiology*. 1970;11(6): 457–475. https://doi.org/10.1111/j.1751-1097.1970.tb06017.x.

22. Joliot P, Kok B. Oxygen evolution in photosynthesis. In: *Energetics of Photosynthesis*. Elsevier; 1975. p. 387–412.

23. Chitnis PR. Photosystem I: Function and Physiology. *Annual Review of Plant Physiology and Plant Molecular Biology*. 2001;52(1): 593–626. https://doi.org/10.1146/annurev.arplant.52.1.593.

24. Golbeck JH. Photosystem I. The Light-Driven Plastocyanin:Ferredoxin Oxidoreductase. *Photosynthetica*. 2007;45(4): 488–488. https://doi.org/10.1007/s11099-007-0083-4.

25. Chen M, Liu X, He Y, Li N, He J, Zhang Y. Diversity Among Cyanobacterial Photosystem I Oligomers. *Frontiers in Microbiology*. 2022;12: 781826. https://doi.org/10.3389/FMICB.2021.781826.

26. Baymann F, Brugna M, Mühlenhoff U, Nitschke W. Daddy, where did (PS)I come from? *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2001;1507(1): 291–310. https://doi.org/10.1016/S0005-2728(01)00209-2.

27. Fromme P, Mathis P. Unraveling the Photosystem I Reaction Center: A History, or the Sum of Many Efforts. *Photosynthesis Research*. 2004;80(1): 109–124. https://doi.org/10.1023/B:PRES.0000030657.88242.e1.

28. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauß N. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature*. 2001;411(6840): 909–917. https://doi.org/10.1038/35082000.

29. Cherepanov DA, Semenov AY, Mamedov MD, Aybush A V, Gostev FE, Shelaev I V, et al. Current state of the primary charge separation mechanism in photosystem I of cyanobacteria. *Biophysical Reviews*. 2022;14(4): 805–820. https://doi.org/10.1007/s12551-022-00983-1.

30. Pålsson LO, Flemming C, Gobets B, van Grondelle R, Dekker JP, Schlodder E. Energy Transfer and Charge Separation in Photosystem I: P700 Oxidation Upon Selective Excitation of the Long-Wavelength Antenna Chlorophylls of Synechococcus elongatus. *Biophysical Journal*. 1998;74(5): 2611–2622. https://doi.org/10.1016/S0006-3495(98)77967-6.

31. Santabarbara S, Kuprov I, Hore PJ, Casal A, Heathcote P, Evans MCW. Analysis of the Spin-Polarized Electron Spin Echo of the [P700+A1-] Radical Pair of Photosystem I Indicates That Both Reaction Center Subunits Are Competent in Electron Transfer in Cyanobacteria, Green Algae, and Higher Plants. *Biochemistry*. 2006;45(23): 7389–7403. https://doi.org/10.1021/bi060330h.

32. Joliot P, Joliot A. In Vivo Analysis of the Electron Transfer within Photosystem I: Are the Two Phylloquinones Involved? *Biochemistry*. 1999;38(34): 11130–11136. https://doi.org/10.1021/bi990857c.

33. Mitsuhashi K, Tamura H, Saito K, Ishikita H. Nature of Asymmetric Electron Transfer in the Symmetric Pathways of Photosystem I. *The Journal of Physical Chemistry B*. 2021;125(11): 2879–2885. https://doi.org/10.1021/acs.jpcb.0c10885.

34. Toporik H, Dobson Z, Mazor Y. The Photosystem I Complex of Oxygenic Photosynthesis. 2021; https://doi.org/10.1016/b978-0-12-819460-7.00253-x.

35. Mirkovic T, Ostroumov EE, Anna JM, van Grondelle R, Govindjee, Scholes GD. Light Absorption and Energy Transfer in the Antenna Complexes of Photosynthetic Organisms. *Chemical Reviews*. 2017;117(2): 249–293. https://doi.org/10.1021/acs.chemrev.6b00002.

36. Cheng YC, Fleming GR. Dynamics of Light Harvesting in Photosynthesis. *Annual Review of Physical Chemistry*. 2009;60(1): 241–262. https://doi.org/10.1146/annurev.physchem.040808.090259.

37. Croce R, Van Grondelle R, Van Amerongen H, Van Stokkum I. *Light Harvesting in Photosynthesis*. Boca Raton: Taylor & Francis/CRC Press, 2017. | Series: Foundations of biochemistry and biophysics: CRC Press; 2018. https://doi.org/10.1201/9781351242899.

38. Forster Th. Energiewanderung und Fluoreszenz. *Die Naturwissenschaften*. 1946;33(6): 166–175. https://doi.org/10.1007/BF00585226.

39. Förster T. Energy migration and fluorescence. *Journal of Biomedical Optics*. 2012;17(1): 011002. https://doi.org/10.1117/1.JBO.17.1.011002.

40. Tan LM, Yu J, Kawakami T, Kobayashi M, Wang P, Wang-Otomo ZY, et al. New Insights into the Mechanism of Uphill Excitation Energy Transfer from Core Antenna to Reaction Center in Purple Photosynthetic Bacteria. *The Journal of*

Physical Chemistry Letters. 2018;9(12): 3278–3284. https://doi.org/10.1021/acs.jpclett.8b01197.

41. Schmitt FJ, Hüls A, Moldenhauer M, Friedrich T. How electron tunneling and uphill excitation energy transfer support photochemistry in Halomicronema hongdechloris. *Photosynthesis Research*. 2024;159(2–3): 273–289. https://doi.org/10.1007/s11120-023-01064-4.

42. Kosugi M, Kawasaki M, Shibata Y, Hara K, Takaichi S, Moriya T, et al. Uphill energy transfer mechanism for photosynthesis in an Antarctic alga. *Nature Communications*. 2023;14(1): 730. https://doi.org/10.1038/s41467-023-36245-1.

43. Nürnberg DJ, Morton J, Santabarbara S, Telfer A, Joliot P, Antonaru LA, et al. Photochemistry beyond the red limit in chlorophyll f–containing photosystems. *Science*. 2018;360(6394): 1210–1213. https://doi.org/10.1126/science.aar8313.

44. Saito K, Tamura H, Ishikita H. Superexchange Electron Transfer and Protein Matrix in the Charge-Separation Process of Photosynthetic Reaction Centers. *The Journal of Physical Chemistry Letters*. 2024;15(36): 9183–9192. https://doi.org/10.1021/acs.jpclett.4c02232.

45. Raszewski G, Renger T. Light Harvesting in Photosystem II Core Complexes Is Limited by the Transfer to the Trap: Can the Core Complex Turn into a Photoprotective Mode? *Journal of the American Chemical Society*. 2008;130(13): 4431–4446. https://doi.org/10.1021/ja7099826.

46. Braslavsky SE, Fron E, Rodríguez HB, Román ES, Scholes GD, Schweitzer G, et al. Pitfalls and limitations in the practical use of Förster's theory of resonance energy transfer. *Photochemical & Photobiological Sciences*. 2008;7(12): 1444–1448. https://doi.org/10.1039/b810620g.

47. Ruban A V. Evolution under the sun: optimizing light harvesting in photosynthesis. *Journal of Experimental Botany*. 2015;66(66): 7–23. https://doi.org/10.1093/jxb/eru400.

48. Pan X, Cao P, Su X, Liu Z, Li M. Structural analysis and comparison of light-harvesting complexes I and II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2020;1861(4): 148038. https://doi.org/10.1016/j.bbabio.2019.06.010.

