Studies of light harvesting complexes from purple photosynthetic bacteria

Rachel Margaret Mulvaney

BSc (Hons)

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

School of Molecular, Cellular and Systems Biology

College of Medical, Veterinary and Life Sciences

University of Glasgow

2013

Abstract

In this thesis light harvesting complexes, the LH2 and core complexes, from several different species of purple photosynthetic bacteria have been analysed both functionally and structurally. Purified monomeric core complexes from *Rhodopseudomonas (Rps.) palustris* have been used to isolate and identify the putative Protein W. This information was then used to create a Protein W deletion mutant. A low-resolution crystal structure of the monomeric core complex from *Allochromatium (Alc.) vinosum* is presented which suggests that the LH1 complex is a complete ellipse, unlike the core complex from *Rps. palustris*.

It has previously been shown that some species are able to synthesise LH2 complexes that have different NIR absorption spectra. For example, strains 7050 and 7750 of *Rps. acidophila* can express both the B800-850 and B800-820 LH2 complexes, whilst strain 10050 only expresses the B800-850 LH2 complex despite evidence to suggest that this strain contains multiple LH2 genes (*pucBA* genes). It is this homogeneity that has made the LH2 complexes from this strain structurally amenable. Here, genomic DNA from *Rps. acidophila* strain 10050 has been isolated and sequenced using the next generation sequencing (NGS) technique, Illumina sequencing. So far 8 *pucBA* gene pairs were identified arranged into 2 distinct operons, one containing B800-850 *pucBA* genes and *pucC*, the putative Bchl transporter that is essential for efficient LH2 expression. The second operon contains B800-820 *pucBA* gene pairs only. Analysis of the protein products of the B800-850 type *pucBA* gene pairs has shown that none of these proteins match the sequence for the LH2 that is expressed by *Rps. acidophila* strain 10050. The crystal structure of the LH2 complex from the culture of *Rps. acidophila* used to isolate the genomic DNA was resolved to 2.05 Å from crystals of the LH2 complex. This structure shows that the protein sequence of the LH2 complex has not changed. Hence, not all the *pucBA* gene pairs have been identified in the genome sequence data. Currently mate-pair

sequencing is being completed to fill in the gaps of sequence data and to complete the genome sequence.

LH2 complexes contain carotenoid (Car) and Bchl molecules. In this thesis, the energy transfer mechanisms between Car and Bchl molecules have been investigated using 2-dimensional electronic spectroscopy (2DES). This technique splits the emission and excitation events on 2-dimensions, which can make the less populated 'dark' states more visible as overlapping peaks can be separated. Car moleucles are not seen as theoretically efficient in photosynthesis. This is due to short life times of the excited state S_2. However, the Car used in photosynthesis have conjugated carbon tails with ≥9 π electrons. According to calculations by Tavan and Schulten, these molecules have the propensity to contain additional excited states that lie below the S_2 state that can be involved in energy transfer and increase the efficiency of energy transfer between the Car and Bchl molecules. For the first time an intermediate Car electronic state has been directly observed and shown to be involved in energy transfer between the Car and Bchl molecules.

Table of contents

List of tables

List of figures

Acknowledgments

Firstly, I would like to thank Professor Richard Cogdell for his guidance, supervision, enthusiasm and all the wonderful opportunities that he has given me throughout my PhD. It has been a pleasure to work for Richard and I have cherished every moment. I would also like to thank Dr Alastair Gardiner for everything. He is the heart of Team Cogdell and his constant support and belief has helped me through some of the harder times and his bad jokes are always a source of amusement.

I would also like to thank the BBSRC for awarding me a 4 year studentship.

I have been privileged to work with the people in Team Cogdell: Dr Aleksander Roszak, Dr Sarah Henry, Dr Anne-Marie Carey, Dr Heiko Lockstein, June Southall, Khuram Ashraf, Laura Cranston, Kirsty Hacking and honorary Cogdell group member Dr Alette Brinth. I would also like to thank previous members of the lab: Dr Tatas Brotosudarmo and Dr Nichola Picken. It is the friendliest and most supportive lab, and I am sorry to be leaving. I would like to extend special thanks to Khuram Ashraf and Dr Alette Brinth for all the late night trouble-shooting (especially when I had lost all my patience), Dr Anne-Marie Carey for staying up late with me to check through this thesis, as well as Dr Aleksander Roszak for tirelessly helping me with data collection, processing and some very useful discussions.

During my PhD I was lucky enough to collaborate with numerous groups. I would like to thank Dr Greg Scholes and Dr Evgeny Ostroumov (University of Toronto) for an amazing collaboration using 2DES, Professor Neil Hunter and Dr Philip Jackson (University of Sheffield) for helping me identify candidate Protein W proteins, Dr Graham Hamilton and Dr Pawel Herzyk (University of Glasgow) for the ongoing NGS of the *Rps. acidophila* strain 10050 genome and finally Dr

Niall Fraser (University of Dundee) for useful discussions, friendship and the occasional beer.

My family has been a constant source of support, especially throughout the writing period, and as such are a large part of this thesis, which wouldn’t be here without them. Finally, I would like to thank my partner Ben. He has supported me throughout my A-levels, undergraduate degree and now my PhD. He has been a huge support, followed me the length of the country, always believed in me, picked me up when I didn't think I could do it anymore and never stopped me from chasing my dreams in-spite of his own, and I am eternally grateful.

Author’s declaration

This thesis is an original composition which describes work performed entirely by myself unless otherwise cited or acknowledged. Its contents have not previously been submitted for any other degree. The research for this thesis was performed between January 2010 and August 2013.

Signature .....

Printed name .....

Abbreviations

2D	2-dimensional
2DEAS	2-dimensional evolutionary absorption spectra
2DES	2-dimensional electronic spectroscopy
3D	3-dimensional
A	Acceptor
Alc.	*Allochromatium*
Amp	Ampicillin
APB	Anoxygenic photosynthetic bacteria
B800	Bulk Bchl with a maximal absorbance at 800 nm
B820	Bulk Bchl with a maximal absorbance at 820 nm
B850	Bulk Bchl with a maximal absorbance at 850 nm
$B_{A/B}$	Monomeric Bchl on branch A or B
Bchl	Bacteriochlorphyll
B-factors	Indication of mobility of an atom
Blc.	*Blastochloris*
CAR	Carotenoid
CD	Circular dichroism
CMC	Critical micelle concentration
D	Donor
DE52	Diethylaminoethyl cellulose
DLS	Diamond light source
DNA	Deoxyribonucleic acid
DMNG-3	Decyl 3-maltose neopentyl glycol
DMSO	Dimethyl sulfoxide
DS	Down stream
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ESA	Excited state absorption
ET	Energy transfer
FWHM	Full width half maximum
GPCRs	G-protein coupled receptors
GSB	Ground state bleach
H	Heavy subunit of RC
HA/B	Bacteriopheophytin on branch A or B of RC
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
HL	High light
I	Spectral overlap of D emission and A absorption
IAA	Isoamyl alcohol
IC	Internal conversion
IC21	Internal conversion from S_2 to S_1
ICM	Intra cytoplasmic membrane
IMAC	Immobilised metal affinity chromatography
JB	JellyBody REFMAC5

k	Strength of interactions between D and A
Kan	Kanamycin
KO	Knock-out
L	Light subunit of RC
LB	Luria Broth
LC	Liquid chromatography
LC-MS-MS	Liquid chromatography tandem Mass spectrometry
LDAO	Lauryldimethylamine oxide
LH1	Light harvesting 1 complex
LH1:RC	Light harvesting 1 complex:reaction centre or Core complex
LH2	Light harvesting 2 complex
LL	Low light
LLG	Log-likelihood gain
M	Medium-subunit of RC
MAG	Monoacylglycerol
MCS	Multiple cloning site
MES	2-ethanesulfonic acid
MR	Molecular replacement
MWCO	Molecular weight cut off
N	Length of the conjugated carbon chain in Car molecules
NGS	Next generation sequencing
NIR	Near infra red
NM	Nonyl-maltoside
NMR	Nuclear magnetic resonance
NOPA	Non collinear optical parametric amplifier
OD	Optical density
OR	Optical ratio
$P_{A/B}$	Special pair of Bchl molecules on brance A or B of RC
PCR	polymerase chain reaction
PEG	Polyethylene glycol
Pi	Isoelectric point
PNSB	Purple non-sulphur bacteria
PS^{-}	Photosynthetic minus phenotype
PS^{+}	Photosynthetic positive phenotype
PSB	Purple sulphur bacteria
PSU	Photosynthetic unit
Q/QH_2	Quinone/Quinol exchange
$Q_{A/B}$	Quinone binding site on branch A or B of the RC
R	Distance between the centre of D to the centre of A molecules
Rba.	*Rhodobacter*
RC	Reaction centre
Rca.	*Rhodobaccia*
R_{factor}	Reliability factor
RFZ	*Rotational Z- factor*
RG	Rhodopin glucoside
ROS	Reactive oxygen species
RPM	Rate per minute

Rps.	*Rhodopseudomonas*
Rs.	*Rhodospirillum*
RT	Room temperature
SCX	Strong cation exchange
SDG	Sucrose density gradients
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	Stimulated emission
SEC	Size exclusion chromatography
SMS	Single molecule spectroscopy
SOE	Splice overlap extension
succs$^-$	C-Succinate media without casamino acids
Tch.	*Thermochromatium*
TFA	Trifluoroacetic acid
TFZ	Translational Z-factor
UHR-TOF	Ultra-high resolution time of flight
US	Up stream
WT	Wild type
β-DDM	β- Dodecyl maltoside
β-DM	β- Decyl maltoside
τd	Life time of Donor molecule
φD	Flouresence quantum yield

1. Introduction

Photosynthesis is one of the most important biological processes and has evolved to convert solar energy efficiently into chemical energy within a wide range of ecological niches. The sun provides $4.3x10^{20}$ J of energy every hour at the Earth's surface, which is more than mankind currently requires in a year [1-3]. As such there is currently a great deal of interest to try to learn from photosynthesis with respect to harvesting solar energy to produce fuels, especially fuels for transport. In order to design artificial photosynthetic mimics and devices capable of producing a solar fuel we first need to understand exactly how photosynthesis works. Once this is understood artificial systems can be designed with factors that enhance the efficiency of the reaction and methods for generating sustainable and scalable artificial systems in mind. Great efforts are, therefore, being made to understand the exact molecular details of photosynthesis [1,4-6].

1.1. Overview of bacterial photosynthesis

Photosynthesis has been historically associated with oxygen evolution [7]. In bacterial photosynthesis, however, this is not always the case. Purple photosynthetic bacteria were first observed by Sir E. Ray Lankester in 1873 when water containing decaying matter created the appropriate conditions for purple bacteria to grow [7,8]. Since then purple and green photosynthetic bacteria have been subjected to intense scientific investigation.

By comparing the basic photosynthetic reactions of green plants and green sulphur bacteria (GSB), van Neil proposed a general equation for photosynthesis (Figure 1-1). He observed that there was a key similarity in that all the reactions were examples of photochemical carbon dioxide reduction using different electron donors. It was shown that this equation satisfied the metabolism of *all* photosynthetic species

studied. Oxygen would be evolved if water was the electron donor, whereas, if sulphide was the electron donor, then elemental sulphur would be produced [7,9].

General equation:

$$CO_2 + 2H_2A \xrightarrow{\text{Light}} (CH_2O) + H_2O + 2A$$

Figure 1-1 van Neil's general photosynthetic equation. The reactions are all photochemical carbon dioxide reduction reactions. The general equation can be satisfied by the presence of different hydrogen donors.

1.2. Purple photosynthetic bacteria: systematics

Anaerobic photosynthetic bacteria (APB) are incredibly diverse group of organisms that include green and purple photosynthetic bacteria. Purple photosynthetic bacteria are very metabolically versatile and can grow both chemoheterotrophically and photosynthetically. These bacteria can adjust their growth habit depending on the oxygen tension, light intensity, temperature and substrates present in their immediate environments [10]. The genera of purple photosynthetic bacteria are divided into those that can utilise reduced sulphur compounds (purple sulphur photosynthetic bacteria) as electron donors and those that do not, rather they use reduced carbon compounds as electron donors, e.g., succinate (purple non-sulphur photosynthetic bacteria) [9,11]. This allows the bacteria to thrive in a multitude of different environments. Other photosynthetic bacteria, such as green sulphur bacteria grow in the absence of oxygen and are strictly anaerobic. It is this versatility that allows photosynthetic bacteria to thrive in conditions where similar organisms cannot function [9].

APB have adapted to their environments. They are usually found growing photosynthetically deeper in the water column where oxygen and light are limited due to the presence of plants, cyanobacteria and algae closer to the surface. Cyanobacteria and algae contain chlorophyll molecules that absorb in the blue and red wavelengths of visible light.

Therefore, APB have evolved their own pigments; bacteriochlorophyll (Bchl) molecules that absorb the near infra red (NIR) wavelengths of light and carotenoid (Car) molecules that absorb the green wavelengths of light. These pigments are organised into photosynthetic units (PSU). Classically a PSU has been defined as the number of Bchl molecules per reaction centre (RC). The RC is the pigment protein complex within which electron transfer and charge separation first occurs (See section 1.3.1). APB are typically able to vary the size and composition of their PSUs in response to a variety of environmental factors such as light intensity. For example, when cells are grown at lower light conditions, the size of the PSU increases to compensate, and the intracytoplasmic membranes (ICM), where these complexes are found, become more numerous allowing the total number of PSUs to increase. This increases the surface area of membranes so that more light is absorbed and these bacteria can successfully compete with other organisms.

Purple photosynthetic bacteria are anoxygenic gram-negative phototrophs. They belong to the phyla *proteobacteria*. The purple non-sulphur bacteria (PNSB) belong to the *α* and *β* classes, whilst the purple sulphur bacteria (PSB) belong to the *γ* class of *proteobacteria* [12,13] (Table 1-1).

Generally, PNSB are metabolically versatile. As already mentioned, they can grow chemoheterotrophically, photosynthetically and in some cases fermentatively. The 'switch' that controls the transition between growing by photosynthesis or respiration is the partial pressure of oxygen. If the pressure is higher and more oxygen is present the PNSB will grow chemoheterotrophically [10,11].

PNSB generally are prevalent in habitats where there is decaying matter, such as sewage and in waste lagoons. Although the PNSB have been defined by their inability to utilise sulphur, if sulphide is present at low levels (<0.5 mM) some PNSB will grow and the oxidised elemental

sulphur is deposited outside the cell. Within the *γ-proteobacteria* there are two distinct families (Table 1-1)

Taxonomy/Phylogeny	Genus	Morphology
Purple non-sulphur bacteria	*Rhodobaca (Rca.)*	Cocci to short rods
α-proteobacteria	*Rhodobacter (Rba.)*	Rods
	Rhodovulum (Rdv.)	Rods-cocci
	Rhodopseudomonas (Rps.)	Budding rods
	Rhodoblastus (Rbl.)	Budding rods
	Blastochloris (Blc.)	Budding rods
	Rhodospirillum (Rsp.)	Spirilla
	Rhodovibrio (Rhv.)	Vibrio
β-proteobacteria	*Rhodocyclus (Rcy.)*	Curled vibrios
	Rhodoferax (Rfx.)	Rods, vibrios
	Rubrivivax (Rvi.)	Rods, curved rods
Purple sulphur bacteria	*Allochromatium (Alc.)*	Rods
γ-proteobacteria	*Ameobacter (Amb.)*	Cocci in plates or clumps
Family *Chromatiaceae*	*Chromatium (Chr.)*	Rods
	Halochromatium (Hch.)	Rods
	Marichromatium (Mch.)	Rods
	Rhabdochromatium (Rbc.)	Rods
	Thermochromatium (Tch.)	Rods
	Thioalkalicoccus (Tac.)	Cocci
	Thiobaca (Tba.)	Rods
Family *Ectothiorhodospiraceae*	*Ectothiorhodospira (Ect.)*	Vibrio to spirilla
	Halorhodospira (Hlr.)	Vibrio to spirilla
	Thiorhodospira (Trs.)	Vibrio to spirilla
	Ectothiorhodospirus (Ets.)	Rods

Table 1-1 Genera of anoxygenic purple photosynthetic bacteria taken from Madigan and Jung, 2008 [11].

Chromatiaceae and *Ectothiorhodospiraceae*. These two families are separated by the intracellular and extracellular storage of sulphur respectively. This production of extracellular elemental sulphur inhibits the formation of large blooms of PNSB, whilst PSB can grow in large blooms and microbial mats at positions in lakes where the light and sulphur concentrations are optimal.

1.3. Photosynthetic antenna systems

When purple photosynthetic bacteria are grown anaerobically in the light highly invaginated photosynthetic membranes are synthesised called intra-cytoplasmic membranes (ICM). ICM are continuous with but

distinct from the inner membrane and contain all the molecular machinery required to harvest the energy of a photon, produce and store chemical energy [14,15]. Within the ICM (Figure 1-2) there are generally two types of antenna complexes, the molecules responsible for the initial photon capture, called light harvesting 2 (LH2) and light harvesting 1 (LH1) complexes. The LH1 complexes are associated with the reaction centre (RC) in a 1:1 stoichiometry. [14]. The LH1:RC complex is more commonly called the core complex. The ICM also contains all the components required for photosynthetic electron transport and ATP synthesis such as the cytochrome bc_1 complex, dehydrogenases (that oxidise succinate or other reduced carbon compounds) and the ATP synthase. These complexes are involved in completing the cyclic electron transfer reactions that generate the proton-motive force and ultimately result in the synthesis of ATP.

In most purple photosynthetic bacteria the PSU is composed of the LH2 and core complexes [14]. There are some species however, such as *Rhodospirillum (Rs.) rubrum*, which only contain LH1:RC complexes. In these species the size of the PSU is fixed and they respond to changes in light intensity just by altering the number of PSUs present per cell [16,17]. In contrast, species such as *Rba. sphaeroides*, which also contain LH2 complexes are not only able to respond to changes in light intensity by increasing the number of PSUs present per cell, but are also able to vary the size of the PSUs by regulating the ratio of LH2 to the core complexes.

Interestingly it has also been shown through mutational analyses that the minimal PSU possible is the RC alone, although RC-only mutants can only grow photosynthetically at very high light intensities. Mutants have also been generated that contain LH2 and RC-only or core complex only [19]. These mutants grow photosynthetically at a faster rate than RC-only mutants. Previously, work showed that in *Rba. sphaeroides* the amount of LH2 complexes synthesised was very variable in response to

changes in its environment, such as light intensity and its wavelength [14,19,20].