49. Jansson S. The light-harvesting chlorophyll ab-binding proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 1994;1184(1): 1–19. https://doi.org/https://doi.org/10.1016/0005-2728(94)90148-1.

50. Bryant DA, Gisriel CJ. The structural basis for light harvesting in organisms producing phycobiliproteins. *The Plant Cell*. 2024;36(10): 4036–4064. https://doi.org/10.1093/plcell/koae126.

51. Staheli CF, Barney J, Clark TR, Bowles M, Jeppesen B, Oblinsky DG, et al. Spectroscopic and Photophysical Investigation of Model Dipyrroles Common to Bilins: Exploring Natural Design for Steering Torsion to Divergent Functions. *Frontiers in Chemistry*. 2021;9: 628852.

52. Zhang J, Ma J, Liu D, Qin S, Sun S, Zhao J, et al. Structure of phycobilisome from the red alga Griffithsia pacifica. *Nature*. 2017;551(7678): 57–63. https://doi.org/10.1038/nature24278.

53. Gantt E, Conti SF. Ultrastructure of Blue-Green Algae. *Journal of Bacteriology*. 1969;97(3): 1486–1493. https://doi.org/10.1128/jb.97.3.1486-1493.1969.

54. Gantt E, Lipschultz CA. Structure and phycobiliprotein composition of phycobilisomes from Griffithsia pacifica. *Journal of Phycology*. 1980;16(3): 394–398. https://doi.org/10.1111/j.1529-8817.1980.tb03051.x.

55. Bryant DA, Glazer AN, Eiserling FA. Characterization and structural properties of the major biliproteins of Anabaena sp. *Archives of Microbiology*. 1976;110(1): 61–75. https://doi.org/10.1007/BF00416970.

56. Guglielmi G, Cohen-Bazire G, Bryant DA. The structure of Gloeobacter violaceus and its phycobilisomes. *Archives of microbiology*. 1981;129: 181–189.

57. Jiang HW, Wu HY, Wang CH, Yang CH, Ko JT, Ho HC, et al. A structure of the relict phycobilisome from a thylakoid-free cyanobacterium. *Nature Communications*. 2023;14(1): 8009. https://doi.org/10.1038/s41467-023-43646-9.

58. Gisriel CJ, Shen G, Brudvig GW, Bryant DA. Structure of the antenna complex expressed during far-red light photoacclimation in Synechococcus sp. PCC 7335. *Journal of Biological Chemistry*. 2024;300(2): 105590. https://doi.org/https://doi.org/10.1016/j.jbc.2023.105590.

59. Rakhimberdieva MG, Boichenko VA, Karapetyan N V., Stadnichuk IN. Interaction of Phycobilisomes with Photosystem II Dimers and Photosystem I Monomers and Trimers in the Cyanobacterium Spirulina platensis. *Biochemistry*. 2001;40(51): 15780–15788. https://doi.org/10.1021/bi010009t.

60. Watanabe M, Semchonok DA, Webber-Birungi MT, Ehira S, Kondo K, Narikawa R, et al. Attachment of phycobilisomes in an antenna–photosystem I supercomplex of cyanobacteria. *Proceedings of the National Academy of Sciences*. 2014;111(7): 2512–2517. https://doi.org/10.1073/pnas.1320599111.

61. Zhang X, Xiao Y, You X, Sun S, Sui SF. In situ structural determination of cyanobacterial phycobilisome–PSII supercomplex by STAgSPA strategy. *Nature Communications 2024 15:1*. 2024;15(1): 1–11. https://doi.org/10.1038/s41467-024-51460-0.

62. Zlenko D V, Elanskaya I V, Lukashev EP, Bolychevtseva Y V, Suzina NE, Pojidaeva ES, et al. Role of the PB-loop in ApcE and phycobilisome core function in cyanobacterium Synechocystis sp. PCC 6803. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*. 2019;1860(2): 155–166.

63. Zheng Z, Li X, Wei P, Zhang X, Zhang T, Zhang Z, et al. Molecular glue for phycobilisome attachment to photosystem II in Synechococcus sp. PCC 7002. *Proceedings of the National Academy of Sciences*. 2025;122(4). https://doi.org/10.1073/pnas.2415222122.

64. Espinoza-Corral R, Iwai M, Zavřel T, Lechno-Yossef S, Sutter M, Červený J, et al. Phycobilisome protein ApcG interacts with PSII and regulates energy transfer in Synechocystis. *Plant Physiology*. 2024;194(3): 1383–1396. https://doi.org/10.1093/plphys/kiad615.

65. Akhtar P, Balog-Vig F, Han W, Li X, Han G, Shen JR, et al. Quantifying the Energy Spillover between Photosystems II and I in Cyanobacterial Thylakoid Membranes and Cells. *Plant And Cell Physiology*. 2024;65(1): 95–106. https://doi.org/10.1093/pcp/pcad127.

66. Björn LO, Papageorgiou GC, Blankenship RE, Govindjee. A viewpoint: Why chlorophyll a? *Photosynthesis Research*. 2009;99(2): 85–98. https://doi.org/10.1007/s11120-008-9395-x.

67. Lichtlé C. Effects of nitrogen deficiency and light of high intensity on Cryptomonas rufescens (Cryptophyceae). *Protoplasma*. 1979;101(4): 283–299. https://doi.org/10.1007/BF01276969.

68. Arnon DI, McSwain BD, Tsujimoto HY, Wada K. Photochemical activity and components of membrane preparations from blue-green algae. I. Coexistence of two photosystems in relation to chlorophyll a and removal of phycocyanin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 1974;357(2): 231–245. https://doi.org/https://doi.org/10.1016/0005-2728(74)90063-2.

69. Mullineaux CW, Liu LN. Membrane Dynamics in Phototrophic Bacteria. *Annual Review of Microbiology*. 2020;74(1): 633–654. https://doi.org/10.1146/annurev-micro-020518-120134.

70. Perez-Boerema A, Engel BD, Wietrzynski W. Evolution of Thylakoid Structural Diversity. *Annual Review of Cell and Developmental Biology*. 2024;40(1): 169–193. https://doi.org/10.1146/annurev-cellbio-120823-022747.

71. Olive J, Ajlani G, Astier C, Recouvreur M, Vernotte C. Ultrastructure and light adaptation of phycobilisome mutants of Synechocystis PCC 6803. *Biochimica et*

Biophysica Acta (BBA) - Bioenergetics. 1997;1319(2–3): 275–282. https://doi.org/10.1016/S0005-2728(96)00168-5.

72. Casella S, Huang F, Mason D, Zhao GY, Johnson GN, Mullineaux CW, et al. Dissecting the Native Architecture and Dynamics of Cyanobacterial Photosynthetic Machinery. *Molecular Plant*. 2017;10(11): 1434–1448. https://doi.org/10.1016/j.molp.2017.09.019.

73. Rast A, Schaffer M, Albert S, Wan W, Pfeffer S, Beck F, et al. Biogenic regions of cyanobacterial thylakoids form contact sites with the plasma membrane. *Nature Plants*. 2019;5(4): 436–446. https://doi.org/10.1038/s41477-019-0399-7.

74. Sacharz J, Bryan SJ, Yu J, Burroughs NJ, Spence EM, Nixon PJ, et al. Sub-cellular location of FtsH proteases in the cyanobacterium Synechocystis sp. 6803 suggests localised PSII repair zones in the thylakoid membranes. *Molecular Microbiology*. 2015;96(3): 448–462. https://doi.org/10.1111/mmi.12940.