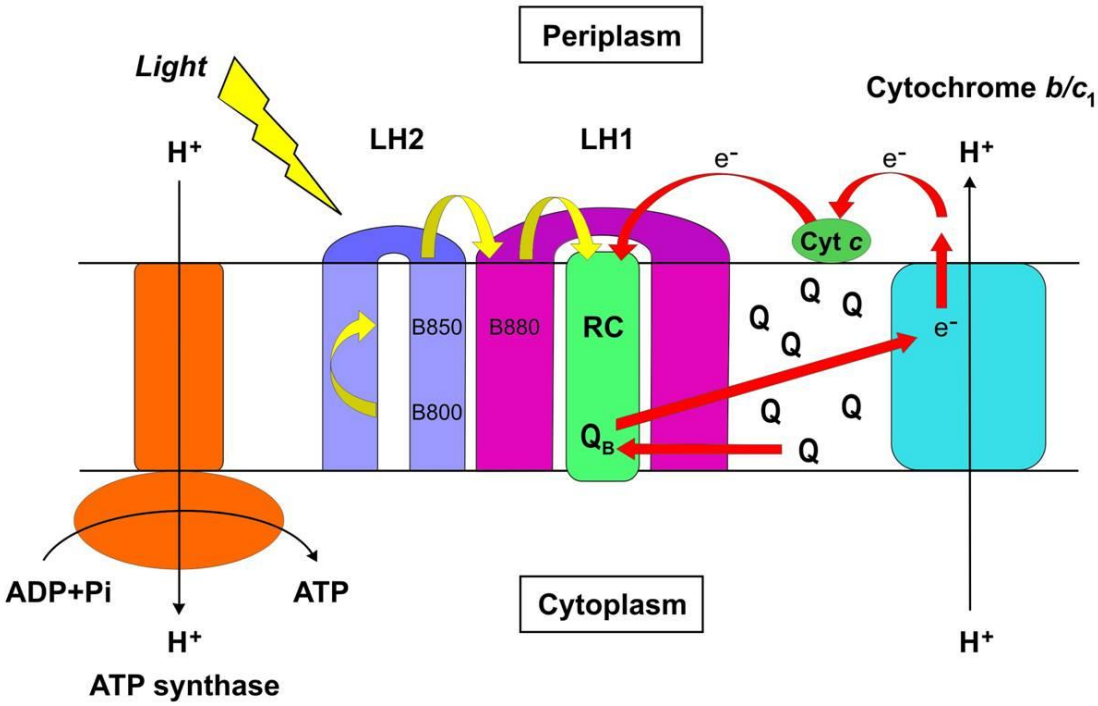


Figure 1-2 [18] Schematic of a bacterial photosynthetic membrane. Light harvesting (LH) complexes within the membrane harvest the energy of a photon and funnel the energy down an energy gradient until it reaches the reaction centre (RC). At the RC electron transfer reactions are initiated at the Bchl special pair. The electron is passed to nearby cofactors down a redox gradient until it reaches a quinone molecule (Q_B). When the quinone is reduced by the first electron, charge separation is generated across the membrane. The quinone is subsequently further reduced and doubly protonated. The resultant quinol molecule then diffuses through the membrane to the cytochrome b/c_1 complex where it is oxidised and a proton motive force is generated which is used by ATP synthase to synthesise an ATP molecule.

1.3.1. Structure of RC

The RC from *Blastochloris (Blc.) viridis* (previously *Rhodospeudomonas viridis*) was the first membrane protein to have its structure solved using X-ray crystallography (section 1.4) [21]. This ground-breaking research changed the field of membrane protein biochemistry and the Nobel prize was awarded to Michel and co-workers [22]. The RC was

chosen from *Blc. viridis* as a structural target as, although it is a membrane protein, it was also known that the LH1:RC complex formed 2 dimensional crystals in the membrane [22]. They therefore hoped that these RCs would also have a strong tendency to crystallise. The complex is also coloured, available in large quantities, robust and denaturation can be clearly followed by a change in the absorption spectrum. For these reasons it was viewed as a good protein target [22] although, at that time, membrane proteins were considered impossible to crystallise and to obtain high-resolution structural data with. The gamble paid off as the structure was solved to 3 Å resolution [21].

Figure 1-3 shows the structure of the *Blc. viridis* RC as modeled in PyMol [23] using the PDB file 1PRC. The complex is composed of 4 subunits, H (heavy), M (medium), L (light) and the cytochrome-*c* subunit. The L and M subunits bind the following cofactors; 2 strongly coupled Bchl molecules ($P_{A/B}$), 2 monomeric Bchl molecules ($B_{A/B}$), 2 Bacteriopheophytin molecules (Bpheo) ($H_{A/B}$), 2 quinone molecules ($Q_{A/B}$) and an Fe^{2+} ion (Figure 1-4). The cofactors were clearly visible in the electron density maps for this structure and the subsequent cofactor map was shown to agree with the results obtained from time-resolved spectroscopic investigations of the electron transfer reactions conducted in *Rhodobacter* (*Rba.) sphaeroides*, [21,24-27].

The RC is 130 Å in height from the tip of the cytochrome-*c* subunit to the H-subunit and contains elliptical axes of 70 Å and 30 Å (Figure 1-3). The L and M subunits differ in length but have a very similar fold composed of 5 transmembrane α-helices (A-E) arranged beside one

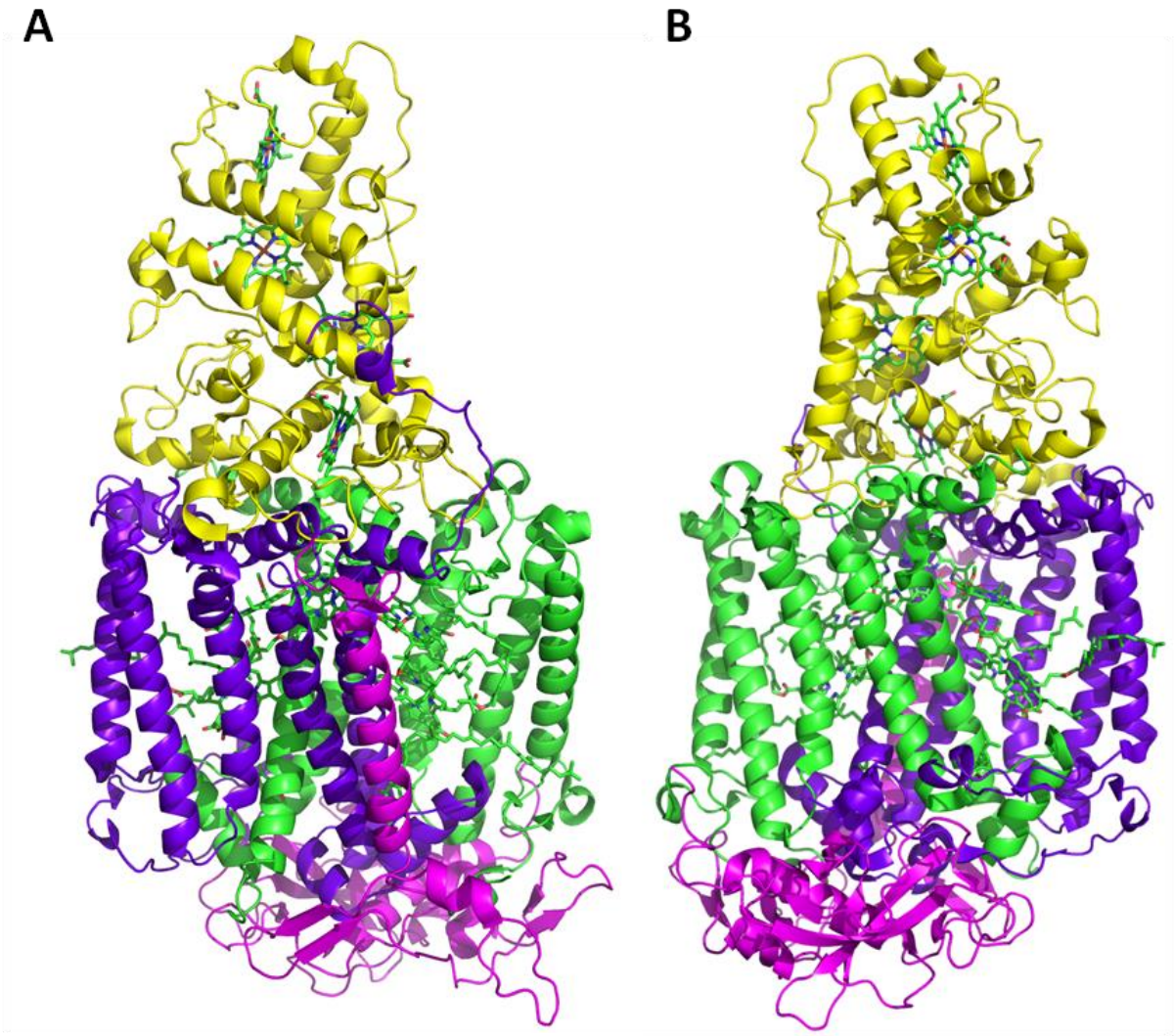


Figure 1-3 Structure of the RC from *Blc. viridis*. Modeled in PyMol (1PRC) from 2 angles. The structure was solved by [21].The RC is composed of Cytochrome-c subunit (yellow), L-subunit (green), M-subunit (blue) and the H-subunit (magenta). The L and M-subunits are integral membrane proteins each composed of 5 transmembrane helices. The cytochrome-c subunit is electrostatically attached to the periplasmic side of the membrane, whilst the H-subunit inserts 1 transmembrane helix (A). It is also visible in (A) that a section of the M-subunit extends in to the cytochrome which appears to be involved in the tight binding between the two. There is an axis of pseudosymmetry between the L and M subunits which is easiest to see in image B.

another in a concave manner in the order; A,B,C,E,D. This arrangement produces an axis of pseudo-symmetry around the L and M subunits. This pseudo-symmetry, however, is broken as there are major structural differences in close proximity to the contact sites of the H- and cytochrome-*c* subunits [21].

The cytoplasmic and periplasmic faces of the L and M subunits form flat surfaces that interact with the H-subunit and cytochrome-*c* respectively.

The M-subunit is 50 amino acids longer than the L-subunit. These extra residues are involved in additional contacts to the cytochrome-*c* and the H-subunit. It can be seen in Figure 1-3 that the M-subunit extends into the cytochrome-*c* subunit with an α-helical 'arm'. Additional antiparallel β-sheets are also formed at the H-contact surface.

The cytochrome-*c* subunit is mainly composed of α-helices that run parallel to the plane of the 4 heme groups. There are additional antiparallel β-sheets that are involved in contacting the L-subunit [21].

The H-subunit contains 1 transmembrane α-helix. The rest of the protein is folded into a soluble globular domain at the cytoplasmic side of the complex [21,28].

The structure of the RC showed that the pseudo-twofold symmetry generated by the L and M subunits extended to their cofactors, which can be split into 2 branches, A and B. The cofactors in branch A are associated with the L subunit, whereas the cofactors in branch B are associated with the M subunit (Figure 1-4). These cofactors are very carefully positioned by the protein scaffold to ensure that they are separated from each other at a distance that is optimal for electron transfer reactions [21,24,29-31].

At the periplasmic side of the RC on the symmetry axis of the L- and M-subunits lie the $P_{A/B}$ Bchl molecules. They are non-covalently bound in a hydrophobic environment and positioned in such a way that their macrocyclic rings are almost parallel causing the Bchl molecules to be strongly electronically coupled (Figure 1-3). [22,28,32]. The Mg^{2+} ions in the centre of the Bchl molecules of $P_{A/B}$ and $B_{A/B}$ are liganded to His residues in the L and M subunits [21,28,32]. The B_A molecule lies in close proximity to $P_{A/B}$. As a result of a large number of studies the following electron transfer pathway has been generally accepted [33]. It was concluded that once the $P_{A/B}$ molecules were excited to a singlet state, $P_{A/B}$*, an electron is transferred to the B_A molecule in

approximately 3.5 ps. The $P^+B_A^-$ state is very short-lived and decays with a rate constant of 0.9 ps and the electron is further transferred to H_A. From H_A^- the electron is transferred to Q_A in 220 ps. The delay of the electron transfer here is thought to be due to larger distances between the donor and acceptor and results in a build-up of the $P^+H_A^-$ radical pair, making it readily detectable. From Q_A, the electron is then further transferred to Q_B in 100 μs. The transfer of the electron from Q_A to Q_B is influenced by the presence of Fe^{2+}, but the mechanism is not understood. When a single electron and proton have been accepted by the quinone bound in the Q_B binding site there is a charge separation across the membrane (P^+Q^-). The quinone molecule needs to accept 2 electrons and 2 protons to become fully reduced, so the system is reset by the cytochrome-*c* subunit which reduces the P^+ back to P. Another excitation results in further electron transfer, fully reducing the quinone bound in the Q_B binding pocket. Once it has been fully protonated, it loses its affinity for the Q_B binding pocket and is released from the RC [21,33-38].

Once the structure of the *Blc. viridis* RC was published, the coordinates were used in a molecular replacement (MR) strategy to solve the structure of the RC from *Rba. sphaeroides* by the Feher group [32,39-41]. It was important to solve the structure of the RC from this species as it is genetically amenable. Since the L and M subunits from *Rba. sphaeroides* and *Blc. viridis* share 50-60% sequence homology, it was no surprise that the structures were very similar. The main differences between the two are that the RC from *Rba. sphaeroides* doesn't contain the periplasmic cytochrome-*c* subunit, contains Bchl-*a* instead of the Bchl-*b* found in *Blc. viridis*, and ubiquinone, instead of menaquinone as the primary electron acceptor Q_A [28].

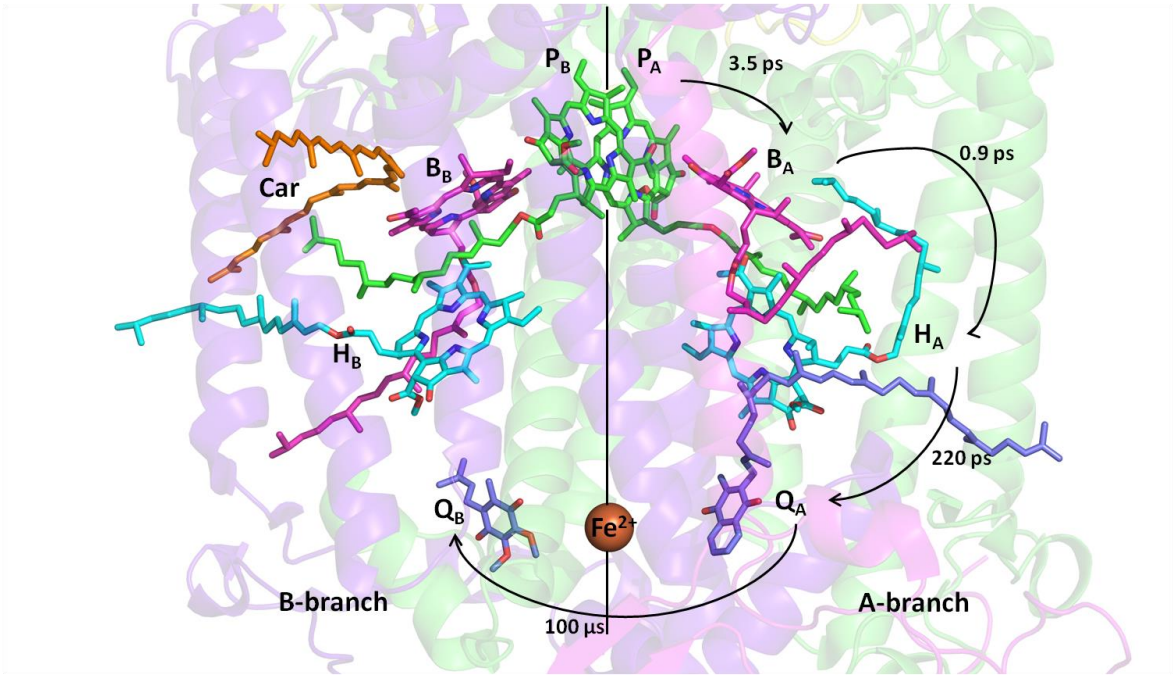


Figure 1-4 The cofactors of the RC (modeled in PyMol - 1PRC) from *Blc. viridis*. The cartoon of the protein scaffold is visible in the background. It is clear to see from this that the 2 branches of cofactors, B and A are more associated with M (purple) and L (green) subunits respectively. The A branch carries out the electron transfer. The special pair of Bchl ($P_{A/B}$) which lie on the symmetry axis (black line) accept an electron and become very powerful reducing agents allowing an electron to be transferred to the neighbouring cofactors down a redox gradient. 2 electrons are required to produce the fully reduced ubiquinol (QH_2). P_A P_B refer to the special pair of Bchl according to the branch they are more associated with. B_A and B_B are the monomeric Bchl, H_A and H_B are the bacteriopheophytin molecules, Q_A and Q_B are the quinone binding pockets. In *Blc. viridis* these bind ubiquinone and menaquinone respectively. Image adapted from [21,38].

1.3.2. LH2 complexes

The peripheral antenna complexes, the LH2 complexes, are usually composed of 8-9 dimeric subunits, but have also been observed to contain 12-13 subunits in *Allochromatium (Alc.) vinosum* LH2 complexes [42]. These subunits are heterodimers made up of small hydrophobic α and β apoproteins: the αβ-subunit. Both the α and β apoproteins form single membrane spanning α-helices [43-46]. The αβ-subunits are encoded by the *puc* operon [47,48]. The *puc* operon usually contains *pucBAC*, where *pucC* encodes a putative Bchl transporter that is involved in the expression of the LH2 complexes, *pucBA* encode the β and α apoproteins respectively. In some species, however, multiple *puc* operons have been observed, not all of which contain *pucC* and some that contain multiple *pucBA* gene pairs [44,49-

51]. These additional *puc* operons have been found dispersed over the genome. [44,50,51]. In some species, such as *Rps. palustris*, *Rps. acidophila* (strains 7050 and 7750) and *Alc. vinsoum*, these additional genes are thought to encode polypeptides that produce structural and spectroscopic variants of LH2 complexes when the bacteria are grown at different light intensities and temperatures. The spectral variants appear to be determined by a change in the Bchl binding sites within the protein scaffold that change the Bchl site energies [49,52,53]. Despite *Rps. acidophila* strain 10050 containing multiple *puc* genes, only one type of LH2 is expressed under all tested growth conditions. This is the most studied LH2 complex as it is spectrally and structurally well understood, and was one of the first membrane proteins to have its structure solved by X-ray crystallography [43,54]. Figure 1-5 shows the structure of *Rps. acidophila* strain 10050 LH2 complex. It is composed of 9 αβ-subunits that form a hollow cylindrical structure with an outer diameter of 34 Å formed by the β-apoproteins and an inner diameter of 18 Å formed by the α-apoproteins [43].

The αβ-heterodimers form membrane spanning α-helices that coordinate 3 Bchl-*a* molecules and 1 carotenoid molecule, rhodopin glucoside (RG). The protein complexes are essentially protein scaffolds that hold the pigments in the appropriate positions and distances from one another for efficient energy transfer. As such there are minimal protein-protein interactions between the αβ apoproteins within the transmembrane part of the dimeric subunits but prominent protein-pigment interactions [43,54,55].

The 3 Bchl-*a* molecules in the heterodimer unit are arranged into 2 different rings within the nonameric complex (Figure 1-5). One ring contains 18 excitonically coupled Bchl-*a* molecules that have a Q_Y absorption maxima at 850 nm (B850). These Bchl-*a* molecules are coordinated through their central Mg^{2+} ions by the conserved αHis-31

and βHis-30 residues and also form hydrogen bonds to αTrp-45 as well as αTyr-44 on the adjacent α-apoprotein.

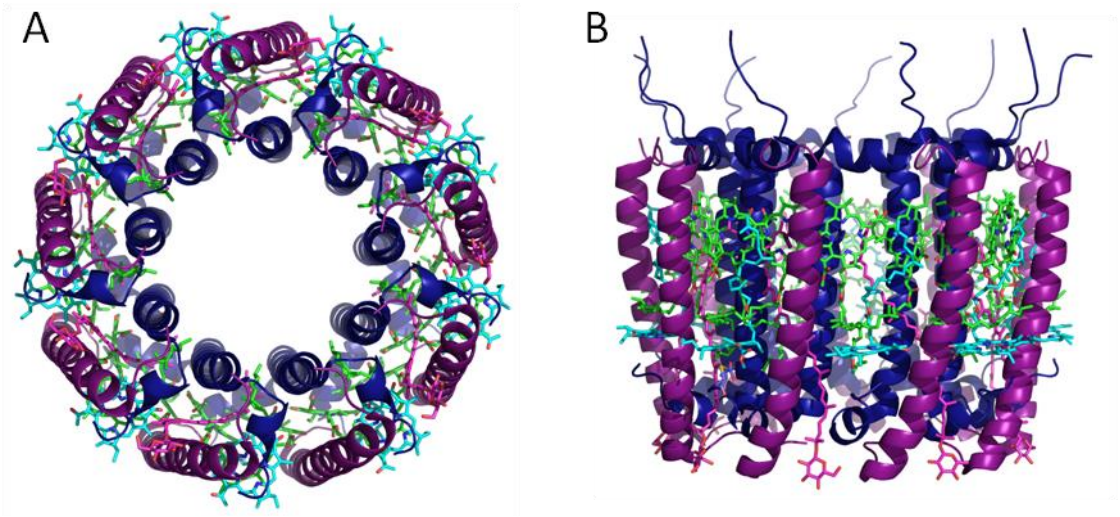


Figure 1-5 Structure of B800-850 LH2 from *Rps. acidophila* strain 10050 [43] as modelled in PyMol using 2FKW PDB file [56]. A: view from the periplasmic side of the membrane. The inner helices (blue) are composed of the α-apoproteins, whilst the outer helices are composed of the β-apoproteins. Between these two proteins, pigments are sandwiched. Each αβ subunit contains 3 Bchl-*a* molecules and 1 carotenoid molecule. The LH2 complex is formed of 9 αβ-subunits with a 9-fold symmetry forming a circular 'doughnut' structure. B: Looking through the membrane. The α-apoproteins are slightly longer than the β-apoproteins. The helices of these apoproteins bend at the periplasmic face and form hydrogen bonds with the adjacent subunits B850 molecules. These connections hold the complexes together as there is very little interaction between the two apoproteins themselves.