75. MacGregor-Chatwin C, Sener M, Barnett SFH, Hitchcock A, Barnhart-Dailey MC, Maghlaoui K, et al. Lateral Segregation of Photosystem I in Cyanobacterial Thylakoids. *The Plant Cell*. 2017;29(5): 1119–1136. https://doi.org/10.1105/tpc.17.00071.

76. Block MA, Dorne AJ, Joyard J, Douce R. Preparation and characterization of membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts. I. Electrophoretic and immunochemical analyses. *Journal of Biological Chemistry*. 1983;258(21): 13273–13280. https://doi.org/10.1016/S0021-9258(17)44112-3.

77. Garab G, Yaguzhinsky LS, Dlouhý O, Nesterov S V., Špunda V, Gasanoff ES. Structural and functional roles of non-bilayer lipid phases of chloroplast thylakoid membranes and mitochondrial inner membranes. *Progress in Lipid Research*. 2022;86: 101163. https://doi.org/10.1016/j.plipres.2022.101163.

78. Zheng L, Li Y, Li X, Zhong Q, Li N, Zhang K, et al. Structural and functional insights into the tetrameric photosystem I from heterocyst-forming cyanobacteria. *Nature Plants*. 2019;5(10): 1087–1097. https://doi.org/10.1038/s41477-019-0525-6.

79. Silva SC, Ferreira ICFR, Dias MM, Barreiro MF. Microalgae-Derived Pigments: A 10-Year Bibliometric Review and Industry and Market Trend Analysis. *Molecules*. 2020;25(15): 3406. https://doi.org/10.3390/molecules25153406.

80. Vigani M, Parisi C, Rodríguez-Cerezo E, Barbosa MJ, Sijtsma L, Ploeg M, et al. Food and feed products from micro-algae: Market opportunities and challenges for the EU. *Trends in Food Science & Technology*. 2015;42(1): 81–92.

81. Shah MR, Lutzu GA, Alam A, Sarker P, Kabir Chowdhury MA, Parsaeimehr A, et al. Microalgae in aquafeeds for a sustainable aquaculture industry. *Journal of Applied Phycology*. 2018;30(1): 197–213. https://doi.org/10.1007/s10811-017-1234-z.

82. Mofijur M, Rasul MG, Hassan NMS, Nabi MN. Recent Development in the Production of Third Generation Biodiesel from Microalgae. *Energy Procedia*. 2019;156: 53–58. https://doi.org/https://doi.org/10.1016/j.egypro.2018.11.088.

83. Dolganyuk V, Belova D, Babich O, Prosekov A, Ivanova S, Katserov D, et al. Microalgae: A Promising Source of Valuable Bioproducts. *Biomolecules*. 2020;10(8): 1153. https://doi.org/10.3390/biom10081153.

84. Mohsenpour SF, Hennige S, Willoughby N, Adeloye A, Gutierrez T. Integrating micro-algae into wastewater treatment: A review. *Science of The Total Environment*. 2021;752: 142168. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.142168.

85. Lardon L, Hélias A, Sialve B, Steyer JP, Bernard O. Life-Cycle Assessment of Biodiesel Production from Microalgae. *Environmental Science & Technology*. 2009;43(17): 6475–6481. https://doi.org/10.1021/es900705j.

86. Cotton CAR, Douglass JS, De Causmaecker S, Brinkert K, Cardona T, Fantuzzi A, et al. Photosynthetic Constraints on Fuel from Microbes. *Frontiers in Bioengineering and Biotechnology*. 2015;3. https://doi.org/10.3389/fbioe.2015.00036.

87. Morillas-España A, Lafarga T, Sánchez-Zurano A, Acién-Fernández FG, Rodríguez-Miranda E, Gómez-Serrano C, et al. Year-long evaluation of microalgae production in wastewater using pilot-scale raceway photobioreactors: Assessment of biomass productivity and nutrient recovery capacity. *Algal Research*. 2021;60: 102500. https://doi.org/https://doi.org/10.1016/j.algal.2021.102500.

88. Bellan A, Bucci F, Perin G, Alboresi A, Morosinotto T. Photosynthesis Regulation in Response to Fluctuating Light in the Secondary Endosymbiont Alga Nannochloropsis gaditana. *Plant and Cell Physiology*. 2020;61(1): 41–52. https://doi.org/10.1093/pcp/pcz174.

89. Davis GA, Kanazawa A, Schöttler MA, Kohzuma K, Froehlich JE, Rutherford AW, et al. Limitations to photosynthesis by proton motive force-induced photosystem II photodamage. *eLife*. 2016;5. https://doi.org/10.7554/eLife.16921.

90. Chen M, Blankenship RE. Expanding the solar spectrum used by photosynthesis. *Trends in Plant Science*. 2011;16(8): 427–431. https://doi.org/https://doi.org/10.1016/j.tplants.2011.03.011.

91. Chen M, Schliep M, Willows RD, Cai ZL, Neilan BA, Scheer H. A red-shifted chlorophyll. *Science*. 2010;329(5997): 1318–1319.

92. Miyashita H, Ikemoto H, Kurano N, Adachi K, Chihara M, Miyachi S. Chlorophyll d as a major pigment. *Nature*. 1996;383(6599): 402. https://doi.org/10.1038/383402a0.

93. Hoober JK, Eggink LL, Chen M. Chlorophylls, ligands and assembly of light-harvesting complexes in chloroplasts. *Photosynthesis Research*. 2007;94(2): 387–400. https://doi.org/10.1007/s11120-007-9181-1.

94. Shpilyov A V, Zinchenko V V, Grimm B, Lokstein H. Chlorophyll a phytylation is required for the stability of photosystems I and II in the cyanobacterium Synechocystis sp. PCC 6803. *The Plant Journal*. 2013;73(2): 336–346. https://doi.org/https://doi.org/10.1111/tpj.12044.

95. Scheer H. An Overview of Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications. In: Grimm B, Porra RJ, Rüdiger W, Scheer H (eds.) *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*. Dordrecht: Springer Netherlands; 2006. p. 1–26. https://doi.org/10.1007/1-4020-4516-6_1.

96. Li Y, Scales N, Blankenship RE, Willows RD, Chen M. Extinction coefficient for red-shifted chlorophylls: Chlorophyll d and chlorophyll f. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2012;1817(8): 1292–1298. https://doi.org/https://doi.org/10.1016/j.bbabio.2012.02.026.

97. Li Y, Cai ZL, Chen M. Spectroscopic Properties of Chlorophyll f. *The Journal of Physical Chemistry B*. 2013;117(38): 11309–11317. https://doi.org/10.1021/jp402413d.

98. Viola S, Roseby W, Santabarbara S, Nürnberg D, Assunção R, Dau H, et al. Impact of energy limitations on function and resilience in long-wavelength Photosystem II. Kramer DM, Kornmann B, Huang J, Romero E (eds.) *eLife*. 2022;11: e79890. https://doi.org/10.7554/eLife.79890.

99. Murakami A, Miyashita H, Iseki M, Adachi K, Mimuro M. Chlorophyll d in an Epiphytic Cyanobacterium of Red Algae. *Science*. 2004;303(5664): 1633–1633. https://doi.org/10.1126/science.1095459.

100. Ohashi S, Miyashita H, Okada N, Iemura T, Watanabe T, Kobayashi M. Unique photosystems in Acaryochloris marina. *Photosynthesis Research*. 2008;98(1–3): 141–149. https://doi.org/10.1007/s11120-008-9383-1.