The residues present at positions α44 and α45 appear to affect the absorbance maxima of the aggregated Bchl molecules and changes from αTyr-44 to Phe or Met change the Bchl absorbance maxima from 850 nm (B850) to 820 nm (B820) [44,55]. The Bchl molecules that are coordinated at the α44-45 positions have their bacteriochlorin rings arranged perpendicular to the plane of the membrane [20,43,55]. The second ring is composed of monomeric Bchl-*a* molecules, coordinated by the N-terminal αfMet-1. The bacteriochlorin ring for these Bchl-*a* molecules lies parallel to the membrane plane, and has a Q_Y absorption maxima at 800 nm (B800). The shift in the Q_Y maxima is controlled by 2 things: site energy and the excitonic coupling. The site energy of the Bchl molecules is controlled by the protein environment. For example the B850 Bchl-*a* molecules are held within hydrophobic environments surrounded by aromatic residues, whereas the B800 Bchl-*a* molecules

are located in relatively polar environments. These changes in residues in the Bchl-*a* binding sites change the hydrogen bonding network to the Bchl molecules, hence change the site energy and shift in the maximal absorbance energy of the individual Bchl-*a* molecules. The red shift of B850 relative to B800 is exacerbated as the B850 molecules are orientated in a large, excitonically coupled aggregate. The larger the aggregate, the greater the red shift [20,43,55]. The β-B850 phytyl chain bends to contact the B800 phytyl chain (Figure 1-6 B) connecting the two Bchl-*a* rings [43].

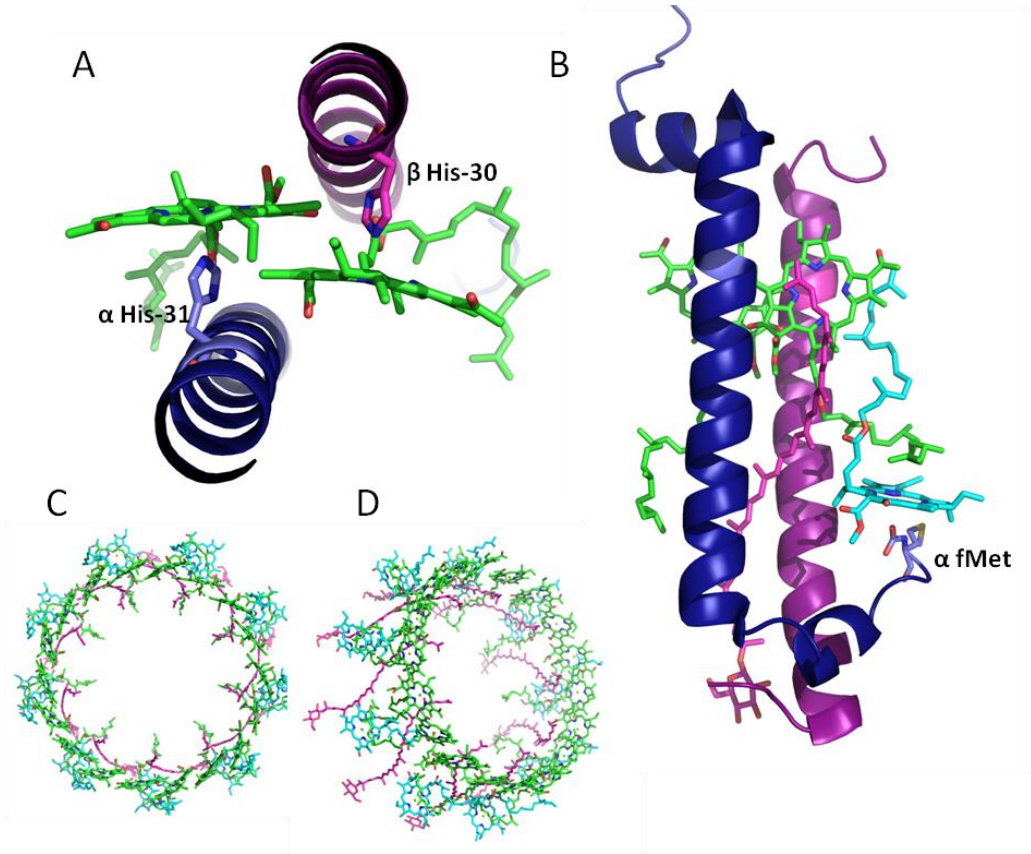


Figure 1-6 A: Coordination of B850 Bchl-*a* Mg^{2+} by the Nε of the conserved His residues αHis-31 and βHis-30 the α-apoprotein is shown in dark blue, and the β-apoprotein is shown in pink. The B850 molecules (green) are orientated in such a way that the bacteriochlorin rings lie perpendicular to the membrane plane and form an overlapping ring of electronically coupled molecules in the complete structure B: Coordination of B800 (cyan) Mg^{2+} via the O αfMet-1. The bacteriochlorin ring is positioned parallel to the plane of the membrane. In pink the carotenoid molecule (Car) RG is shown. It lies in an extended orientation, which contacts both B800 and B850 (green) molecules. The phytyl chains of the 2 B850 molecules lie in different orientations. The phytyl chain of the B850 molecule is more associated with the β-apoprotein bends to contact the phytyl chain of the B800 moleucle. C and D show the overall cofactor structure, which is composed of 2 rings of Bchl molecules, aggregates and monomers, with an intertwined Car molecule.

The RG molecule in B800-850 LH2 complex from *Rps. acidophila* lies in-between the α and β apoproteins in an extended conformation spanning the depth of the membrane (Figure 1-6 B-D). This extended conformation allows the RG to intercalate between the B800 and B850 Bchl-*a* molecules and is in van der Waals contact with them and the apoprotein subunits [55]. Carotenoid molecules have 3 functions in LH2 complexes; photoprotection by preventing the accumulation of reactive oxygen species (ROS) by quenching the Bchl triplet excited state and dissipating excess energy, structural integrity and assembly of photosynthetic complexes and finally in light harvesting and energy transfer (ET) [45,57]. They absorb wavelengths of light that are inaccessible to Bchl molecules and so increase the cross section for absorption of LH complexes.

An example of the B800-850 LH2 absorption spectrum from *Rps. acidophila* is shown in Figure 1-7. There are two clear peaks at 850 and 800 nm corresponding to the Q_Y transition of the strongly coupled and monomeric Bchl-*a* molecules respectively. The 590 nm peak corresponds to the Q_X transition of Bchl-*a* molecules, whilst the Soret peak at 390 nm corresponds to the absorbance of all Bchl-*a* molecules. The three peaks at 450-550 nm are due to the RG absorbance.

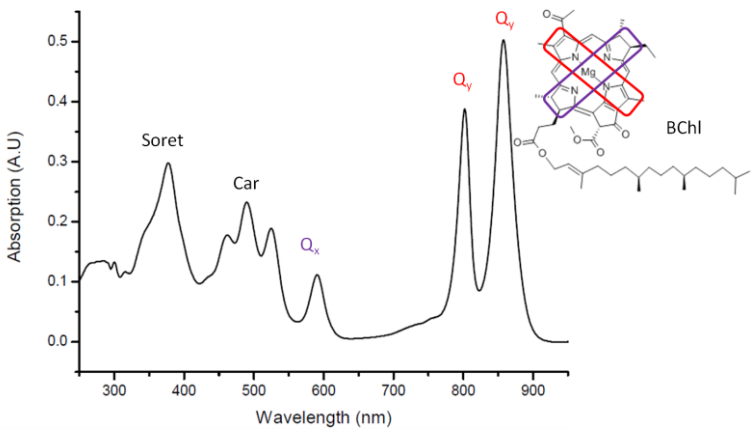


Figure 1-7 Absorption spectrum of B800-850 LH2 complex from *Rps. acidophila* strain 10050. Q_Y absorption of Bchl-*a* at 850 nm (B850) and at 800 nm (B800). Q_X absorption at 590 nm from all Bchl-*a* molecules present. The Soret band at ~390 nm is composed of the Q_Y and Q_X transitions from all Bchl-*a* molecules present. The carotenoid absorption is composed of 3 well defined peaks ranging from 450 - 550 nm.

1.3.3. LH1:RC 'core' complex

Structural studies of the LH1:RC 'core' complex have proved more elusive than the RC or LH2 complexes. As a result, only low resolution structures of these complexes have been solved [18,58,59] and efforts are still being made to obtain crystals of these complexes [60,61]. Core complexes have been shown to exist as monomers and dimers [18,58,59,62,63] but they always exist in a 1:1 ratio of LH1:RC.

1.3.3.1. *puf* operon and *pufX*

Purple photosynthetic bacteria contain the *puf* operon. This operon is composed of the genes that encode the α and β apoproteins of the LH1 (*pufBA*) complex and the L & M (and cytochrome-*c* if present) subunits of the RC (*pufLM(C)*). There are differences in the architecture of this operon between different species (Figure 1-8), for example, the *Rhodobacter* genus contains 2 additional genes; *pufQ* and *pufX*. Strains lacking *pufQ* and *pufX* produce a photosynthetic minus (PS^-) phenotype. The lack of *pufQ* results in a complete loss of expression of the *puf* operon as it encodes a Bchl biosynthesis protein [64], whereas $pufX^-$ impairs photosynthetic growth [62,65-68]. Some species, such as *Allochromatium* (*Alc.*) *vinsoum* and *Marichromatium purpatum* have additional *pufBA* genes downstream from *pufC*. For *Alc. vinosum* it has been shown that all 3 sets of *pufBA* genes are transcribed [69] but it is not currently known if all are expressed, and if so, in what growth conditions.

It was first concluded that PufX is required for photosynthetic growth in the *Rhodobacter* genus in 1989 [70]. A $pufLMX^-$ strain of *Rba. sphaeroides* was generated that presented a PS^- phenotype. When this mutant strain was complemented with the *pufLM* genes photosynthetic growth was not restored, however, aerobic growth was comparable to the wild type (WT) strains [65]. Similar studies in *Rba. capsulatus* showed that $pufX^-$ strains were PS^+, however growth rates were

significantly reduced relative to the WT [71]. Spontaneous suppressor mutants evolved from *pufX*$^-$ strains of *Rba. sphaeroides* in which the PS$^+$ phenotype was restored. Analysis of these mutants showed that point mutations had occurred in the *pufBA* genes and resulted in a decreased level of LH1 subunits surrounding the RC [62,66,68,72]. This led to the conclusion that PufX was only required if LH1 was present as RC$^+$ LH1$^-$ strains were PS$^+$ and cyclic electron transfer occurred as normal [14,19].

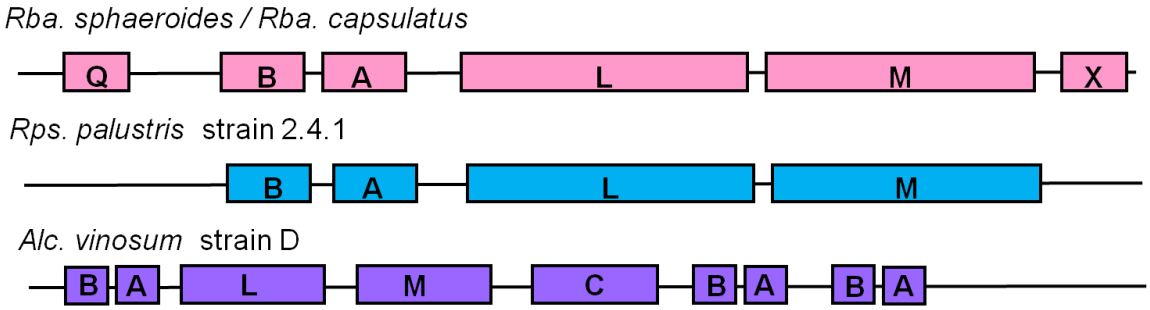


Figure 1-8 Schematic comparing the different *puf* operon architecture in different strains of purple photosynthetic bacteria: *Rhodobacter (Rba.) sphaeroides*, *Rba. capsulatus*, *Rhodopseudomonas (Rps.) palustris*, and *Allochromatium (Alc.) vinosum*. Only the *Rhodobacter* genera have the additional *pufX* gene downstream of *pufM*. This gene has been shown to encode a protein that indirectly facilitates efficient cyclic electron transfer within the PSU. No similar gene or protein product has been isolated in the other species. *pufQ* is a Bchl biosynthetic gene, *pufBA* encode the α and β apoproteins of the LH1 complex, *pufLM(C)* encode the RC L, M (and cytochrome-*c*) subunits. The H subunit is encoded by *puhA* approximately 40 kb from the *puf* operon [65].

It has been shown that PufX has an affinity with the α-apoproteins over the β-apoproteins of the LH1 complex. This affinity means that the presence of full length PufX inhibits the α-α reconstitution but not β-β apoprotein reconstitution [67]. Full length PufX proteins only inhibit α-α reconstitution from their own species, [67,73], however, chemically synthesised fragments of PufX proteins from either *Rba. sphaeroides* or *Rba. capsulatus* showed that hydrophobic transmembrane fragments of PufX exerted an inhibitory affect in cross-species experiments [73]. A slight inhibitory affect was also observed with *Rs. rubrum*. This is very interesting as this species produces a monomeric core complex with no orthologous *pufX* gene present [73], and has also been hypothesised, through low-resolution modeling, that the LH1 ring is complete and

closed around the RC [58,74]. Circular Dichroism (CD) studies provided evidence that this transmembrane fragment of PufX contains 47% α-helical secondary structure, which is similar to the α/β apoproteins of the LH1 complex [73].

Analysis of electron transfer within core complexes provided an explanation for the PS^-/reduced photosynthetic growth rates of *pufX*$^-$ strains. Using light-induced absorption spectroscopy [68,72] it was shown that *pufX*$^-$ strains were capable of generating charge separation as normal. With successive flashes of light it was observed that cyclic electron transfer was inhibited relative to WT [68]. Manipulation of the redox potential of the system also showed that PufX is required for cytochrome-bc_1 turnover and efficient Q/QH_2 exchange [66,72]. These experiments concluded that PufX appears to be important in the organisation of the functional PSU by disrupting the LH1 ring [62,68,72]. If the role of PufX in *Rba. sphaeroides* and *Rba. capsulatus* is structural organisation and disruption of the LH1 ring to allow cyclic electron transfer to occur, are there orthologous proteins in other species of purple photosynthetic bacteria or another route for the Q/QH_2 exchange?

1.3.3.2. Structure of PufX

Recombinantly expressed PufX from *Rba. sphaeroides* was analysed in organic and detergent solutions in order to elucidate the structure using solution nuclear magnetic resonance (NMR) spectroscopy (Figure 1-9) [75,76]. The solution showed a single α-helix that spans the membrane once. The hydrophobic stretch of residues in the core of PufX is 34 residues long (position 21-55), whereas a typical transmembrane helix is ~23 residues in length. The structure showed that this long hydrophobic stretch was accommodated by enforcing a bend in the helix. The C-terminus is disordered and exposed on the periplasmic side of the membrane, which allows for the first 12 residues of *Rba. sphaeroides* PufX to be post-translationally cleaved [67,73,75,76].

A series of N-terminal truncations showed that the presence of the bend in PufX is very important in facilitating the formation of dimeric core complexes in the *Rba.* genera [76]. The inside of this bend is populated by small hydrophobic Glycine (G) residues in a conserved GXXXGXXXG motif. When the G residues were mutated to Leucine (L), no affect was observed on dimer formation [77] and it was concluded that GXXXGXXXG was a pseudo-motif.

G29 is a conserved residue found on the inside bend of PufX. Mutants were generated at this position incorporating increasing sizes of hydrophobic side chains (G29A, G29V, G29I & G29L) and the formation of dimeric core complexes was monitored. The G29A mutant resulted in reduced populations of dimeric core complexes, while the formation of dimers was abolished with G29V, G29I and G29L substitutions [76]. This lead to the conclusion that G29 is required for the formation of dimeric core complexes.

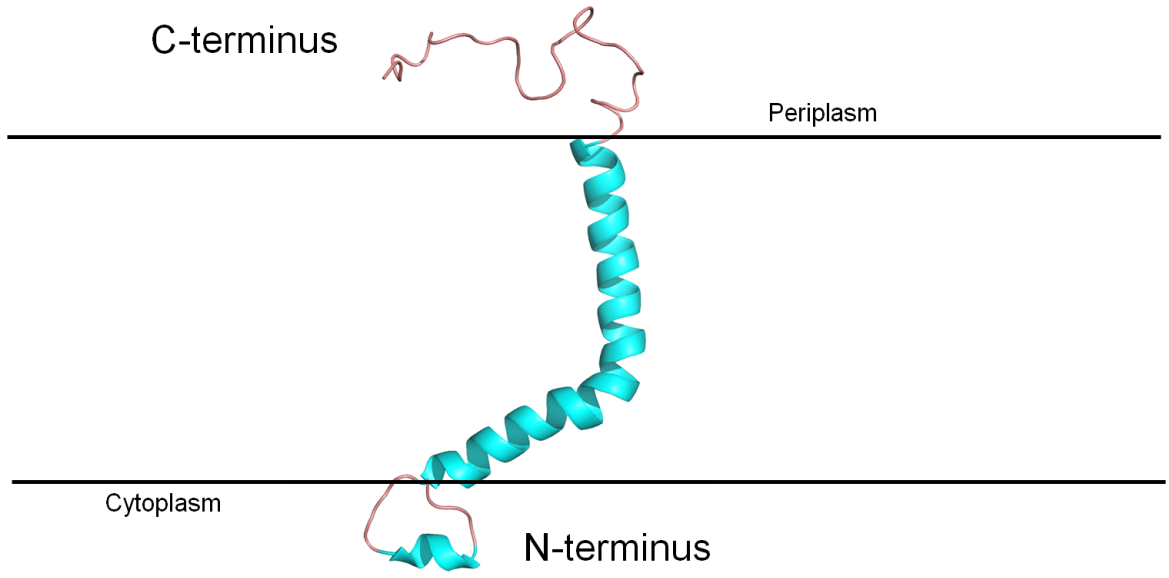


Figure 1-9 Structure of PufX from *Rba. sphaeroides*. Modelled in PyMol using the 1NRG PDB.code [75]. The disordered C-terminus is exposed on the periplasmic side of the membrane and is post-translationally cleaved early on in the membrane insertion process [77]. The hydrophobic fragment of PufX composed of 34 residues forms 2 helices that are connected by a well-ordered helical bend. The N-terminus is on the cytoplasmic side of the membrane. It is very important in forming the bend of the helix and facilitating the formation of dimeric core complexes [75,76].

1.3.3.3. Structure of LH1:RC 'core' complex

So far only low resolution structures have been elucidated for both dimeric and monomeric LH1:RC 'core' complexes. The LH1 complexes are composed of similar αβ-subunits to the LH2 complexes. Dimeric core complexes have been viewed using electron microscopy (EM) within the native membrane [63] and from 2-dimensional (2D) crystals [59] both from *Rba. sphaeroides*. These analyses both presented an elongated 'S' shaped structure composed of 2 'C' shaped LH1 complexes surrounding the RCs (Figure 1-10).