101. Loughlin P, Lin Y, Chen M. Chlorophyll d and Acaryochloris marina: Current status. *Photosynthesis Research*. 2013;116(2–3): 277–293. https://doi.org/10.1007/S11120-013-9829-Y.

102. Petrova AA, Casazza AP, Shelaev I V., Gostev FE, Aybush A V., Nadtochenko VA, et al. Role of pheophytin a in the primary charge separation of photosystem I from

Acaryochloris marina: Femtosecond optical studies of excitation energy and electron transfer reactions. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2023;1864(3): 148984. https://doi.org/10.1016/j.bbabio.2023.148984.

103. Hamaguchi T, Kawakami K, Shinzawa-Itoh K, Inoue-Kashino N, Itoh S, Ifuku K, et al. Structure of the far-red light utilizing photosystem I of Acaryochloris marina. *Nature Communications*. 2021;12(1): 2333. https://doi.org/10.1038/s41467-021-22502-8.

104. Hu Q, Miyashita H, Iwasaki I, Kurano N, Miyachi S, Iwaki M, et al. A photosystem I reaction center driven by chlorophyll d in oxygenic photosynthesis. *Proceedings of the National Academy of Sciences*. 1998;95(22): 13319–13323. https://doi.org/10.1073/pnas.95.22.13319.

105. Antonaru LA, Cardona T, Larkum AWD, Nürnberg DJ. Global distribution of a chlorophyll f cyanobacterial marker. *The ISME Journal*. 2020;14(9): 2275–2287. https://doi.org/10.1038/s41396-020-0670-y.

106. Gan F, Zhang S, Rockwell NC, Martin SS, Lagarias JC, Bryant DA. Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light. *Science*. 2014;345(6202): 1312–1317. https://doi.org/10.1126/science.1256963.

107. Trinugroho JP, Bečková M, Shao S, Yu J, Zhao Z, Murray JW, et al. Chlorophyll f synthesis by a super-rogue photosystem II complex. *Nature plants*. 2020;6(3): 238–244.

108. Ming-Yang H, Gaozhong S, P CD, Chi Z, A BD. Light-dependent chlorophyll f synthase is a highly divergent paralog of PsbA of photosystem II. *Science*. 2016;353(6302): aaf9178. https://doi.org/10.1126/science.aaf9178.

109. Zhao C, Gan F, Shen G, Bryant DA. RfpA, RfpB, and RfpC are the Master Control Elements of Far-Red Light Photoacclimation (FaRLiP). *Frontiers in Microbiology*. 2015;6: 1303. https://doi.org/10.3389/fmicb.2015.01303.

110. Ho MY, Gan F, Shen G, Zhao C, Bryant DA. Far-red light photoacclimation (FaRLiP) in Synechococcus sp. PCC 7335: I. Regulation of FaRLiP gene expression. *Photosynthesis Research*. 2017;131(2): 173–186. https://doi.org/10.1007/s11120-016-0309-z.

111. Gisriel CJ, Shen G, Flesher DA, Kurashov V, Golbeck JH, Brudvig GW, et al. Structure of a dimeric photosystem II complex from a cyanobacterium acclimated to far-red light. *Journal of Biological Chemistry*. 2023;299(1): 102815. https://doi.org/10.1016/j.jbc.2022.102815.

112. Gisriel C, Shen G, Kurashov V, Ho MY, Zhang S, Williams D, et al. The structure of Photosystem I acclimated to far-red light illuminates an ecologically important acclimation process in photosynthesis. *Science Advances*. 2020; https://doi.org/10.1126/sciadv.aay6415.

113. Kaucikas M, Nürnberg D, Dorlhiac G, Rutherford AW, van Thor JJ. Femtosecond Visible Transient Absorption Spectroscopy of Chlorophyll f-Containing Photosystem I. *Biophysical Journal*. 2017;112(2): 234–249. https://doi.org/https://doi.org/10.1016/j.bpj.2016.12.022.

114. Mascoli V, Bhatti AF, Bersanini L, van Amerongen H, Croce R. The antenna of far-red absorbing cyanobacteria increases both absorption and quantum efficiency of Photosystem II. *Nature Communications*. 2022;13(1): 3562. https://doi.org/10.1038/s41467-022-31099-5.

115. Mascoli V, Bersanini L, Croce R. Far-red absorption and light-use efficiency trade-offs in chlorophyll f photosynthesis. *Nature Plants*. 2020;6(8): 1044–1053. https://doi.org/10.1038/s41477-020-0718-z.

116. Kato K, Shinoda T, Nagao R, Akimoto S, Suzuki T, Dohmae N, et al. Structural basis for the adaptation and function of chlorophyll f in photosystem I. *Nature Communications*. 2020;11(1): 238. https://doi.org/10.1038/s41467-019-13898-5.

117. Hastings G, Makita H, Agarwala N, Rohani L, Shen G, Bryant DA. Fourier transform visible and infrared difference spectroscopy for the study of P700 in photosystem I from Fischerella thermalis PCC 7521 cells grown under white light and far-red light: Evidence that the A–1 cofactor is chlorophyll f. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2019;1860(6): 452–460. https://doi.org/10.1016/j.bbabio.2019.04.002.

118. Cherepanov DA, Shelaev I V, Gostev FE, Aybush A V, Mamedov MD, Shen G, et al. Evidence that chlorophyll f functions solely as an antenna pigment in far-red-light photosystem I from Fischerella thermalis PCC 7521. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2020;1861(5): 148184. https://doi.org/https://doi.org/10.1016/j.bbabio.2020.148184.

119. Zouni A, Witt HT, Kern J, Fromme P, Krauss N, Saenger W, et al. Crystal structure of photosystem II from Synechococcus elongatus at 3.8 Å resolution. *Nature*. 2001;409(6821): 739–743. https://doi.org/10.1038/35055589.

120. Ferreira KN, Iverson TM, Maghlaoui K, Barber J, Iwata S. Architecture of the Photosynthetic Oxygen-Evolving Center. *Science*. 2004;303(5665): 1831–1838. https://doi.org/10.1126/science.1093087.

121. Loll B, Kern J, Saenger W, Zouni A, Biesiadka J. Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature*. 2005;438(7070): 1040–1044. https://doi.org/10.1038/nature04224.

122. Kamiya N, Shen JR. Crystal structure of oxygen-evolving photosystem II from Thermosynechococcus vulcanus at 3.7-Å resolution. *Proceedings of the National Academy of Sciences*. 2003;100(1): 98–103. https://doi.org/10.1073/pnas.0135651100.

123. Kühlbrandt W. The Resolution Revolution. *Science*. 2014;343(6178): 1443–1444. https://doi.org/10.1126/science.1251652.

124. Cheng Y. Single-particle cryo-EM—How did it get here and where will it go. *Science*. 2018;361(6405): 876–880. https://doi.org/10.1126/science.aat4346.

125. Burton-Smith RN, Watanabe A, Tokutsu R, Song C, Murata K, Minagawa J. Structural determination of the large photosystem II–light-harvesting complex II supercomplex of Chlamydomonas reinhardtii using nonionic amphipol. *Journal of Biological Chemistry*. 2019;294(41): 15003–15013. https://doi.org/https://doi.org/10.1074/jbc.RA119.009341.