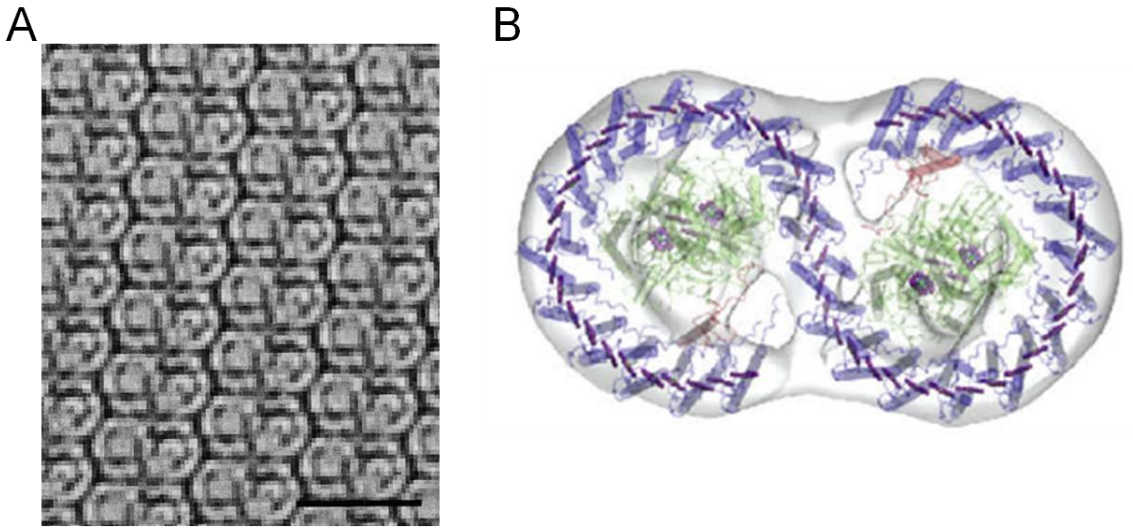


Figure 1-10 Low resolution structure of dimeric LH1:RC:PufX 'core' complex from *Rba. sphaeroides*. A: Filtered image of masked Fourier Transform of electron micrograph, 20 Å Bar = 200 Å. Regular array of dimeric core complexes show an 'S' shape superstructure [63]. B: 8.5 Å cryoEM structure of *Rba. sphaeroides* dimeric core. 'S' shaped LH1 complex is in blue. The red helices show extra density that has been proposed to be where PufX is located. In the middle of the 2 'C' shaped LH1 complexes is the RC [59].

Resolving the structural model of the dimeric core at 8.5 Å was aided by the crystal structure of the RC from *Rba. sphaeroides* [40] that allowed the precise positioning of the RC within the LH1 complex [59]. In each individual 'C' shaped LH1 complex, 14 αβ LH1 subunits were assigned. The density for 12 subunits (subunits 2-13) was well defined and formed the arc of the LH1 complex. Subunits 1 and 14 were positioned near the dimer interface, however, the densities at these positions were not well defined and couldn't be seen at lower resolutions of negatively stained dimers. The positions of these subunits in the final structure

was chosen due to spatial arrangements and distances between the subunits [59].

The position of PufX outwith the arc of the αβ subunits was assigned to an area with well defined density comparable to that of the αβ subunits. This position would allow interactions between PufX and the α-apoprotein, and lies close to the Q_B binding site within the RC [59].

At the recent PS16 conference in St. Louis, MO, USA, the Hunter group presented an unpublished 8 Å model of the dimeric core complex generated from a low-resolution X-ray crystallographic data set. The model was produced using MR. A search model was generated composed of the RC from *Rba. sphaeroides*. The LH1 complex was modelled against *Rs. molishianum* LH2 complex, the EM model discussed previously and the solution structure of PufX. This low-resolution model showed similar overall features, however the PufX proteins were in line with the α-apoprotein ring instead of lying in between the LH1 complex and the RC.

A more recent structure of a dimeric core complex from *Rhodobaca* (*Rca.*) *bogoriensis* (Figure 1-11) elucidated by single particle EM [78] provides a slightly different picture of the dimeric core complex organisation. The low resolution models of *Rba. sphaeroides* and *Rca. bogoriensis* are, at first glance, similar. However, there are some very interesting differences. In *Rca. bogoriensis* 13 αβ subunits were assigned to form each discontinuous arc of the LH1 complex with an additional 2 subunits positioned at the dimer interface. Whereas, in *Rba. sphaeroides* the LH1 complex appears to adopt a continuous 'S' shaped LH1 complex. The most striking difference between the two structures is the presence of 4 PufX proteins per dimer in *Rca. bogoriensis*, relative to 2 PufX proteins in *Rba. sphaeroides.* These differences could be species related or due to low resolution modelling. There is currently no known stoichiometry of LH1:RC:PufX for *Rca. bogoriensis* [78]. Although this has been put forward, Professor Neil

Hunter disagrees with this due to the differences in stoichiometry (personal communication). This needs to be further investigated before any real conclusions can be made. These apparent differences probably reflect over interpretation of low-resolution data in the case of the single particle reconstructions.

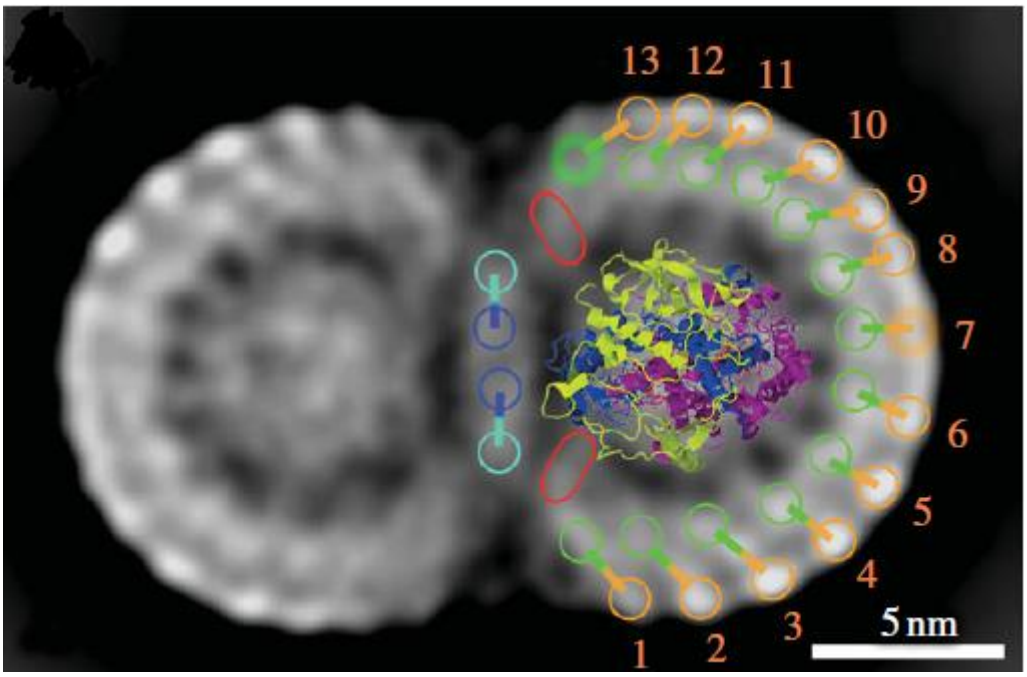


Figure 1-11 Structure of *Rca. bogoriensis* dimeric LH1:RC:PufX core complex elucidated by single particle EM [78]. Structure shows 2 clear semi-circles of αβ LH1 subunits (orange and green circles) with 2 more αβ subunits at the interface (blue and cyan circles). 2 PufX molecules are shown in red spheres. The cartoon structure in the middle of the LH1 complex is the RC.

Monomeric core complexes have also been modelled using EM. An 8.5 Å model [58,74] of the *Rs. rubrum* core complex showed that the LH1 complex forms a complete ring composed of 16 αβ LH1 subunits that surround the RC continuously [58,74]. There is some controversy over this structure as Karrasch used reconstituted LH1 complexes. As discussed in section 1.3.2.1, LH1 reconstitution is inhibited by the presence of PufX. Reconstitution using only α and β apoproteins will therefore only form a complete LH1 ring [58]. More recent studies using native core complexes also presented a closed LH1 complex surrounding the RC [74], however, as the monomeric core complex has rotational freedom it is possible that at this resolution the electron

micrographs have produced structures with continuous LH1 complexes as an artifact of rotational averaging. A continuous LH1 complex would create a barrier to the Q/QH_2 exchange and cyclic electron transfer.

Molecular dynamic simulations have been performed using the low resolution *Rs. rubrum* core complex model and suggests that a hydrophobic pathway exists through the complex and LH1 ellipse. This pathway could allow the diffusion of QH_2 with a minor protein backbone deviation of 1.6 Å in approximately $\leq 8 \times 10^{-3}$ s [79]. This suggests that a protein orthologous to PufX may not be required for efficient cyclic electron transfer in monomeric core complexes, however, this work was carried out on a low resolution model and as such caution would be advised when analysing the data.

To date, the highest resolution model of a core complex is of the monomeric core complex from *Rps. palustris* strain 2.1.6 [18]. The structure was solved by X-ray crystallography to 4.8 Å. At this resolution atomic details cannot be elucidated, however, secondary structures and the positions of the subunits are able to be clearly assigned. The coordinates of the *Rba. sphaeroides* RC [40] were used for MR to provide phase information. This low-resolution structure was very interesting as, similar to *Rs. rubrum*, *Rps. palustris* does not appear to contain an orthologous *pufX* gene. In the structure, an elliptical LH1 complex was observed, with dimensions of 110 Å by 95 Å for the outer ring of β apoproteins, formed by 15 αβ subunits, instead of the 16 αβ-subunits observed in *Rs. rubrum*. In place of the 16th subunit an additional peptide was modelled in the α-ring, however, there was no density in the putative 16th subunit position in the β-apoprotein ring. The density for the 16th α-subunit was clear, yet showed an α-helix that adopts a disjointed position within the α-apoprotein ring and would not fit into a discontinuous ellipse. This peptide, denoted Protein W, disrupts the ellipse of the LH1 complex and is positioned on the opposite side of the RC with respect to the H-subunit. Both these subunits break the

two-fold crystallographic symmetry and are thought to be involved in the structural organisation of the complex. Protein W is shown to be located close to the Q_B binding site and creates a gap in the LH1 α-helix pallisade that is large enough for a quinol molecule to diffuse through. It was, therefore, hypothesised that Protein W is orthologous to PufX [18,45,80]. To date it has not been possible to isolate Protein W from purified core complexes, hence, many questions remain regarding its location within the genome, the level of sequence homology to PufX and whether it is integral to efficient cyclic electron transfer.

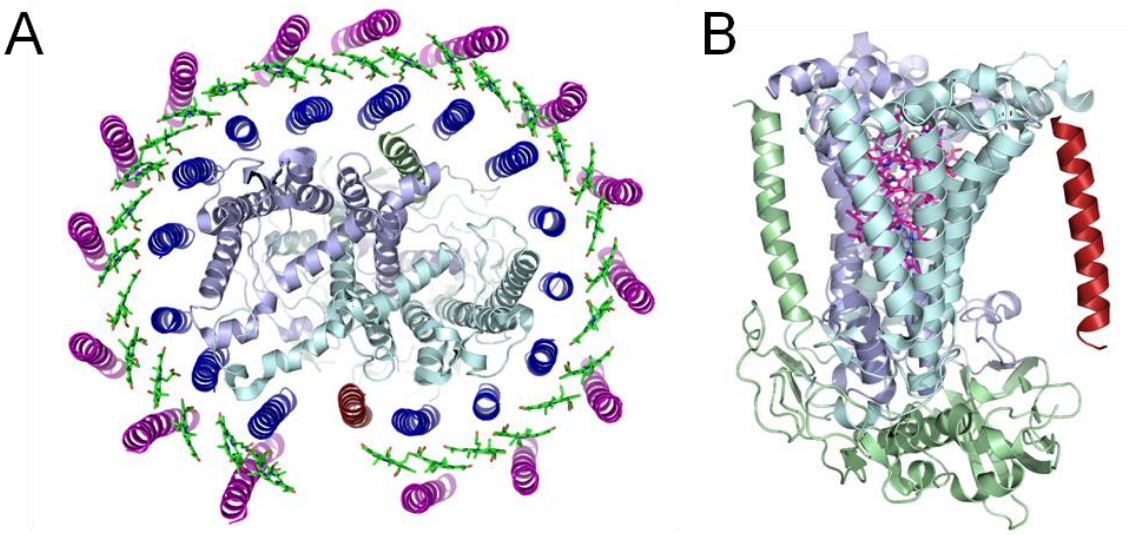


Figure 1-12 Structure of monomeric core complex from *Rps. palustris* modelled in PyMol using 1PYH PDB coordinates [18]. A: view from periplasmic side of the membrane looking through the complex. Magenta - β apoproteins, blue - α apoproteins, red- Protein W. The pastel colours in the middle of the LH1 ellipse represent the RC. This view shows how Protein W disrupts the LH1 ellipse and provides a gap in the complex large enough for a quinol molecule to diffuse through. B: Through the membrane view of the RC and Protein W. Pale green - H-subunit of RC, red - Protein W of LH1 complex. Pale blue - L subunit of RC, pale purple - M subunit of RC. This view shows H and W opposite one another, breaking the pseudo-symmetry of the RC [18]

1.4. Energy transfer

Electronic energy transfer (EET) or Förster energy transfer is a transition dipole-dipole interaction between a donor molecule and an acceptor molecule (Equation 1). The strength of these interactions (k) is dependent on the distance from the centre of the donor to the centre of the acceptor molecules (R), the spectral overlap of the donor emission

and the acceptor absorption (I), the fluorescence quantum yield (φD) and the lifetime (τD) of the donor molecule [81]. This type of ET can occur over relatively large distances between donor and acceptor molecules with weak coupling.

$$k = \frac{\varphi D}{\tau D}\left(\frac{R_0}{R}\right)^6 = \frac{\varphi D}{\tau D}\left(\frac{8.785 \times 10^{-25} I}{n^4 R^6}\right)$$

Equation 1: General Förster equation. *k:* strength of these interactions, *R:* distance from the centre of the donor to the centre of the acceptor molecules, *I:* spectral overlap of the donor emission and the acceptor absorption, *φD:* fluorescence quantum yield, *τD:* lifetime of the donor molecule.

In bacterial LH complexes the chromophores are organised very close together, within 10 Å of one another, and in many cases are touching. These short distances enable very strong coupling and an overlap of their electronic wave functions. The overlap of the wave functions between donor and acceptor molecules can allow ET by electron exchange between these molecules via the Dexter mechanism. Figure 1-13 compares Förster and Dexter ET. It is clear from this schematic that triplet states, common in photosynthesis, can only be transferred or annihilated using the Dexter ET mechanism as Coulombic exchange would violate the Wigner spin conservation law [82].

However, recently the traditional Förster and Dexter mechanisms of EET have been shown to fail to predict the measured times for EET in photosynthetic antenna complexes [83,84]. As a result both these mechanisms have been modified by Gregory Scholes and co-workers to account for intermediate strength coupling [83]. These new features complicate the EET mechanisms in LH complexes, and as such they are still under intense investigation.

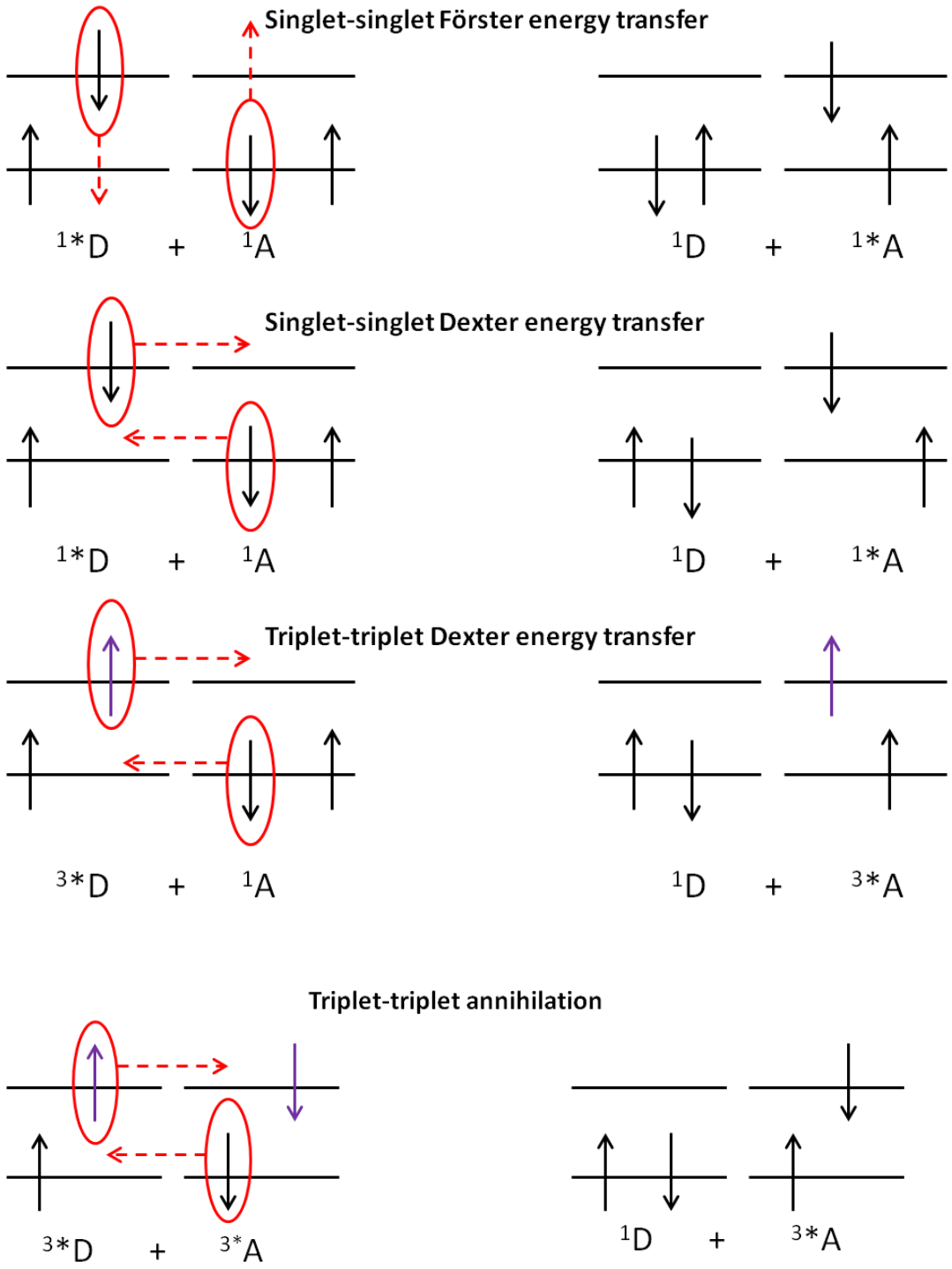


Figure 1-13 Schematic comparing Förster and Dexter ET mechanisms. Förster ET occurs when two molecules, a donor and acceptor, are close enough for coulomb interactions to occur, and for their emission and absorption spectra to overlap. Energy is therefore transferred from the donor to the acceptor. Dexter ET occurs when the donor and acceptor molecules are within 10 Å of each other, and their wavefunction as well as their spectra overlap. This allows an electron from the donor to be exchanged in to the acceptor molecules electron cloud. This is very useful for triplet state sensitisation and annihilation.

1.4.1. Photosynthetic pigments and energy transfer

When B800 is excited in the LH2 complex the excitation energy is rapidly transferred to the B850 molecules. At room temperature (RT) the rate constant for this is 0.7 ps. Even at liquid helium temperatures this rate only slows down to about 2 ps. [85]. Single molecule spectroscopy (SMS) has been especially useful in providing evidence about the correct way to think of the Bchl molecules in the B800 and B850 absorption bands. SMS has shown that there are narrow lines at B800 region in the fluorescence excitation spectrum which show that the B800 are largely monomeric. The broad absorption bands in the B850 region clearly show the strong excitonic nature of the B850 bands. From the first approximation, it is possible to interpret that energy is transferred from B800 monomeric Bchl molecules to the B850 aggregated Bchl molecules, which, behave as a collective supermolecule [57,85,86]. The energy is then transferred from the excited B850 molecules to the B875 molecules of the LH1 complex. It is hypothesised that the rate of ET is maximised as the bacteriochlorin rings in LH2 and LH1 are organised to be as close as possible to each other at the same depth within the membrane. This aligns the Q_Y transition dipole moments so that they are favourable for ET from LH2 to LH1 [20].

Carotenoids (previously discussed in section 1.3.1) are molecules that contain extended linear conjugated π-electron systems that can differ in conjugation length (*N*). Carotenoids absorb light at wavelengths inaccessible to Bchl molecules and, importantly, where the intensity of solar radiation is maximal [87]. This strong absorption occurs in the visible region usually between 450 and 550 nm. The photochemical behaviour of carotenoids is relatively complicated. The strong one photon absorption that gives rise to the well-known absorption band between 450 and 550 nm is due to the S_2 state. The lower lying S_1 state is one photon forbidden for reasons of symmetry. Figure 1-15 shows the standard model for the carotenoid excited states [88] [87,89-91] and is composed of two low lying excited singlet states, S_2

($1^1B_u^+$) and S_1 ($2^1A_g^-$) and a ground state S_0 ($1^1A_g^-$). Although one photon excitations of the S_1 state are symmetry forbidden, they can still be populated via internal conversion (IC) from the S_2 state, or by a two photon excitation. As such the S_1 state is poorly populated relative to the S_2 state and is termed a 'dark' state [57,87,89-93]. It has been predicted by Tavan and Schulten that there are additional electronic states in carotenoid molecules, some of which lie in between the S_2 and S_1 states [88]. It has been very difficult to try to observe these states and their importance due to the incredibly short lifetimes of these states.

The S_2 state is relatively short lived, and as such, carotenoids are not obvious candidates as light harvesting molecules. However, ET from carotenoid to Bchl-*a* molecules occurs in 200-300 fs with near 100% efficiency in some species [57,86,87,89,94]. It has also been observed that increasing the *N* conjugation length produces an increase in the energy gap between the S_2 and S_1 energy states, ranging from 5800 cm^{-1} to 6600 cm^{-1} for $N = 9$ to $N = 11$ carotenoids, whereas the lifetime of the S_2 state decreases from 330 fs to 130 fs [95]. This should drastically reduce the ET efficiency between the carotenoid and Bchl molecules, however, experimentally this has not been observed.

Figure 1-15 shows the energy levels of the carotenoid and Bchl molecules. The dynamics of this system are ultrafast and so very short laser pulses were required to delineate the exact pathways of IC and ET present [94]. Once the system was excited and the S_2 state became populated, energy was transferred to the Q_X of B850 and B800 Bchl molecules in 200 and 300 fs respectively. However, a competing channel was observed, IC between S_2 - S_1 (IC_{21}) that occurred after approximately 100 fs. This energy was not necessarily lost to light harvesting as, if the S_1 energy level is higher than that of the Q_Y energy level, ET can occur in approximately 30 ps [57,94,96]. Once ET to the

B850 Q_Y has occurred, the energy decays into the ground state in ~1 ns [57,94].

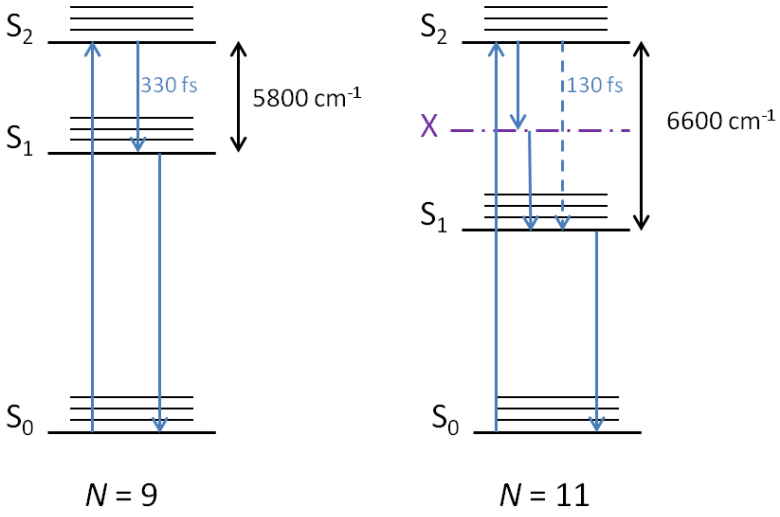


Figure 1-14 Schematic energy levels of carotenoids with *N* = 9 and *N* = 11. Each energy level has vibrational states, represented by thinner lines above the energy levels. Energy is transferred from S_0 to the symmetry allowed S_2 state. The lifetime of this state is 330 fs and 130 fs for *N* = 9 and *N* = 11 carotenoids respectively. The energy gap between S_2 and S_1 increases with increasing *N*. It has been proposed [88] that intermediate energy states lie between S_2 and S_1 that might act as an intermediate step in IC_{21} in such circumstances.

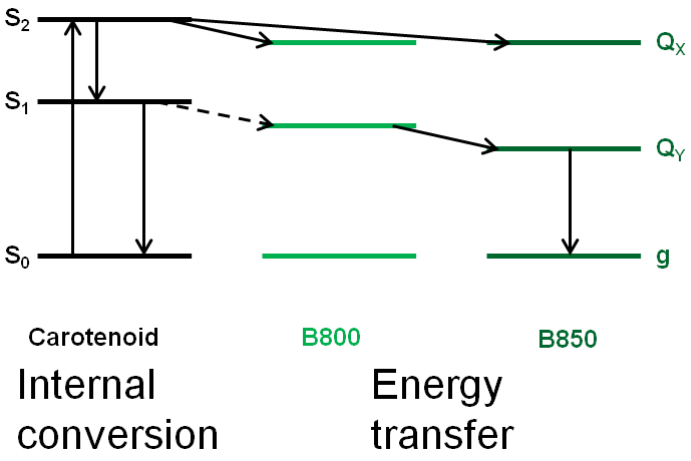


Figure 1-15 Schematic energy level diagram of a carotenoid molecule and B800 and B850 Bchl-*a* molecules. Once excited the carotenoid is promoted from ground state to the first excited singlet state, S_2. This state is short lived. In 200 - 300 fs energy is transferred to B850 and B800 Q_X transitions respectively. It can also decay to S_1 via IC in 100 fs. Although energy transfer between Carotenoid and Bchl molecules generally occurs through S_2 - Q_X pathway, in some cases, if the energy level of the S_1 is higher than the Q_Y, S_1 - Q_Y pathway is also observed with rates of 30 ps. Energy is transferred to the B850 Q_Y transition in 800 fs, where it decays to ground state in 1 ns.

Calculations of carotenoid excited states proposed that carotenoid molecules with $N \geq 9$ could have additional excited states [88] in

between the S_2 and S_1 energy states. This could explain why ET in LH2 complexes containing carotenoids with N = 11 is so efficient as they could act to bridge the increasing energy gaps. In 2002, ultrafast spectroscopy was used to probe the early events of carotenoid relaxation and an additional peak was indirectly observed [90]. Time delays showed that the S_2 state of β-carotene decayed very rapidly, and within 50 fs a new peak appeared at a lower energy. The S_1 energy level appeared with a 50 fs delay, and so it was hypothesised that this new peak was caused by an intermediate electronic carotenoid state that lies in between the S_2 and S_1 states [89,90]. This hypothesis was further corroborated by testing carotenoids with N = 5, 9, 13 and 15 [91]. The results from these experiments concluded that N = 5 fulfilled the three-level model and no intermediates were observed. Whilst N = 9, 13 and 15 all produced additional peaks that were suggested to be intermediate carotenoid electronic states, therefore requiring a four-level model to explain IC [89,96] [91].

The existence of intermediate carotenoid electronic states is still very controversial and as yet they have not been directly or conclusively observed.

1.5. Principles of 2 dimensional electronic spectroscopy (2DES)

2-dimensional electronic spectroscopy (2DES) is a technique analogous to 2-dimensional nuclear magnetic resonance (2D-NMR) [97,98]. 2DES correlates excitation and emission energies as a function of a time delay between excitation and emission events, within the bandwidth of the laser pulse. It offers more detailed observations of light harvesting efficiency relative to steady state and time resolved absorptive and emission spectra as it spreads out the time evolution of the spectrum in 2 dimensions. This enhances spectral resolution and allows the pathways of relaxation to be directly observed. Another benefit of this new technique is that the 'photo-congestion' from the many combined

IC and ET events is reduced. This allows more 'subtle', or less populated states, such as the controversial intermediate states (section 1.4), to be observed [98].

Figure 1-16 shows a schematic of a 2DES experimental set up. Briefly, a Ti:Sapphire laser emits ultrafast pulses, 150 fs, that are amplified and tuned to the energy of excitation using a non-collinear parametric amplifier (NOPA). The resulting beam is split using diffractive optics into 4 individual beams arranged in a box geometry. Delays are incorporated by passing the beams through wedges.

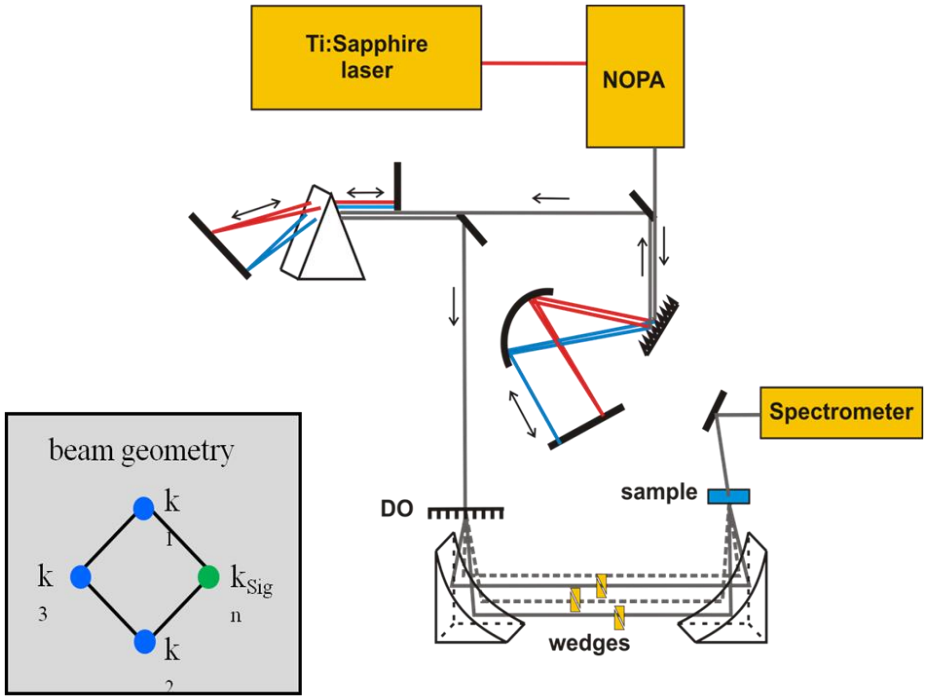


Figure 1-16 Schematic of 2DES experimental set up. Ti:Sapphire laser generates ultra-fast laser pulses that are amplified and tuned using a non-collinear optical parametric amplifier (NOPA). The beam was split using diffractive optics generating 4 beams that were arranged in a box geometry. Time delays are generated by passing the beams through wedges. The signal is detected by a spectrophotometer.

Figure 1-17 outlines the different laser pulses used in 2DES. Pulses k_1 and k_2 excite the system. This allows coherent ET between the cofactors, such as Bchl molecules, to occur. A delay is incorporated between k_2 and k_3 . This allows the evolution of excited states and ET to be observed and kinetics of the system extrapolated. Pulse k_3 detects the signal which is recorded using a spectrophotometer. The delay

between k_2 and k_3 can be controlled. The evolution of the signal can be monitored by varying the delay between k_2 and k_3. The signals are heterodyne detected. In heterodyne detection, the beam k_3 and k_{LO} are temporally and spatially overlapped. This allows detection of the phase of the signal, as well as isolating the real (absorptive) signal from the imaginary (dispersive) signal, therefore amplifying the signal.

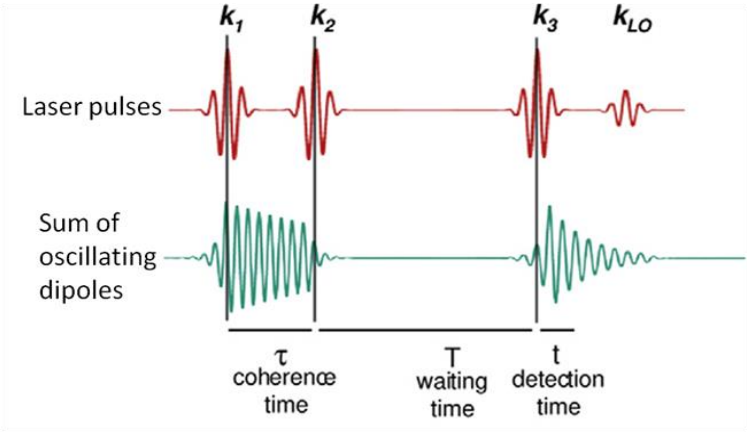


Figure 1-17 Schematic showing the 4 laser pulses k_{1-3} and k_{LO}. The first 2 laser pulses, k_1 and k_2, excite the system and allow it to enter coherence. Different delay times of k_2 are used to monitor the evolution of IC and ET within the system. Pulse *k3* detects an emission signal. The time between k_2 and k_3 is called the waiting time and is where the photophysics of the system is monitored. k_{LO} is a reference beam that amplifies and provides phase information on the system. Image taken from [98].

1.6. Crystallisation of membrane proteins

The pigment-protein complexes that form the PSU within the ICM are integral membrane proteins. Just as with any protein, structural information is required in order to gain a better understanding of their function. However, membrane proteins are notoriously difficult to work with and used to be considered 'impossible' to crystallise [99-101]. In 1985 the first membrane protein structure was resolved to 3 Å using X-ray crystallography [21]. Since then, despite being difficult to work with, the number of high resolution structures of membrane proteins has increased exponentially (Figure 1-18). There are currently 415 unique membrane protein structures published and 1292 structures in total (2nd August 2013, http://blanco.biomol.uci.edu). However, this is

still a small number compared to the total number of solved structures of water soluble proteins, 91397.

The main difficulty of working with membrane proteins is that they are, by definition, embedded in a lipid membrane bilayer. In order to extract them detergents are required to disrupt the native membrane (Figure 1-19), as discussed in section 1.5.1. However, solubilisation is a relatively harsh process and often results in the denaturation of the membrane protein under study. It can be challenging to find suitable detergents that are strong enough to solubilise the required membrane protein yet mild enough to leave its structure intact. [99,100,102]. Detergents have often been shown to induce aggregation, destabilisation and loss of function of some membrane proteins [99-101]. However, homologous proteins from different species can behave differently in the same detergent environments, and so a common technique employed by structural biochemists is to test a range of homologous protein targets from different species in order to find one that is stable when solubilised.

By testing homologues of the protein of interest it can be possible to isolate a more stable and robust protein for structural studies with the same function. More recently systematic site directed mutagenesis has also been used to engineer thermostable mutants of a membrane protein of interest, for example the β-adrenergic receptor [103]. Another common problem that is usually encountered when working with membrane proteins is that they are not typically abundant. Recombinant host systems for over-expressing membrane proteins, such as in *Escherichia (E.) coli* and yeast, have improved dramatically in recent years. This has led to an improved success rate for the recombinant overexpression of low abundance proteins of interest. However, there is often still a bottleneck in structure elucidation due to problems with construct design, establishing the optimum growth conditions for over-expression and correct protein folding

[100,101,104]. Purple photosynthetic LH complexes are generally naturally abundant, coloured and have well defined absorbance spectra that allows the stability and function of the complexes to be followed throughout the purification and crystallisation procedures [22].

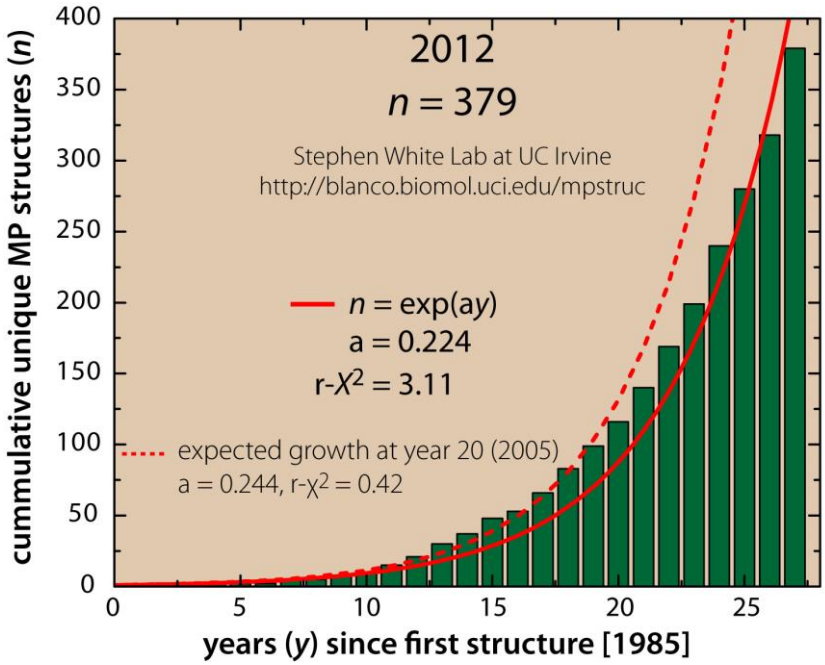


Figure 1-18 Graph showing the number of unique membrane protein structures deposited in the Protein DataBank since 1985 (http://blanco.biomol.uci.edu/mpstruc/listAll/list).

1.6.1. Detergents

Detergents are amphipathic molecules that self assemble to form micelles. The hydrophobic tails of the detergents surround the hydrophobic sections of membrane proteins allowing the polar head-groups of the detergents to face the aqueous solution so that the membrane proteins are then soluble in an aqueous solution [99]. The micelle size of detergents depends on the alkyl chain length (the longer the chain length, the bigger the micelle). Detergents with larger micelles are milder and less destabilising than detergents with smaller micelles, as less of the protein is exposed to solvents. Harsh, short chain detergents can induce a loss of function and protein aggregation and, therefore, are generally not used in membrane protein crystallisations [100,105].

Membrane protein crystal formation is dependent on having a portion of the protein outside the detergent micelle in the soluble phase, as it is these exposed hydrophilic domains that form the intermolecular contacts within the crystal lattice. This means that although larger micelle detergents can be more stabilising, they may not lead to crystal formation or high resolution structures as they will cover the hydrophilic regions of the proteins that are required for crystal formation. It was recently shown through an extensive study, that a range of membrane proteins crystallised in β-dodecylmaltoside (β-DDM), a C12 mild non-ionic detergent, showed poor diffraction (~4 - 8 Å). When the same membrane proteins were purified in N,N-dimethylamine-N-oxide (LDAO), a C12 zwitterionic harsh detergent with a smaller micelle, high resolution data was obtained for the ammonia transporter AmtB (1.9 Å) [105]. As a first step membrane proteins are often solubilised in β-DDM. Although there is no general rule for membrane proteins, β-DDM has been shown to be a good 'first choice' for membrane transporters especially. Protein detergent complexes can be analysed by size exclusion chromatography (SEC) and functional assays to assess the proteins uniformity and stability respectively. The width, size and symmetry of the size exclusion chromatogram peaks are analysed. If sharp, symmetrical monodisperse peaks are obtained, the protein is deemed to be crystallisation quality, although, this is not true for all cases [101,105-107]. If the protein is not monodisperse then either different constructs or homologues should be tested [104].