126. Nagao R, Kato K, Kumazawa M, Ifuku K, Yokono M, Suzuki T, et al. Structural basis for different types of hetero-tetrameric light-harvesting complexes in a diatom PSII-FCPII supercomplex. *Nature Communications*. 2022;13(1): 1764. https://doi.org/10.1038/s41467-022-29294-5.

127. Shen L, Huang Z, Chang S, Wang W, Wang J, Kuang T, et al. Structure of a C2S2M2N2-type PSII–LHCII supercomplex from the green alga Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences*. 2019;116(42): 21246–21255. https://doi.org/10.1073/pnas.1912462116.

128. Jiang HW, Gisriel CJ, Cardona T, Flesher DA, Brudvig GW, Ho MY. Structure and evolution of Photosystem I in the early-branching cyanobacterium Anthocerotibacter panamensis. *bioRxiv*. 2024; 2010–2024.

129. Gisriel CJ, Flesher DA, Shen G, Wang J, Ho MY, Brudvig GW, et al. Structure of a photosystem I-ferredoxin complex from a marine cyanobacterium provides insights into far-red light photoacclimation. *Journal of Biological Chemistry*. 2022;298(1): 101408. https://doi.org/https://doi.org/10.1016/j.jbc.2021.101408.

130. Gisriel CJ, Shen G, Ho MY, Kurashov V, Flesher DA, Wang J, et al. Structure of a monomeric photosystem II core complex from a cyanobacterium acclimated to far-red light reveals the functions of chlorophylls d and f. *Journal of Biological Chemistry*. 2022;298(1): 101424. https://doi.org/10.1016/j.jbc.2021.101424.

131. Hussein R, Graça A, Forsman J, Aydin AO, Hall M, Gaetcke J, et al. Cryo–electron microscopy reveals hydrogen positions and water networks in photosystem II. *Science*. 2024;384(6702): 1349–1355.

132. Campbell MG, Cheng A, Brilot AF, Moeller A, Lyumkis D, Veesler D, et al. Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy. *Structure*. 2012;20(11): 1823–1828. https://doi.org/https://doi.org/10.1016/j.str.2012.08.026.

133. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*. 2013;10(6): 584–590. https://doi.org/10.1038/nmeth.2472.

134. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*. 2017;14(4): 331–332. https://doi.org/10.1038/nmeth.4193.

135. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*. 2015;192(2): 216–221. https://doi.org/https://doi.org/10.1016/j.jsb.2015.08.008.

136. Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*. 2019;2(1): 218. https://doi.org/10.1038/s42003-019-0437-z.

137. Bepler T, Morin A, Rapp M, Brasch J, Shapiro L, Noble AJ, et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*. 2019;16(11): 1153–1160. https://doi.org/10.1038/s41592-019-0575-8.

138. Tegunov D, Cramer P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods*. 2019;16(11): 1146–1152. https://doi.org/10.1038/s41592-019-0580-y.

139. Wlodawer A, Dauter Z. `Atomic resolution': a badly abused term in structural biology. *Acta Crystallographica Section D Structural Biology*. 2017;73(4): 379–380. https://doi.org/10.1107/S205979831700225X.

140. Yip KM, Fischer N, Paknia E, Chari A, Stark H. Atomic-resolution protein structure determination by cryo-EM. *Nature*. 2020;587(7832): 157–161. https://doi.org/10.1038/s41586-020-2833-4.

141. Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMGE, et al. Single-particle cryo-EM at atomic resolution. *Nature*. 2020;587(7832): 152–156. https://doi.org/10.1038/s41586-020-2829-0.

142. Tiemeijer PC, Bischoff M, Freitag B, Kisielowski C. Using a monochromator to improve the resolution in TEM to below 0.5Å. Part I: Creating highly coherent monochromated illumination. *Ultramicroscopy*. 2012;114: 72–81. https://doi.org/10.1016/j.ultramic.2012.01.008.

143. Gubbens A, Barfels M, Trevor C, Twesten R, Mooney P, Thomas P, et al. The GIF Quantum, a next generation post-column imaging energy filter. *Ultramicroscopy*. 2010;110(8): 962–970. https://doi.org/10.1016/j.ultramic.2010.01.009.

144. McMullan G, Faruqi AR, Henderson R. Direct Electron Detectors. In: 2016. p. 1–17. https://doi.org/10.1016/bs.mie.2016.05.056.

145. Kuijper M, van Hoften G, Janssen B, Geurink R, De Carlo S, Vos M, et al. FEI's direct electron detector developments: Embarking on a revolution in cryo-TEM. *Journal of Structural Biology*. 2015;192(2): 179–187. https://doi.org/https://doi.org/10.1016/j.jsb.2015.09.014.

146. Guo H, Franken E, Deng Y, Benlekbir S, Singla Lezcano G, Janssen B, et al. Electron-event representation data enable efficient cryoEM file storage with full preservation

of spatial and temporal resolution. *IUCrJ*. 2020;7(5): 860–869. https://doi.org/10.1107/S205225252000929X.

147. Müller H, Maßmann I, Uhlemann S, Hartel P, Zach J, Haider M. Aplanatic imaging systems for the transmission electron microscope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2011;645(1): 20–27. https://doi.org/10.1016/j.nima.2010.12.091.

148. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Microbiology*. 1979;111(1): 1–61. https://doi.org/https://doi.org/10.1099/00221287-111-1-1.

149. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*. 2017;14(3): 290–296. https://doi.org/10.1038/nmeth.4169.

150. Punjani A, Zhang H, Fleet DJ. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods*. 2020;17(12): 1214–1221. https://doi.org/10.1038/s41592-020-00990-8.

151. Rubinstein JL, Brubaker MA. Alignment of cryo-EM movies of individual particles by optimization of image translations. *Journal of Structural Biology*. 2015;192(2): 188–195. https://doi.org/https://doi.org/10.1016/j.jsb.2015.08.007.

152. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*. 2010;66(4): 486–501. https://doi.org/10.1107/S0907444910007493.

153. Adams PD, Afonine P V, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D*. 2010;66(2): 213–221. https://doi.org/10.1107/S0907444909052925.

154. Wojdyr M. GEMMI: A library for structural biology. *Journal of Open Source Software*. 2022;7(73): 4200. https://doi.org/10.21105/joss.04200.

155. Gisriel CJ, Wang J, Brudvig GW, Bryant DA. Opportunities and challenges for assigning cofactors in cryo-EM density maps of chlorophyll-containing proteins. *Communications Biology*. 2020;3(1): 408. https://doi.org/10.1038/s42003-020-01139-1.

156. Gisriel CJ, Huang HL, Reiss KM, Flesher DA, Batista VS, Bryant DA, et al. Quantitative assessment of chlorophyll types in cryo-EM maps of photosystem I acclimated to far-red light. *BBA Advances*. 2021;1: 100019. https://doi.org/https://doi.org/10.1016/j.bbadva.2021.100019.

157. Sottatipreedawong M, Kazmi AA, Vercellino I. How Cryo-EM Revolutionized the Field of Bioenergetics. *Microscopy and Microanalysis*. 2024; ozae089. https://doi.org/10.1093/mam/ozae089.

158. Shen L, Gao Y, Tang K, Qi R, Fu L, Chen JH, et al. Structure of a unique PSII-Pcb tetrameric megacomplex in a chlorophyll d–containing cyanobacterium. *Science Advances*. 2024;10(8): eadk7140.