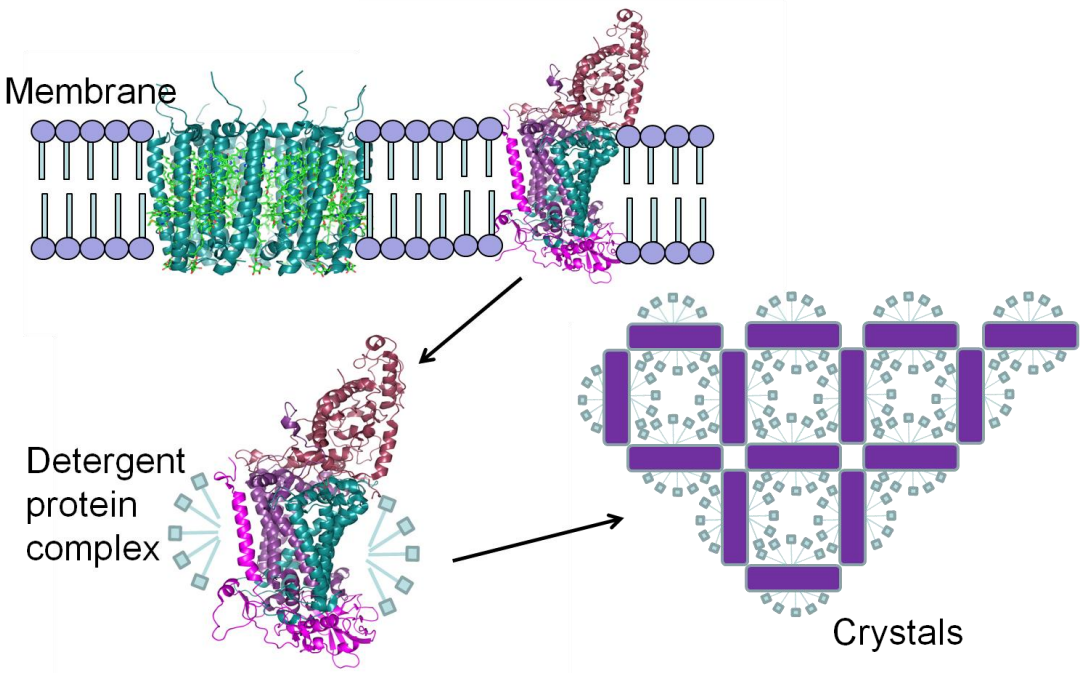


Figure 1-19 Schematic showing membrane protein isolation using detergents and crystal formation. Membrane proteins are imbedded in a phospholipid bilayer. To extract the proteins detergent molecules are required. Detergent molecules are amphipathic and self assemble to form micelles which surround the hydrophobic sections of the membrane proteins. Detergents have varying head groups and alkyl chain lengths. Different head groups are favoured by different proteins. The alkyl chain length dictates the size of the micelle. The larger the alkyl chain, the bigger the micelle. Larger micelles occlude more of the soluble sections of the membrane proteins and so are undesirable for membrane protein crystallisation, as crystals are formed by regular protein to protein contacts which are formed by the soluble sections of the protein outside the detergent micelle.

1.6.2. Crystal growth

1.6.2.1. Vapour diffusion

Crystals are regular arrays of precipitated material. In order for crystals to grow the solutions should be supersaturated. Proteins, however, are more difficult to crystallise. The solubility of the protein has to be lowered, and is driven out of solution using precipitants such as salts and polyethyleneglycol (PEG). Vapour diffusion is one of the most widely used techniques for crystal growth [108]. There are two types of vapour diffusion: sitting and hanging drop (Figure 1-20). As membrane proteins are in a detergent solution, the surface tension of the solution is reduced. This means that, if large volumes of the protein detergent complex are required, the hanging drop vapour diffusion method may not work as the drop would fall in to the reservoir. The sitting drop

method was designed to counter this difficulty. Both these techniques have a protein drop containing a small amount of the precipitant. The concentration of the precipitant in the protein drop is just less than that required to form crystals. Surrounding the sitting or hanging drop is a reservoir containing the precipitant solution at a higher concentration than that of the protein drop, causing the protein drop to equilibrate against the reservoir solution. This abstraction of water from the drop to the reservoir causes the protein solution to reach saturation. Usually the protein will simply precipitate out of solution as an amorphous aggregate, however, under very specific conditions once the protein solution is supersaturated and enters the labile zone, protein molecules can aggregate and form nuclei. If the nuclei are stable and remain solid, more molecules of protein can be recruited onto its growing surfaces. As the solid crystal grows the concentration of the protein decreases and the system reenters the metastable zone, where the crystals can continue growing [108,109]. Ideally, the protein is precipitated in a slow and controlled manner, with the system close to the metastable phase in non-denaturing conditions. This slows down the rate of nucleation and crystal growth, which should aid in the production of larger single crystals with a regular array of protein molecules in the unit cell. Crystal quality depends on the regularity and strength of the protein-protein interactions within the crystal, which can be affected by the detergent, pH and type of buffer present, ionic concentration, temperature and additives [109]. Additives or small molecules are used to decrease the micelle size and to shift the phase diagram to allow nucleation at protein concentrations, buffer types, salt concentrations etc., that were previously unavailable [22].

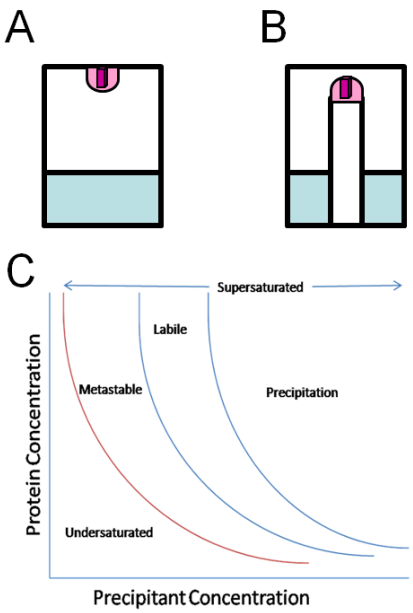


Figure 1-20 A: hanging drop vapour diffusion method. The protein drop is suspended from a cover slip. Below the protein drop is a reservoir containing precipitant such as PEG-400 or ammonium sulphate. B: sitting drop vapour diffusion method. The protein drop sits on a pedestal with reservoir solution surrounding it. In both A & B the protein drop equilibrates with the reservoir solution causing the protein drop to concentrate as the water evaporates. C: phase diagram. The red line is the solubility curve. If the system is in the undersaturated and metastable phases the protein drop will remain clear. As the protein drop concentrates and become supersaturated it enters the labile phase. In this phase stable nucleation occurs. As crystals grow from the nucleation the protein concentration decreases and can reenter the metastable phase where crystal growth can continue.

1.6.3. Automation and robotics in crystallisation

Crystallisation, optimisation and X-ray diffraction experiments have been improved with the advent of high-throughput automation and crystallisation robots. These new techniques have provided the means to test nano-volumes of purified protein using sparse matrix crystallisation screens that test 96 different crystallisation conditions per tray allowing many more conditions to be tested with a smaller sample size. Automation and robotics has increased the reproducibility of crystallisation experiments as the volumes of viscous solutions, such as PEGs and glycerol, are carefully measured by weight. The temperature and humidity of the robot during the experiment set-up is also carefully monitored and maintained. Automated crystal mounting and the addition of Pilatus (Dectris) detectors at synchrotrons allows more crystals to be tested and images collected in a shorter time frame,

whereas microfocus beamlines allow smaller, better ordered crystals to be analysed [99,101,110,111].

1.7. Thesis aims

There were 4 aims of this thesis. Firstly, to isolate, purify and crystallise monodisperse protein detergent complexes of monomeric core complexes from different strains of purple photosynthetic bacteria. If a high-resolution structure of the monomeric core complex can be elucidated the remaining questions about Q/QH_2 exchange and the putative Protein W or PufX. Secondly, to isolate and characterise the hypothesised Protein W from purified core complexes from *Rps. palustris* strain 2.1.6. If Protein W can be identified within the genome it can be deleted, mutated or even swapped with PufX. This would allow comparisons between the function of Protein W and PufX. Thirdly, to analyse how many *pucBA* gene pairs *Rps. acidophila* strain 10050 has, and if any of these genes would translate B800-820 type LH2 complexes. This will be done by sequencing the genome using the next generation sequencing (NGS) technique Illumina in collaboration with Dr Graham Hamilton and Dr Pawel Herzyk. It is important to have the whole genome sequenced as the *pucBA* gene pairs can be arranged throughout the genome and are not necessarily arranged into distinct operons. Finally, as discussed in section 1.4.1 the existence of additional electronic states in Car molecules will be investigated as these states appear to be very important in efficient energy transfer. To do so 2DES (see section 1.5) will be used to excite the S2 and QX peaks of the LH2 complex to see if any additional states can be observed. This work will be a collaboration with Dr Evgeny Ostroumov and Dr Gregory Scholes at the University of Toronto.

2. Materials & Methods

2.1. Cell culture, growth and storage

All cultures were stored both in glycerol stocks (50 % (v/v) glycerol in a 1:1 ratio with culture) at -80°C and in agar stabs at 20°C (Figure 2-1).

Cultures were regularly checked for possible contamination by streaking the cultures on 1.5% C-succinate agar plates (*Rps. acidophila* 10050 & *Rps. palustris*) or 1.5% Pfennig's medium I agar plates (*Alc. vinosum*), grown both phototrophically (BD GasPak™ EZ anaerobe pouch system) and aerobically in the dark to check for contamination of oxygenic organisms. Several single colonies were used to innoculate liquid culture to ensure that no single point mutations were selected.

Rhodopseudomonas (Rps.) palustris strain 2.1.6*, Rps. acidophila* 10050 & *Allochromatium (Alc.) vinosum* strain D were cultured in C-succinate medium [112], Pfennig's medium [113] and Pfennig's medium I [114] respectively (see Appendices). Bacterial cultures stored in glycerol stocks were grown initially in 5 ml of liquid media in airtight sterile universals. Bacterial cultures stored in agar stabs had liquid media filled to the top of the universal to allow cultures to grow. When the culture was fully grown, it was subsequently sub-cultured into larger volumes of liquid media. Large scale growth was carried out in sterilised 10 L pots. The culture was stirred constantly to ensure all the cells received the same average illumination. Liquid cultures were grown for 48 hours under anaerobic conditions exposed to 60 µmol m^{-2} s^{-1} of incandescent light (Figure 2-2). Light intensity was measured using a LI-250A light meter *LI-COR®*. *Rps. palustris* and *Rps. acidophila* 10050 were grown at 30°C, whilst *Alc. vinosum* was grown at 40°C. The cells were harvested for further use as described in section 2.2.

Figure 2-1 Agar stabs of the species of photosynthetic bacteria stored on a laboratory shelf until further use.

Figure 2-2 Liquid cultures growing phototrophically under incandescent light (60 µmol m^{-2} s^{-1}) at 30 °C

2.2. Preparation of chromatophores

The cells were harvested by centrifugation at 5000 g for 30 minutes at 4°C in a Beckman J6B centrifuge, or stored at -80°C until required. The cell pellet was resuspended in 20 mM MES, 100 mM KCl pH 6.8 and centrifuged at 5000 g for 30 minutes at 4°C. The washed cell pellet was finally resuspended and homogenised in 20 mM Tris-HCl pH 8.0 supplemented with 2 mg DNAse + 1 mM $MgCl_2$. Cells were lysed by passage through a French Pressure cell at 950 Kpsi. Membranes were isolated from the lysate by ultracentrifugation at 256,860 g for 60

minutes at 4°C. Pelleted membranes were resuspended in 20 mM Tris-HCl pH 8.0 to an optical density (OD) at 850 nm of 50 cm^{-1} or 25 cm^{-1} as required. Isolated membranes were either used immediately or flash frozen in liquid nitrogen and stored at -20°C for later use.

2.3. Solubilisation of LH complexes

In order to isolate and purify the light harvesting complexes, the membranes had to be solubilised. Different detergents were tested for use in solubilising *Alc. vinosum* as the LH1:RC complexes had previously been shown to be sensitive to treatment with harsh detergents, such as LDAO (personal communication, Dr Nichola Picken). The concentration of the different detergents used to solubilise the membranes was optimised in small scale experiments using 1 ml of membranes at an $OD_{850\ nm}$ = 25 cm^{-1} and increasing amounts of detergent (DDM, DM, DMNG-3) ranging from the CMC concentration to 15 times the CMC concentration. The type and concentration of detergent that appeared to produce stable protein was selected for large scale protein purification (2% DDM or 0.5 % DMNG-3)

Membranes from *Rps. palustris* and *Rps. acidophila* 10050 were adjusted to an $OD_{850\ nm}$ = 50 cm^{-1} and were solubilised with 1% (v/v) LDAO [43]. The optimisation tests showed that the best conditions for optimal solubilisation of *Alc. vinosum* membranes at an $OD_{850\ nm}$ = 25 cm^{-1} were either, 2% (w/v) β-DDM, 2% (w/v) DM or 0.5% (w/v) decylmaltose neopentyl glycol-3 (DMNG-3). All solubilisations were incubated for 30 minutes stirring at RT. Unsolubilised material was removed by centrifugation at 208,000 g for 30 minutes at 4°C.

2.4. Purification of light harvesting complexes

2.4.1. Sucrose density gradients

The solubilised light harvesting complexes were initially fractionated by the use of discontinuous sucrose density gradients (SDG) (Figure 2-3).

The SDGs were first optimised as the different species and different detergents used affected the density of the complexes, and therefore, also the separation of the light harvesting complexes. Optimisation of SDGs was carried out on a small scale (11 ml Beckman tubes) where the volumes of the different concentrations of sucrose were varied in order to obtain separation of the LH2 and LH1:RC. 2 ml of solubilised membranes were layered on top of the sucrose solutions in Beckman ultracentrifuge tubes and centrifuged at 256,860 g for 14 hours in a Beckman Ti70 rotor. Once the SDGs were in the centrifuge a continuous gradient is produced.

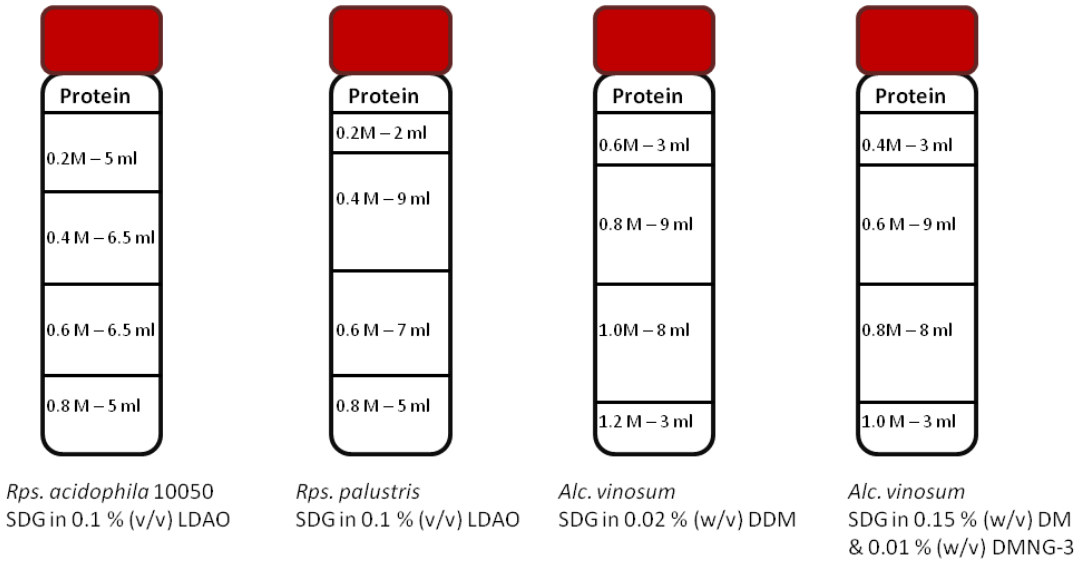


Figure 2-3 Schematic of SDGs showing the optimised conditions for the best separation of the LH2 and LH1:RC complexes.

After centrifugation, typically, 3 bands were observed. The top band corresponds to free pigments, the middle band to the LH2 complexes and the bottom band to the LH1:RC complexes. Both the LH2 and LH1:RC were removed from the gradient using a Pasteur pipette and retained for further purification by anion exchange chromatography.

2.4.2. Anion exchange chromatography

Diethylaminoethyl cellulose (DE52) resin (Whatman) was equilibrated with 3 column volumes of 20 mM Tris-HCl pH 8.0 in a gravity flow

column. Either LH2 or LH1:RC isolated from sucrose density gradients was gently layered on to the column. The bound protein was washed with 20 mM Tris-HCl pH 8.0 containing detergent at its CMC concentration (Buffer A), and eluted with increasing concentrations of NaCl (0 – 100 mM NaCl for LH2 & 0 – 300 mM for LH1:RC) in the same buffer. *Alc. vinosum* strain D core complexes had 10 mM ATP, 20 mM KCl and 50 mM $MgCl_2$ added to the purification buffer to remove the GroEL chaperonin. 0.5 ml fractions were collected and analysed by absorbance spectroscopy (Shimadzu UV-1700 PharmaSpec) for purity and stability of the complexes. The purity of the complexes was estimated by measuring the ratio of the Bchl maxima (approx 880 nm (LH1:RC) or 850 nm (LH2)) to the protein content (280 nm). This ratio is called the optical ratio (OR). Samples with an OR 880:280 >1.8 for LH1:RC (888:280 >1.8 for *Al. vinosum*) or 850:280 >2.5 for LH2 were pooled and concentrated to 1 ml using a Millipore 50,000 molecular weight cut-off (MWCO) centricon (sartorious filter) for further purification by size exclusion chromatography (SEC).

For purification of *Alc. vinosum* strain D core complexes a secondary anion exchange column was required to remove the GroEL contamination by changing the pH of the resin and buffer to a pH value below its theoretical pI value, hence allowing the GroEL to remain bound to the resin whilst the core complex is eluted. The resin was equilibrated with 20 mM K_2HPO_4 pH 5.2. The pool of protein purified by the initial anion exchange chromatography step was buffer exchanged into 20 mM K_2HPO_4 pH 5.2, 0.02% β-DDM (Buffer B) using a Millipore 50,000 MWCO centricon. The protein was washed 5 times with Buffer B. The immobilised protein was eluted from the column using increasing volumes of NaCl (0-100 mM NaCl) in Buffer B. The first band of protein that eluted was retained and further purified by SEC.

2.4.3. Size exclusion chromatography (SEC)

SEC was used to further purify light harvesting complexes previously purified by anion exchange chromatography. An S200 Sepharose column connected to an AKTA prime (GE Healthcare Life Sciences) system with a flow rate of 0.5 ml/min was used. This allowed the light harvesting complexes to be purified on the basis of size. This step also allowed an assessment of how monodisperse the complexes were under these conditions. All SEC was carried out at 4°C.

Absorbance spectroscopy was used to assess the quality of the light harvesting complexes. LH1:RC complexes with an $OR_{880:280\ nm}$ >2.4 ($OR_{888:280\ nm}$ > 2.4 *Al. vinosum)* were pooled and concentrated to 50, 100 or 150 cm^{-1} for crystallization, whilst LH2 complexes with an $OR_{850:280\ nm}$ >3.0 were pooled and stored for 2-dimensional electronic spectroscopy (2DES) (section 2.11). Although the OR is an indicator for purity, it is not a gold standard and purity was also checked by using SDS-PAGE (section 2.4.4)

2.4.4. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to analyse the purity and the polypeptide compositon of the light harvesting complexes. Pre-cast NuPAGE 4–12% Bis-Tris gels (Novex) were used according to the manufacturer's instructions. 1 mM dithiothreitol (DTT) was added to maintain reduced cysteine residues in the RC and prevent the formation of any unwanted disulphide bonds. The complexes were incubated in denaturing buffer and heated at 70°C for 10 minutes. The gels were run in 1x 2-ethanesulfonic acid (MES) buffer pH 7.3 (Novex) at 200 V for 35 minutes using a BioRad power pack. The gels were then stained in either SimplyBlue safe stain (Life Technologies) Sypro Ruby (Invitrogen) or Silver stain (Invitrogen) as per the manufacturer's instructions. The

Novex sharp pre-stained protein standard was used as a protein ladder (220 kDa – 10 kDa).

2.5. Crystallisation of LH1:RC complexes

Crystallisation trials were carried out using the sitting drop vapour diffusion method in 96-well medical research council (MRC) crystallisation plates (Molecular Dimensions). Crystallisation plates were set up using a Cartesian nanovolume crystallisation robot (Figure 2-4) initially using MemGold & MemGold2 sparse matrix crystallisation screens (Molecular Dimensions). Crystals were looked for using a MEIJI EMZ light microscope. Crystals that were three-dimensional with smooth edges and > 20 μm in at least one direction were cryo-protected using well solution supplemented with 40% (v/v) PEG 400 and were cryo cooled in a liquid nitrogen stream (approximately -173.15°C). The crystals were then tested on the University of Glasgow in-house X-ray Diffractometer, Rigaku MicroMax-007 and Mar345 dtb detector. Crystals that presented good diffraction with clear diffraction spots and defined lunes were stored in liquid nitrogen for further testing at Diamond Light Source (DLS).

Optimisations were set up using solutions from Molecular Dimensions where possible (home-made solutions were filter sterilised before use) using a Hamilton robot (Figure 2-5). Promising crystal conditions were optimised by testing the effect of change in pH, buffer concentration, salt concentration and type, and the addition of amphiphiles (Hampton Research and Molecular Dimensions). Purified protein ($OD_{880\ nm}$ = 150 cm^{-1}, 100 cm^{-1} & 50 cm^{-1}) in 20 mM Tris-HCl pH 8.0 + 0.02% (w/v) β-DDM, 0.1% (v/v) LDAO, 0.15% (w/v) DM or 0.01% (w/v) DMNG-3 were mixed in a 1:1 ratio with reservoir solution. The exact conditions tested are outlined in the Appendices. Seeding was also tried as part of the optimisation procedure using the Hampton Seed Beed kit, as described in the manufacturer's instructions.

Figure 2-4 The Cartesian nano-volume robot used to set up crystallization experiments. As the volumes were small the robot contained a humidifier to stop any of the solutions evaporating during the procedure.

Figure 2-5 The Hamilton robot. This robot was used to set up deep-well blocks of both the initial sparse matrix screens and optimizations.

2.6. Crystallisation of the LH2 complex, *Rps. acidophila* 10050.

Purified LH2 complexes were concentrated to an $OD_{850\ nm}$ = 100 cm^{-1} in 20 mM Tris-HCl pH 8.0, 350 mM NaCl, 3.5% (w/v) benzamidine-HCl, 0.1% (v/v) LDAO [43]. Crystallisations were set up in 24-well MRC sitting drop plates. The drop contained a 3:1 ratio of purified LH2 to 4 M K_2HPO_4 to a final volume of 15 µl surrounded by 1 ml of 2.1 M ammonium sulphate pH 9.35 as the reservoir solution. A saturated

sucrose solution was used as a cryoprotectant, however, the solutions had to be equilibrated with 2.1 M ammonium sulphate pH 9.35 over a 48 hour period before use.

2.7. nanoLC-MS-MS

Identification of the peptides in purified core complexes was performed by nanoLC-MS-MS at the University of Sheffield by Dr Philip Jackson. 200 µg of purified core complexes were sonicated in the presence of 120 mM triethylammonium bicarbonate (TEAB) pH 8.5 (Sigma-Aldrich), 1.4% SDS & 7 mM DTT and subsequently incubated at 60°C for 30 minutes. Once the sample had cooled to ambient temperature 9 mM iodoacetamide, an alkylating agent, was added and the sample was further incubated at ambient temperature for 30 minutes. The sample was then diluted with liquid chromatography (LC) grade water to reduce the SDS concentration to 0.1% and supplemented with 4 µg of trypsin. The sample was subsequently incubated overnight at 37°C. The peptides were then dried using a vacuum centrifuge.

The resultant peptides were cleaned using a strong cation-exchange (SCX) separation spin column (C_{18} Spin Tips Sample Prep Kit, Protea). Poros 20 SP (Applied Biosystems) resin was suspended in 50% (v/v) methanol. To generate the spin column a perforated cap was fitted to a standard 1.5 ml reaction tube (Eppendorf) in which a barrier tip, containing 40 µl of the resin suspension, was placed. The resin was packed by centrifugation in a Jencons-PLS spectrafuge at 3.3 g for 3 minutes. The column was washed twice with a cleaning buffer (25% acetonitrile, 0.01 M potassium phosphate pH 3.0 and 1 M KCl) to ensure that any contaminants and solvent were removed. The column was subsequently equilibrated with loading buffer (25% acetonitrile, 0.01 M potassium phosphate pH 3.0) and 0.5 M phosphoric acid pH 2.5-3.0.

The dried peptide sample was resuspended into 200 µl of loading buffer and immobilised on the SCX column. The column was centrifuged at

16.3 g in a microcentrifuge for 10 minutes and washed twice with loading buffer to remove any un-bound material. This step also removed any remaining SDS. The peptides were eluted in 80 µl in elution buffer (25% acetonitrile, 0.01 M potassium phosphate pH 3.0 and 0.5 M KCl) into a Lo-Bind Eppendorf tube and subsequently dried in a vacuum centrifuge.

The dried tryptic fragments were re-dissolved in 7 µl of 0.1% (v/v) triflouroacetic acid (TFA) and 3% (v/v) acetonitrile. An aliquot of 5 µl was analysed by nanoLC-MS-MS using an Ultimate 3000 RSLCnano liquid chromatography system (Dionex, Camberley, UK) with 5 mm x 300 µm trapping and 75 µm x 15 cm analytical PepMap C_{18} reverse-phase columns. A 40-min linear gradient from 94% solvent A (0.1% (v/v) formic acid) to 40% solvent B (0.1% (v/v), 80% (v/v) acetonitrile) at a flow rate of 300 nL/min was used to elute the tryptic peptide fragments. Mass spectra were acquired online using a Maxis UHR-TOF instrument (Bruker Daltonics, Bremen, Germany) operating in line format with automated dependent MS/MS scans. Peak lists for database searching, in the form of Mascot Generic Files (MGFs) were created from the datafiles using a processing script supplied by Bruker. The MGFs were submitted for database searching using Mascot Daemon v. 2.2.0 running with Mascot Server v. 2.2.01 (Matrix Science, London, UK) against the *Rps. palustris* complete proteome database. Search parameters were set to allow for one missed cleavage with S-carbamidomethyl-cysteine as a fixed modification and oxidised methionine as a variable modification. Tolerances were ±0.5 Da for both peptide (MS) and product (MS/MS) ions.

2.8. Engineering knock-out strains

2.8.1. Splice Overlap Extension (SOE) and pk18mobsacB

To engineer knock-out (KO) strains of the candidate Protein W proteins identified by nLC-MS-MS in *Rps. palustris* the splice-overlap extension

protocol was used [115,116]. This required internal and flanking primers to be designed (Table 2-1). Outer primers that lie outwith the KO construct were also designed to allow amplification and sequencing of the KO strain post conjugation. The flanking primers were designed between 300-500 bp upstream (US) and downstream (DS) of the gene of interest, *RPA0067*. The restriction enzyme sites EcoRI and SphI were engineered into these primers to aid the cloning of the KO fragment into the plasmid pk18mobsacB.

Postion	Name	Sequence 5' - 3'
Flanking	0067 US for EcoRI	cata**gaattc**TTTAGCGTCTTGACAGTTCA
Internal	0067 US rev	GTCTTACAACTGCATGTCATTCGCTCG
Internal	0067 DS for	GACATGCAGTTGTAAGACTGCGCATTCTC
Flanking	0067 DS rev Sph1	aaat**gcatgc**AGCAGCTTCCAGTCGAAC
Flanking	0067 outer for	GGTTCAGTGACGCCTATCGGACG
Flanking	0067 outer rev	GGAGACCAGGAACGGCAGTTTG

Table 2-1 Primers used for the generation of RPA0067 KO constructs. Writing in bold blue are the restriction enzyme sequences.

The internal primers were designed to contain a 6 codon overlap composed of 3 codons of the US DNA, and 3 codons of the DS DNA, which extended both into the US and DS sections of DNA. The overlap contained the start and stop codons of *RPA0067* to ensure that all other codons downstream remained in-frame. The generation of the construct is shown schematically in Figure 2-6. The forward flanking primer and reverse internal primer, and reverse flanking and forward internal primers were paired and in two separate PCR reactions amplified the US

and DS fragments of DNA. These two DNA fragments both contained the complementary 6 codon overlap. Using a low and unspecific annealing temperature (37°C) these fragments were annealed and subsequently amplified with a more specific annealing temperature to generate the full length KO construct.

The PCR reactions were completed using Novagen KOD Hot Start DNA polymerase as described in the manufacturer's instructions with the addition of 2 mM $MgSO_4$ and 10% (v/v) dimethyl sulphoxide (DMSO). The sequences were very GC rich and the addition of DMSO was used to relax the secondary structures of the DNA, allowing the primers to anneal and amplify the DNA with a greater efficiency. The reactions contained ~50-125 ng of template DNA: *Rps. palustris* genomic DNA (extracted as outlined in section 2.9) for the US and DS amplifications.

The initial PCR reactions that amplified the US and DS fragments were denatured at 95°C for 15 seconds, annealed at 60°C for 15 seconds and extended at 72°C for 1 minute. These steps were cycled 34 times. The amplified DNA was analysed by gel electrophoresis (section 2.8.3) and was then gel extracted using a Qiagen QiaQuick gel extraction kit. The quantity of DNA measured using a NanoDrop1000 spectrophotometer.

The SOE PCR contained between 50-125 ng of the US/DS DNA fragments. The primers used were the US/DS flanking primers & internal US reverse/internal DS forward primers. No DNA polymerase was added initially. The reactions were incubated in a BioRad PCR thermoCyler (C1000) where they were denatured at 95°C for 10 seconds, un-specifically annealed at 50°C for 20 seconds and extended at 72°C for 60 seconds. These steps were cycled 7 times. This step was to ensure that the overlap DNA sequences annealed to one another prior to amplification. After this step had completed the DNA polymerase was added and the PCR program outlined previously was

completed. This generated the full KO construct which was checked by restriction digest and sequencing by MWGeurofins.

2.8.2. Transformation

The completed KO construct was first ligated in to the cloneJET pJET2.1 Blunt cloning vector (Thermo Scientific) as described in the manufacturer's instructions. 2 µl of the ligation mixture was gently mixed with DH5α competent cells and the mixture was incubated on ice for 15 minutes. To incorporate the plasmid DNA into the competent cells the mixture was incubated at 42°C for 40 seconds. The cells were then placed on ice for 2 minutes for recovery. 250 µl of LB was added to the cells which were subsequently left to grow at 37°C with 180 rpm shaking for 1 hour. pJET 2.1 is a positive selection system as it contains a lethal restriction enzyme gene that is disrupted when a DNA insert is ligated into the multiple cloning site (MCS). The vector also contains an ampicillin (Amp) resistance cassette allowing for the selection of transformations on LB agar plates supplemented with 100 µg ml^{-1} of Amp. The agar plates were incubated at 37°C over night for single colonies to grow.

Colonies of RPA0067KO:pJET2.1 in *E. coli* were grown overnight in LB supplemented with Amp at 37°C with 180 rpm shaking. The plasmid DNA was extracted from these cultures using a Qiagen QiaQuick® miniprep kit. The plasmid DNA was digested with EcoRI and Sph1 restriction enzymes (Fermentas FastDigest) at 37°C for 10 minutes to release the insert DNA from the sub-cloning vector. The digested DNA was analysed by gel electrophoresis and the band at the correct size was gel extracted for ligation into the pk18mobsacB [117] vector that had also been digested with EcoRI and Sph1. The insert and vector were ligated using T4 DNA ligase and incubated at either 4°C overnight or at RT for 4 hours. The ligation mixture of RPA0067KO:pk18mobsacB was then transformed, as described previously, into *E. coli* JM109 cells and grown on LB supplemented with 30 µg ml^{-1} Kanamycin.

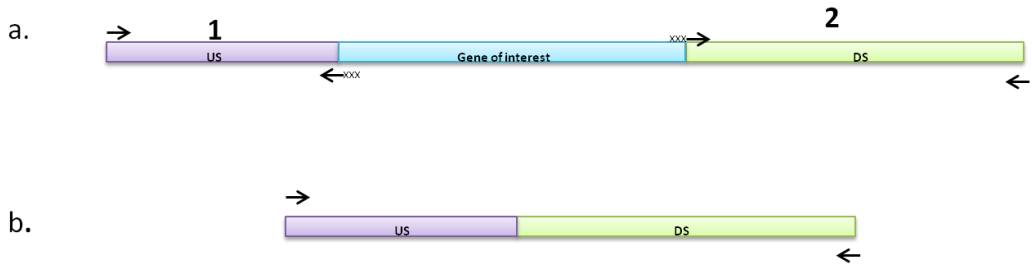


Figure 2-6 Schematic diagram outlining the construction of RPA0067KO DNA fragment. Initially the US and DS fragments are amplified (a1 and a2 respectively) using the corresponding flanking (arrow) and internal (xxxarrow) primers. The US and DS fragments were annealed to each other using the overlap sequence incorporated by the internal primers. The full length construct was amplified using both flanking primers.

Plasmid DNA of RPA0067KO:pk18mobsacB was isolated as described previously. The isolated DNA was sequenced by eurofins MWG operon. In addition to sequencing, fingerprint digests and colony PCR were completed to identify if the insert was the desired KO construct. Once the DNA was sequenced and confirmed to be correct the RPA0067KO:pk18mobsacB construct was transformed into S17-1λpir [118] strain of *E. coli* for conjugation experiments (see section 2.8.4).

2.8.3. Gel electrophoresis

Agarose gel electrophoresis was used to analyse the size of the PCR products and genomic DNA. Gels were made to 1% (w/v) agarose (Roche) in Tris-Acetate-EDTA (TAE) buffer with Syber Safe (Life Technologies). A RunOne (Embi Tec) electrophoresis tank was used at 100 V. A 1 kb DNA ladder (Promega) was run alongside the DNA being tested, as a standard. The DNA bands were visualized using a UVIpro gold gel documentation system (UVItec).

2.8.4. Conjugation

Constructs described in section 2.8 in the pk18mobsacB vector were transformed (as described in section 2.8.2) into S17-1λpir *E. coli* bacterial cells. Overnight liquid cultures of both wild type (WT) *Rps.*

palustris (grown aerobically in the dark, shaking at 30°C) and *E. coli* (S17-1λpir) containing the modified vector were grown. The cultures were analysed by absorbance spectroscopy to ensure they had grown to a sufficient OD ($OD_{650\ nm}$ = > 1.0 cm^{-1} *Rps. palustris* WT and $OD_{600\ nm}$ = > 0.6 cm^{-1} *E. coli*). The concentrations of the cultures were adjusted with sterile Luria broth (LB) to ensure that both donor and recipient strains had the same OD in a final volume of 1 ml. The cultures were centrifuged at 16.3 g in a Jencons-PLS spectrafuge for 1 minute and the pellet was re-suspended in 1 ml of LB.

Ratios of donor:recipient were set up in 1 ml volumes of 1:1, 1:5, 1:10, 1:50, 1:100 and neat *Rps. palustris* as a negative control. The samples were centrifuged at 3.3 g in a Jencons-PLS spectrafuge for 1 minute. The pellet was resuspended in 50 µl of LB and pipetted onto LB agar plates. The samples were not spread as this would minimize contact between the S17-1λpir containing the modified pk18mobsacB vector and the *Rps. palustris* WT. Once the samples had dried on the agar plates, they were incubated at 37 °C for 6 hours.

The dry cells were scraped off the agar plates and re-suspended in 200 µl of C-succinate media without casamino acids ($succs^-$). These samples were then spread onto $succs^-$ agar plates supplemented with 25 µg/ml Kan. *Rps. palustris* was used as a positive and negative control as it should grow when spread on a $succs^-$ agar plate, but should not grow in presence of Kan as it is the presence of pk18mobsacB that confers kan resistance. The plates were grown both chemoheterotrophically and photosynthetically, as previously described in section 2.1. Single colonies from these plates were picked and grown photosynthetically in 8 ml of $succ^-$ supplemented with kan. Once grown, serial dilutions of the cultures were streaked onto $succ^-$ agar supplemented with 10% (w/v) sucrose. pk18mobsacB contains *sacB*, a gene that confers sucrose sensitivity, and so in the presence of sucrose, the vector should fail to thrive, allowing allelic recombination to occur and the RPA0067KO DNA

fragment to became incorporated into the chromosomal DNA of the recipient *Rps. palustris*. The sucrose plates were grown both chemoheterotrophically and photosynthetically and the resultant single colonies were grown in duplicate, one on LB agar supplemented with Kan, the other with no antibiotic present. These plates, called grid plates, were split into 6 x 6 grids to test large numbers of single colonies. The plates were grown chemoheterotrophically.

Colonies that grew only on the grid plates with no antibiotic, were tested by colony PCR using primers designed further US and DS of the flanking primers described in section 2.8 to test whether the gene of interest had been removed. The Taq DNA polymerase JumpStart (Sigma-Aldrich) was used, as described by the manufacturer, using the PCR program used to amplify the full length KO construct previously described in section 2.8.

2.9. Genomic isolation and sequencing

Genomic DNA was extracted as described by Sambrook & Russell [119]. A cell pellet of *Rps. acidophila* 10050 in the log phase of growth was re-suspended in lysis buffer (0.2 M Tris-HCl, 0.5 M NaCl, 0.01 M EDTA, 1% SDS, 1 M sodium acetate) and was incubated for 1 hour at 37°C. The lysis buffer breaks down the cellular membranes allowing the genomic DNA to be released from the cells. 100 µg/ml ProteinaseK was added to break down cellular proteins. The solution was incubated with ProteinaseK at 50°C until the solution became translucent. RNAseA was added to a final concentration of 20 µg/ml and incubated at 37°C for 30 minutes. Phenol:chloroform:IAA (25:24:1) was added in a 1:1 ratio, gently inverting the solutions, forming an emulsion without shearing the genomic DNA. The samples were then centrifuged for 10 minutes in a Jencons-PLS spectrafuge 24D at 16300 g allowing the organic and aqueous phases to reform with the high-molecular weight genomic DNA in the aqueous phase. The resultant aqueous phase was removed carefully and 0.2 volumes of ammonium acetate (10 M) was added with

2 volumes of 100% cold ethanol. The DNA precipitated immediately leaving behind the remaining contaminants. The pellet was removed using a sterile pipette tip and washed twice with 70% ethanol and air dried on the bench. The DNA was resuspended in 30 µl of sterile water. Genomic DNA was further analysed by gel electrophoresis using both uncut and digested DNA. The genomic DNA was digested using Fermentas FastDigest enzymes.