159. Chen M. Chlorophyll Modifications and Their Spectral Extension in Oxygenic Photosynthesis. *Annual Review of Biochemistry*. 2014;83(1): 317–340. https://doi.org/10.1146/annurev-biochem-072711-162943.

160. Tanaka A, Ito H, Tanaka R, Tanaka NK, Yoshida K, Okada K. Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. *Proceedings of the National Academy of Sciences*. 1998;95(21): 12719–12723.

161. Chen M. Chlorophylls d and f: Synthesis, occurrence, light-harvesting, and pigment organization in chlorophyll-binding protein complexes. In: *Advances in botanical research*. Elsevier; 2019. p. 121–139.

162. Wang J, Moore PB. On the interpretation of electron microscopic maps of biological macromolecules. *Protein Science*. 2017;26(1): 122–129. https://doi.org/https://doi.org/10.1002/pro.3060.

163. Marques MA, Purdy MD, Yeager M. CryoEM maps are full of potential. *Current Opinion in Structural Biology*. 2019;58: 214–223. https://doi.org/10.1016/j.sbi.2019.04.006.

164. Wang J. On the appearance of carboxylates in electrostatic potential maps. *Protein Science*. 2017;26(3): 396–402. https://doi.org/10.1002/pro.3093.

165. Karuppasamy M, Karimi Nejadasl F, Vulovic M, Koster AJ, Ravelli RBG. Radiation damage in single-particle cryo-electron microscopy: effects of dose and dose rate. *Journal of synchrotron radiation*. 2011;18(3): 398–412.

166. Tros M, Mascoli V, Shen G, Ho MY, Bersanini L, Gisriel CJ, et al. Breaking the Red Limit: Efficient Trapping of Long-Wavelength Excitations in Chlorophyll-f-Containing Photosystem I. *Chem*. 2021;7(1): 155–173. https://doi.org/https://doi.org/10.1016/j.chempr.2020.10.024.

167. Wang J, Liu J, Gisriel CJ, Wu S, Maschietto F, Flesher DA, et al. How to correct relative voxel scale factors for calculations of vector-difference Fourier maps in cryo-EM. *Journal of structural biology*. 2022;214(4): 107902.

168. Barad BA, Echols N, Wang RYR, Cheng Y, DiMaio F, Adams PD, et al. EMRinger: side chain–directed model and map validation for 3D cryo-electron microscopy. *Nature Methods*. 2015;12(10): 943–946. https://doi.org/10.1038/nmeth.3541.

169. Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*. 1999;10(3): 61–74.

170. Consoli G, Tufail F, Leong HF, Viola S, Davis GA, Medranda D, et al. Locating the Missing Chlorophylls f in Far-red Photosystem I. *bioRxiv*. 2024; https://doi.org/10.1101/2024.08.06.606606.

171. Allakhverdiev SI, Kreslavski VD, Zharmukhamedov SK, Voloshin RA, Korol'kova D V, Tomo T, et al. Chlorophylls d and f and their role in primary photosynthetic processes of cyanobacteria. *Biochemistry (Moscow)*. 2016;81(3): 201–212. https://doi.org/10.1134/S0006297916030020.

172. van Stokkum IHM, Müller MG, Weißenborn J, Weigand S, Snellenburg JJ, Holzwarth AR. Energy transfer and trapping in photosystem I with and without chlorophyll-f. *iScience*. 2023;26(9): 107650. https://doi.org/https://doi.org/10.1016/j.isci.2023.107650.

173. Müller MG, Slavov C, Luthra R, Redding KE, Holzwarth AR. Independent initiation of primary electron transfer in the two branches of the photosystem I reaction center. *Proceedings of the National Academy of Sciences*. 2010;107(9): 4123–4128.

174. Müller MG, Niklas J, Lubitz W, Holzwarth AR. Ultrafast transient absorption studies on photosystem I reaction centers from Chlamydomonas reinhardtii. 1. A new interpretation of the energy trapping and early electron transfer steps in photosystem I. *Biophysical journal*. 2003;85(6): 3899–3922.

175. Şener MK, Park S, Lu D, Damjanović A, Ritz T, Fromme P, et al. Excitation migration in trimeric cyanobacterial photosystem I. *Journal of Chemical Physics*. 2004;120(23): 11183–11195. https://doi.org/10.1063/1.1739400.

176. Reiter S, Kiss FL, Hauer J, de Vivie-Riedle R. Thermal site energy fluctuations in photosystem I: new insights from MD/QM/MM calculations. *Chemical Science*. 2023;14(12): 3117–3131. https://doi.org/10.1039/D2SC06160K.

177. Llansola-Portoles MJ, Li F, Xu P, Streckaite S, Ilioaia C, Yang C, et al. Tuning antenna function through hydrogen bonds to chlorophyll a. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*. 2020;1861(4): 148078.

178. Diers JR, Kirmaier C, Taniguchi M, Lindsey JS, Bocian DF, Holten D. A perspective on the redox properties of tetrapyrrole macrocycles. *Physical Chemistry Chemical Physics*. 2021;23(35): 19130–19140. https://doi.org/10.1039/D1CP01943K.

179. Langley J, Purchase R, Viola S, Fantuzzi A, Davis GA, Shen JR, et al. Simulating the low-temperature, metastable electrochromism of Photosystem I: Applications to Thermosynechococcus vulcanus and Chroococcidiopsis thermalis. *The Journal of Chemical Physics*. 2022;157(12): 125103. https://doi.org/10.1063/5.0100431.

180. Kobayashi M, Akutsu S, Fujinuma D, Furukawa H, Komatsu H, Hotota Y, et al. Physicochemical properties of chlorophylls in oxygenic photosynthesis—succession of co-factors from anoxygenic to oxygenic photosynthesis. In: Dubinsky Z (ed.) *Photosynthesis*. InTech Rijeka, Croatia; 2013. p. 47–90.

181. Zamzam N, Kaucikas M, Nürnberg DJ, Rutherford AW, van Thor JJ. Femtosecond infrared spectroscopy of chlorophyll f-containing photosystem I. *Physical Chemistry Chemical Physics*. 2019;21(3): 1224–1234. https://doi.org/10.1039/C8CP05627G.

182. Shelaev I V, Gostev FE, Mamedov MD, Sarkisov OM, Nadtochenko VA, Shuvalov VA, et al. Femtosecond primary charge separation in Synechocystis sp. PCC 6803 photosystem I. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*. 2010;1797(8): 1410–1420.

183. Melkozernov AN, Lin S, Blankenship RE. Excitation dynamics and heterogeneity of energy equilibration in the core antenna of photosystem I from the cyanobacterium Synechocystis sp. PCC 6803. *Biochemistry*. 2000;39(6): 1489–1498.

184. Savikhin S, Jankowiak R, Golbeck J, van der Est A. The biophysics of photosynthesis. 2014;

185. Di Donato M, Stahl AD, van Stokkum IHM, van Grondelle R, Groot ML. Cofactors involved in light-driven charge separation in photosystem I identified by subpicosecond infrared spectroscopy. *Biochemistry*. 2011;50(4): 480–490.

186. Holzwarth AR, Müller MG, Niklas J, Lubitz W. Ultrafast transient absorption studies on photosystem I reaction centers from Chlamydomonas reinhardtii. 2: mutations near the P700 reaction center chlorophylls provide new insight into the nature of the primary electron donor. *Biophysical journal*. 2006;90(2): 552–565.

187. Gisriel CJ, Bryant DA, Brudvig GW, Cardona T. Molecular diversity and evolution of far-red light-acclimated photosystem I. *Frontiers in Plant Science*. 2023;14. https://doi.org/10.3389/fpls.2023.1289199.

188. MacGregor-Chatwin C, Nürnberg DJ, Jackson PJ, Vasilev C, Hitchcock A, Ho MY, et al. Changes in supramolecular organization of cyanobacterial thylakoid membrane complexes in response to far-red light photoacclimation. *Science Advances*. 2023;8(6): eabj4437. https://doi.org/10.1126/sciadv.abj4437.

189. Kurashov V, Ho MY, Shen G, Piedl K, Laremore TN, Bryant DA, et al. Energy transfer from chlorophyll f to the trapping center in naturally occurring and engineered Photosystem I complexes. *Photosynthesis Research*. 2019;141(2): 151–163. https://doi.org/10.1007/s11120-019-00616-x.

190. Tros M, Bersanini L, Shen G, Ho MY, van Stokkum IHM, Bryant DA, et al. Harvesting far-red light: Functional integration of chlorophyll f into Photosystem I complexes of Synechococcus sp. PCC 7002. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2020;1861(8): 148206. https://doi.org/https://doi.org/10.1016/j.bbabio.2020.148206.

191. Madjet MEA, Müh F, Renger T. Deciphering the Influence of Short-Range Electronic Couplings on Optical Properties of Molecular Dimers: Application to “Special Pairs” in Photosynthesis. *The Journal of Physical Chemistry B*. 2009;113(37): 12603–12614. https://doi.org/10.1021/jp906009j.

192. Wydrzynski T, Satoh K, Freeman J. Photosystem II: the light-driven water: plastoquinone oxidoreductase. 2005; https://link.springer.com/book/10.1007/1-4020-4254-X

193. Komenda J. Autotrophic cells of the Synechocystis psbH deletion mutant are deficient in synthesis of CP47 and accumulate inactive PS II core complexes. *Photosynthesis Research*. 2005;85(2): 161–167. https://doi.org/10.1007/S11120-005-1294-9.

194. Mayes SR, Dubbs JM, Nagy L, Barber J, Vass I, Hideg É. Further Characterization of the psbH Locus of Synechocystis sp. PCC 6803: Inactivation of psbH Impairs QA to QB Electron Transport in Photosystem 2. *Biochemistry*. 1993;32(6): 1454–1465. https://doi.org/10.1021/BI00057A008.

195. Gisriel CJ, Ranepura G, Brudvig GW, Gunner MR. Assignment of chlorophyll d in the ChlD1 site of the electron transfer chain of far-red light acclimated photosystem II supported by MCCE binding calculations. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2024;1865(4): 149496. https://doi.org/10.1016/j.bbabio.2024.149496.

196. Judd M, Morton J, Nürnberg D, Fantuzzi A, Rutherford AW, Purchase R, et al. The primary donor of far-red photosystem II: ChlD1 or PD2? *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2020;1861(10): 148248. https://doi.org/https://doi.org/10.1016/j.bbabio.2020.148248.

197. Gisriel CJ, Cardona T, Bryant DA, Brudvig GW. Molecular evolution of far-red light-acclimated photosystem II. *Microorganisms*. 2022;10(7): 1270.

198. Shen G, Vermaas WFJ. Mutation of Chlorophyll Ligands in the Chlorophyll-Binding CP47 Protein As Studied in a Synechocystis sp. PCC 6803 Photosystem I-less Background. *Biochemistry*. 1994;33(23): 7379–7388. https://doi.org/10.1021/bi00189a044.

199. D'Haene SE, Sobotka R, Bučinská L, Dekker JP, Komenda J. Interaction of the PsbH subunit with a chlorophyll bound to histidine 114 of CP47 is responsible for the red 77K fluorescence of Photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2015;1847(10): 1327–1334. https://doi.org/10.1016/j.bbabio.2015.07.003.

200. Jamali K, Käll L, Zhang R, Brown A, Kimanius D, Scheres SHW. Automated model building and protein identification in cryo-EM maps. *Nature 2024 628:8007*. 2024;628(8007): 450–457. https://doi.org/10.1038/s41586-024-07215-4.

201. Andrizhiyevskaya EG, Chojnicka A, Bautista JA, Diner BA, Grondelle R van, Grondelle R van, et al. Origin of the F685 and F695 fluorescence in photosystem II. *Photosynthesis Research*. 2005; https://doi.org/10.1007/s11120-005-0478-7.

202. Zabret J, Bohn S, Schuller SK, Arnolds O, Möller M, Meier-Credo J, et al. Structural insights into photosystem II assembly. *Nature plants*. 2021;7(4): 524–538.

203. Reinot T, Chen J, Kell A, Jassas M, Robben KC, Zazubovich V, et al. On the Conflicting Estimations of Pigment Site Energies in Photosynthetic Complexes: A Case Study of the CP47 Complex. *Analytical Chemistry Insights*. 2016;11: ACI.S32151. https://doi.org/10.4137/ACI.S32151.

204. Sirohiwal A, Neese F, Pantazis DA. Chlorophyll excitation energies and structural stability of the CP47 antenna of photosystem II: a case study in the first-principles simulation of light-harvesting complexes. *Chemical Science*. 2021; https://doi.org/10.1039/d0sc06616h.

205. de Weerd FL, Palacios MA, Andrizhiyevskaya EG, Dekker JP, van Grondelle R. Identifying the Lowest Electronic States of the Chlorophylls in the CP47 Core Antenna Protein of Photosystem II. *Biochemistry*. 2002;41(51): 15224–15233. https://doi.org/10.1021/bi0261948.

206. Sugiura M, Kato Y, Takahashi R, Suzuki H, Watanabe T, Noguchi T, et al. Energetics in Photosystem II from Thermosynechococcus elongatus with a D1 protein encoded by either the psbA1 or psbA3 gene. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*. 2010;1797(8): 1491–1499.

207. Merry SAP, Nixon PJ, Barter LMC, Schilstra M, Porter G, Barber J, et al. Modulation of Quantum Yield of Primary Radical Pair Formation in Photosystem II by Site-Directed Mutagenesis Affecting Radical Cations and Anions. *Biochemistry*. 1998;37(50): 17439–17447. https://doi.org/10.1021/bi980502d.

208. Soulier N, Laremore TN, Bryant DA. Characterization of cyanobacterial allophycocyanins absorbing far-red light. *Photosynthesis research*. 2020;145(3): 189–207.

209. Gisriel CJ, Elias E, Shen G, Soulier NT, Brudvig GW, Croce R, et al. Structural comparison of allophycocyanin variants reveals the molecular basis for their spectral differences. *Photosynthesis Research*. 2024;162(2–3): 157–170. https://doi.org/10.1007/s11120-023-01048-4.

210. Ho MY, Gan F, Shen G, Bryant DA. Far-red light photoacclimation (FaRLiP) in Synechococcus sp. PCC 7335. II. Characterization of phycobiliproteins produced during acclimation to far-red light. *Photosynthesis research*. 2017;131: 187–202.

211. Gan F, Shen G, Bryant D. Occurrence of Far-Red Light Photoacclimation (FaRLiP) in Diverse Cyanobacteria. *Life*. 2014;5(1): 4–24. https://doi.org/10.3390/life5010004.

212. Xu QZ, Han JX, Tang QY, Ding WL, Miao D, Zhou M, et al. Far-red light photoacclimation: Chromophorylation of FR induced α- and β-subunits of allophycocyanin from Chroococcidiopsis thermalis sp. PCC7203. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2016;1857(9): 1607–1616. https://doi.org/https://doi.org/10.1016/j.bbabio.2016.06.008.

213. Brejc K, Ficner R, Huber R, Steinbacher S. Isolation, crystallization, crystal structure analysis and refinement of allophycocyanin from the cyanobacterium Spirulina platensis at 2.3 Å resolution. *Journal of Molecular Biology*. 1995;249(2): 424–440. https://doi.org/10.1006/jmbi.1995.0307.

214. Gan F, Bryant DA. Adaptive and acclimative responses of cyanobacteria to far-red light. *Environmental microbiology*. 2015;17(10): 3450–3465.

215. Xu QZ, Tang QY, Han JX, Ding WL, Zhao BQ, Zhou M, et al. Chromophorylation (in Escherichia coli) of allophycocyanin B subunits from far-red light acclimated Chroococcidiopsis thermalis sp. PCC7203. *Photochemical & Photobiological Sciences*. 2017;16: 1153–1161.

216. Soulier NT, Bryant DA. Long-wavelength phycobiliproteins. In: *Photosynthesis*. Elsevier; 2023. p. 9–32.

217. Domínguez-Martín MA, Sauer P V., Kirst H, Sutter M, Bína D, Greber BJ, et al. Structures of a phycobilisome in light-harvesting and photoprotected states. *Nature 2022 609:7928*. 2022;609(7928): 835–845. https://doi.org/10.1038/s41586-022-05156-4.

218. Gindt YM, Zhou J, Bryant DA, Sauer K. Spectroscopic studies of phycobilisome subcore preparations lacking key core chromophores: Assignment of excited state energies to the Lcm, β18 and αAP-B chromophores. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 1994;1186(3): 153–162. https://doi.org/10.1016/0005-2728(94)90174-0.

219. Rockwell NC, Lagarias JC. A Brief History of Phytochromes. *ChemPhysChem*. 2010;11(6): 1172–1180. https://doi.org/https://doi.org/10.1002/cphc.200900894.

220. Halliday KJ, Davis SJ. Light-sensing phytochromes feel the heat. *Science*. 2016;354(6314): 832–833. https://doi.org/10.1126/science.aaj1918.

221. Legris M, Klose C, Burgie ES, Rojas CCR, Neme M, Hiltbrunner A, et al. Phytochrome B integrates light and temperature signals in *Arabidopsis*. *Science*. 2016;354(6314): 897–900. https://doi.org/10.1126/science.aaf5656.

222. Legris M, Ince YÇ, Fankhauser C. Molecular mechanisms underlying phytochrome-controlled morphogenesis in plants. *Nature Communications*. 2019;10(1): 5219. https://doi.org/10.1038/s41467-019-13045-0.

223. Rockwell NC, Su YS, Lagarias JC. Phytochrome structure and signaling mechanisms. *Annual Review of Plant Biology*. 2006;57(1): 837–858. https://doi.org/10.1146/annurev.arplant.56.032604.144208.

224. Björling A, Berntsson O, Takala H, Gallagher KD, Patel H, Gustavsson E, et al. Ubiquitous Structural Signaling in Bacterial Phytochromes. *Journal of Physical*

Chemistry Letters. 2015;6(17): 3379–3383. https://doi.org/10.1021/acs.jpclett.5b01629.

225. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000;28(1): 27–30. https://doi.org/10.1093/NAR/28.1.27.

226. Mukougawa K, Kanamoto H, Kobayashi T, Yokota A, Kohchi T. Metabolic engineering to produce phytochromes with phytochromobilin, phycocyanobilin, or phycoerythrobilin chromophore in Escherichia coli. *FEBS Letters*. 2006;580(5): 1333–1338. https://doi.org/10.1016/j.febslet.2006.01.051.

227. Mailliet J, Psakis G, Feilke K, Sineshchekov V, Essen LO, Hughes J. Spectroscopy and a High-Resolution Crystal Structure of Tyr263 Mutants of Cyanobacterial Phytochrome Cph1. *Journal of Molecular Biology*. 2011;413(1): 115–127. https://doi.org/https://doi.org/10.1016/j.jmb.2011.08.023.

228. Burgie ES, Zhang J, Vierstra RD. Crystal Structure of Deinococcus Phytochrome in the Photoactivated State Reveals a Cascade of Structural Rearrangements during Photoconversion. *Structure*. 2016;24(3): 448–457. https://doi.org/https://doi.org/10.1016/j.str.2016.01.001.

229. Burgie ES, Bussell AN, Walker JM, Dubiel K, Vierstra RD. Crystal structure of the photosensing module from a red/far-red light-absorbing plant phytochrome. *Proceedings of the National Academy of Sciences*. 2014;111(28): 10179–10184. https://doi.org/10.1073/pnas.1403096111.

230. Li H, Burgie ES, Gannam ZTK, Li H, Vierstra RD. Plant phytochrome B is an asymmetric dimer with unique signalling potential. *Nature*. 2022;604(7904): 127–133. https://doi.org/10.1038/s41586-022-04529-z.

231. Wahlgren WY, Claesson E, Tuure I, Trillo-Muyo S, Bódizs S, Ihalainen JA, et al. Structural mechanism of signal transduction in a phytochrome histidine kinase. *Nature Communications*. 2022;13(1): 7673. https://doi.org/10.1038/s41467-022-34893-3.

232. Wahlgren WY, Golonka D, Westenhoff S, Möglich A. Cryo-Electron Microscopy of Arabidopsis thaliana Phytochrome A in Its Pr State Reveals Head-to-Head Homodimeric Architecture. *Frontiers in Plant Science*. 2021;12. https://doi.org/10.3389/fpls.2021.663751.

233. Lichtenberg M, Brodersen KE, Kühl M. Radiative Energy Budgets of Phototrophic Surface-Associated Microbial Communities and their Photosynthetic Efficiency Under Diffuse and Collimated Light. *Frontiers in Microbiology*. 2017;8. https://doi.org/10.3389/fmicb.2017.00452.

234. Lichtenberg M, Cartaxana P, Kühl M. Vertical Migration Optimizes Photosynthetic Efficiency of Motile Cyanobacteria in a Coastal Microbial Mat. *Frontiers in Marine Science*. 2020;7. https://doi.org/10.3389/fmars.2020.00359.

Appendix A: Chl *f* candidates in literature

Jordan et al. ID	Kato et al. ID	7S3D	6KMX	9I9L	Mg 2^+ Ligand
A1012	A802			Y	H_2O
A1121	A824	Y		Y	H_2O
A1123	A826	SS	Y	SS	H_2O
A1124	A827		Y		H_2O
A1127	A830		Y		His δ
A1129	A832		Y		His δ
B1207	L202	Y	Y	Y	Gln
B1222	B824		Y		H_2O
B1230	B832	Y		Y	His ε
B1237	A844	Y	Y	Y	H_2O
B1238	B839	Y		Y	H_2O
L1503	L206			Y	H_2O

Table 8 – FR-PSI Chl *f* sites assignments

The table presents an overview of the relevant data processing parameters used.

ID	8EQM	FR-PSII	Mg 2^+ Ligand
B605	Y		H_2O

B608	Y	Y	H_2O
B611	SS		His δ
B614	Y	Y	H_2O
B617		Y	H_2O
C507	Y	Y	His δ

Table 9 – FR-PSII Chl *f* sites assignments

The table presents an overview of the relevant data processing parameters used.