The genome sequencing was carried out by Dr Pawal Herzyk and Dr Graham Hamilton in-house using an Illumina Genome Analyser, a next generation sequencing (NGS) technique (Figure 2-7). Illumina-specific adapters were ligated to the ends of randomly fragmented genomic DNA. This allowed the DNA fragments to become immobilized covalently to the primer rich surface of the flow cell. These primers were specific for the adapters and were used to initiate DNA amplification, which takes place locally. The original template DNA was discarded during a wash step, whilst the newly amplified ssDNA fragments flipped over to hybridize to adjacent primers, initiating bridge amplification. These steps were cycled until a cluster of ssDNA fragments were generated, therefore amplifying the signal. [120]. Sequencing primers were aligned to the adapter sequence of the ssDNA fragments. All four nucleotides and a DNA polymerase were simultaneously added to the flow cell allowing sequencing-by-synthesis to occur [120,121]. Each nucleotide contained a base-specific fluorescent dye on the 3’-OH group which was chemically blocked. Once the appropriate base was incorporated a digital camera within the machine recorded the images in three 100-tile segments. Once the imaging step was complete the system was reset by removing the the fluorescent labels and the chemical blocking group. [120,121].This cycle was repeated 101 times, generating fragments of 101 bp.

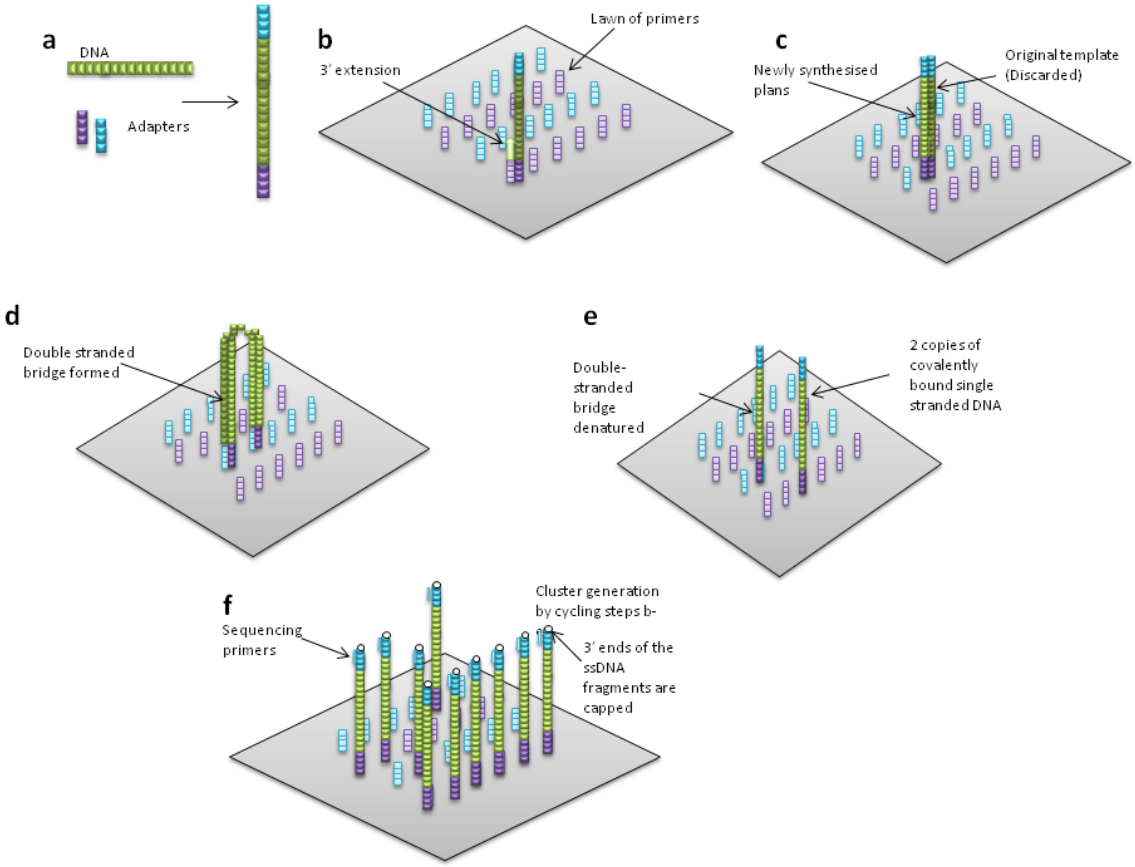


Figure 2-7 A schematic of the Illumina Genome Analyser sequencing technique. a: Randomly fragmented template DNA has adapters ligated to both ends. b: The template DNA covalently binds to the primer rich surface of the flow cell. c: Local amplification of the template is completed, and the template DNA is discarded during a wash step. d: The newly synthesized DNA bends over and binds to a nearby adapter allowing a ds-DNA bridge to be amplified. e:The ds-DNA is denatured forming 2 copies of covalently bound ssDNA fragments. This cycle is repeated until a cluster of DNA fragments are present on the flow cell. This acts to amplify the signal. f: The DNA clusters are capped at the 3' end to avoid any unwanted DNA priming. Sequencing primers that are specific to the adapter sequences are used to amplify the DNA. At this stage all 4 nucleotides (with fluorescent dyes) and DNAsynthase are added at the same time, allowing sequencing by synthesis. Figure adapted from [121] & https://ncifrederick.cancer.gov/atp/cms/wpcontent/uploads/2010/08/CustomerIlluminaTechExplanation.pdf

This method generates millions of short reads that are relatively error-prone with respect to the traditional Sanger method. Illumina technology does, however, reduce costs of sequencing per base. [120]. The high error rate was overcome by trimming the sequence data per fragment from 101 to 85 bp.

Algorithms such as Velvet were used to generate a set of overlapping DNA segments (contig) with as few gaps as possible. Once the genome

was assembled annotation was completed using BLAST (NCBI) and Mauve, multiple genome alignment software [122].

2.10. Isolation and purification of carotenoids

The photosynthetic membranes were extracted with acetone and methanol repeatedly. Between extractions the residual membrane material was pelleted by centrifugation in a Sigma 3K30 SciQuip centrifuge at 478 g for 3 minutes and the supernatants were collected. Once no more pigment was extracted from the membranes the retained supernatants from the acetone washes were diluted with 200 ml petroleum ether and added to a glass separating funnel filled with 500 ml of warm water supplemented with 50 g of NaCl. The solution was mixed before the aqueous and organic phases separated. This allowed any remaining lipids from the carotenoid molecules to be separated as the carotenoid preferentially partitioned in to the petroleum ether organic phase, whilst more polar molecules separated in to the aqueous phase, and the lipids accumulated between the phases. All the extraction methods were carried out under green light.

The isolated carotenoid was then dried using a rotary evaporator (Rotavapor – R, Büchi).

2.10.1. Spheroidene

Extracted carotenoids were further purified using an aluminium oxide column (Aluminium oxide 90 active, neutral Merck). The dried spheroidene was resuspended in 500 µl of petroleum ether and added to a gravity flow aluminium oxide column (60 g aluminium oxide + 600 µl water) equilibrated with petroleum ether. This method of purification utilizes the differences in polarity between pigments. The bound spheroidene was washed with increasing concentrations of diethyl ether from 1% to 15% (v/v) in petroleum ether. The spheroidene eluted at 10% (v/v) diethyl ether. Fractions were collected in 0.5 ml volumes and

were analysed by absorbance spectroscopy to ensure that the spheroidene was pure. It was then dried using rotary evaporation and resuspended in acetone to a concentration of $OD_{451\ nm}$ = 1 mm^{-1} and stored at 4°C in the dark until required.

2.10.2. Rhodopin glucoside

Due to the presence of the glucoside group rhodopin glucoside is polar relative to spheroidene. This meant it could not be purified using the aluminium oxide column as it would remain bound to the column. Rhodopin glucoside in petroleum ether was dried in a glass bulb using a rotary evaporator. The precipitate was washed with methanol to remove any remaining Bchl contaminants. This was repeated until the supernatant was clear. Rhodopin glucoside was then resuspended in acetone and analysed by absorbance spectroscopy (Shimadzu UV-1700 PharmaSpec) to ensure its purity and structural properties. Pure rhodopin glucoside was stored in acetone at $OD_{491\ nm}$ = 1 mm^{-1} at 4 °C in the dark until required.

2.11. 2D electronic spectroscopy

2.11.1. 2DES set up

This work was done in collaboration with the Scholes group at the University of Toronto, Canada, where the experiments took place. The 2 Dimensional electronic spectroscopy (2DES) set up was as described in [123]. A tuneable laser (Ti:Sapphire oscillator) emitted 150 fs duration pulses at 800 nm and seeded a 5 kHz repetition rate regenerative amplifier (Spitfire, Spectra-physics). These pulses seeded a noncollinear parametric amplifier (NOPA) built by the Scholes group. The NOPA converted the 10 mJ pump pulses into 10 mW broad-bandwidth pulses that were compressed to a 13 fs duration. A diffractive optic split the compressed pulse into four beams, k_1, k_2, k_3, and local oscillator (k_{LO}), see Figure 1-17.

The emitted signal was heterodyne detected by the spectrograph (Shamrock, Andor). In heterodyne detection, the beam k_3 and k_{LO} were temporally and spatially overlapped. This allows detection of the phase of the signal, as well as isolating the real (absorptive) signal from the imaginary (dispersive) signal, therefore, amplifying the signal. Figure 1-16 outlines the 2DES set up. Delays were incorporated into k_2 and k_3 and the resultant signals were Fourier transformed to produce the 2 dimensions.

2.11.2. Global analysis

The global analysis was completed by Dr. Evgeny Ostroumov (Ostroumov, Mulvaney *et al.,* 2013) using a home-written Matlab code (MATLAB R2011a, The Mathworks). Global analysis simultaneously analysed large amounts of multi-wavelength data. Using models composed of the instrument response function, the components of the system connected in a kinetic scheme as well as the anisotropy properties of the components, a more powerful spectro-temporal model was produced. Information such as the rate constants and spectra were estimated from the data providing a comprehensive description of the system dynamics [124].

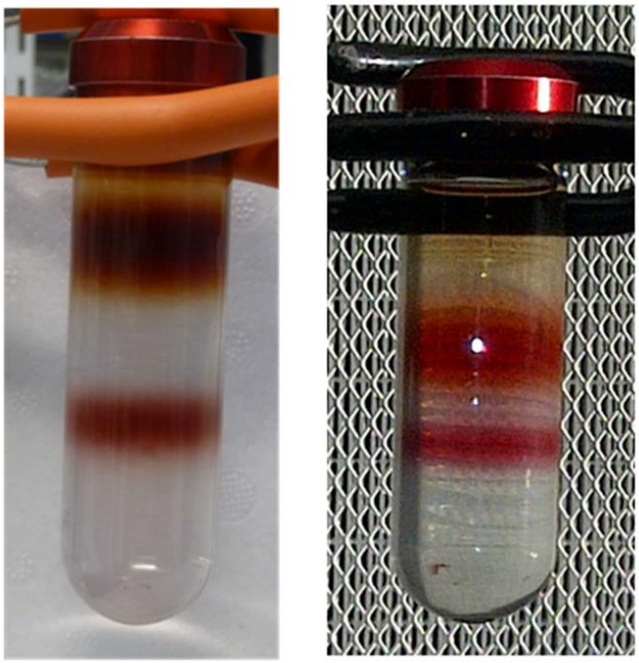
3. Purification, crystallisation and low resolution model of LH1:RC 'core' complex

This chapter describes the purification and crystallisation trials of the LH1:RC 'core' complexes from *Rps. acidophila* 10050, *Rps. palustris* strain 2.1.6 and *Alc. vinosum* strain D. The aim was to obtain 3 dimensional (3D) well-diffracting crystals with which it would be possible to produce a high-resolution structure of a monomeric core complex.

Previously, a 4.8 Å structure of the core complex from *Rps. palustris* was elucidated [18]. At this resolution only overall structural features could be observed. Here, crystallisation trials of the *Rps. palustris* core complex have been attempted to improve upon the existing structural model. Homologues of the core complex of *Rps. palustris* in *Rps. acidophila* and *Alc. vinosum* have also been analysed as they both contain the periplasmic cytochrome-*c* subunit, which increases the soluble portion of the protein complex available to make crystal contacts.

3.1. *Rps. palustris* and *Rps. acidophila*

Purification of these core complexes from *Rps. palustris* and *Rps. acidophila*, completed as outlined in section 2.4 of the Materials & Methods, had been previously optimised by Dr Chris Law and is published in Science [18]. The zwitterionic detergent N,N-dimethyldodecylamine N-oxide (LDAO) was used to solubilise these proteins. Despite this detergent's relatively harsh properties, the core complexes from these species appeared to be stable in LDAO and yield good separation on sucrose density gradients (SDGs). Separation between LH2 and the core complexes on the SDGs was very important as it reduced the amount of LH2 contamination in the core complexes (Figure 3-1). This was aided by downstream purification by anion exchange chromatography and SEC.



Rps. acidophila LDAO

Rps. palustris LDAO

Figure 3-1 SDGs of *Rps. acidophila* and *Rps. palustris* solubilised using the detergent LDAO. The upper 'red' band is the LH2 complex. The denser core complex is the lower band. This was isolated and further purified.

Both *Rps. acidophila* and *Rps. palustris* core complexes bind with a relatively low affinity to the resin, diethylaminoethyl cellulose (DE52), an anion resin equilibrated in 20 mM Tris-HCl pH 8.0. The core complexes were eluted between 100 and 150 mM NaCl in the purification buffer. The samples were analysed by absorbance spectroscopy (Figure 3-2 and Figure 3-3) and SDS-polyacrylamide gel electrophoresis (Figure 3-5), which allows the stability and purity of the protein to be monitored. During purification of the core complex, the absorbance at 280 nm, where aromatic residues of protein absorb, decreases as protein contaminants are reduced. This allows an optical ratio (OR) of the LH1 maxima (880 nm) to protein (280 nm) content to increase, which can be used as a guide to the purity of the samples.

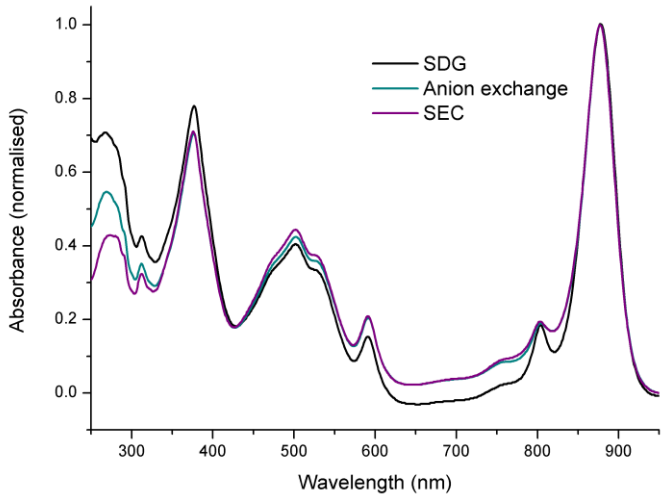


Figure 3-2 Absorbance spectra of *Rps. palustris* core complexes at different stages of purification. The black line shows the spectrum of the core complexes isolated from the SDGs shows that a small fraction of LH2 may be present as the peak at 800 nm is more prominent with respect to the teal and purple lines, which correspond to purified core complexes. There is also a much greater absorbance at 280 nm relative to the other spectra. The spectra of core complexes post anion exchange is shown in cyan, and the post SEC is shown in purple. It is clear from these spectra that during the purification process the 280 and 800 nm peaks were reduced. The spectra were all normalised to an absorbance of 1 A.U at 877 nm, where the LH1 absorbance is maximal.

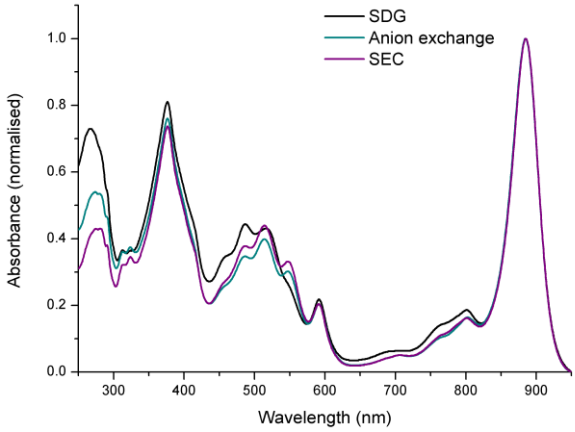


Figure 3-3 Absorbance spectra of *Rps. acidophila* core complexes at different stages of purification. The black line shows the spectrum of the core complexes isolated from the SDGs shows that a small fraction of LH2 may be present as the peak at 800 nm is more prominent with respect to the teal and purple lines, which correspond to purified core complexes. There is also a much greater absorbance at 280 nm relative to the other spectra. The spectra of core complexes post anion exchange is shown in cyan, and the post SEC is shown in purple. It is clear from these spectra that during the purification process the 280 and 800 nm peaks were reduced. The spectra were all normalised to an absorbance of 1 A.U at 885 nm, where the LH1 absorbance is maximal. It is clear that the carotenoid in the core complex sample from the SDG has differences in carotenoid, as the 400 to 550 nm spectra do not overlap. It is possible that the isolated core band was contaminated with free carotenoid or LH2 complexes, generating a hybrid of 2 different carotenoid spectra.

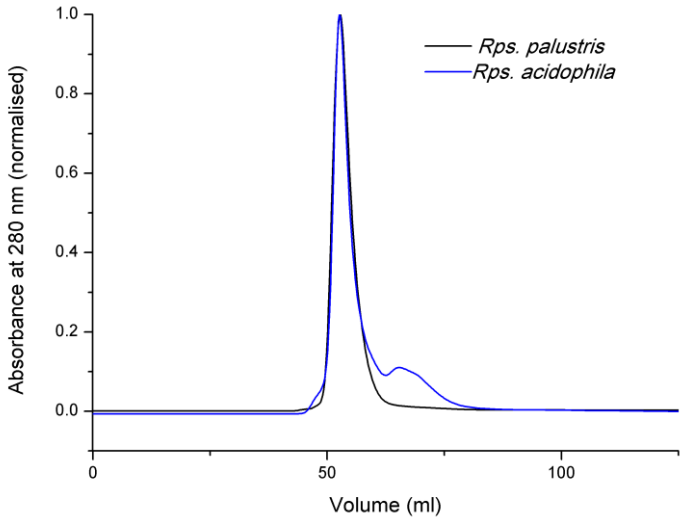


Figure 3-4 Chromatogram from S200 SEC column for the core complexes from *Rps. palustris* and *Rps. acidophila*. *Rps. acidophila*, despite having a prominent elution peak, contains minor humps before and after the main portion of protein elutes. This suggests that the protein from this species is not completely stable, or there are contaminants present. Only the peak fractions were pooled to avoid any contamination with protein that was partially denatured. The chromatogram from *Rps. palustris* core complex is monodisperse.

After anion exchange chromatography, the fractions that had an OR of ≥1.8 were pooled and further purified using an S200 SEC column. The core complexe eluted 12 ml after the void volume. The chromatogram of the core complex from *Rps. palustris* post SEC showed a symmetrical and monodisperse peak, which suggests that the complexes are uniform within the detergent protein complex (Figure 3-4). The core complex from *Rps. acidophila*, however, appears to be less-stable in LDAO as the chromatogram peak presented an additional, less intense, broad peak after the main peak had eluted. This suggests that these core complexes were less stable and undergoing denaturation. When *Rps. acidophil*a membranes were solubilised with 2% (w/v) β-DDM and the extracted core complexes purified as above, a monodisperse peak was obtained.

Purified core complexes were analysed using SDS-polyacrylamide gel electrophoresis (Figure 3-5), which showed that the complexes appeared to be rather pure, as no contaminants were visible when

stained in SimplyBlue stain. *Rps. acidophila* core complexes presented an additional band at 38 kDa, which agreed with the results from chapter 5 section 5.2, suggesting that this species contains an additional subunit, cytochrome-*c*.

Crystallisation trials of the core complexes from *Rps. palustris* began by repeating the crystallisation conditions published by Roszak and colleagues, 2003. Purified core complexes in LDAO were detergent exchanged into 1% (w/v) sucrose monocholate.

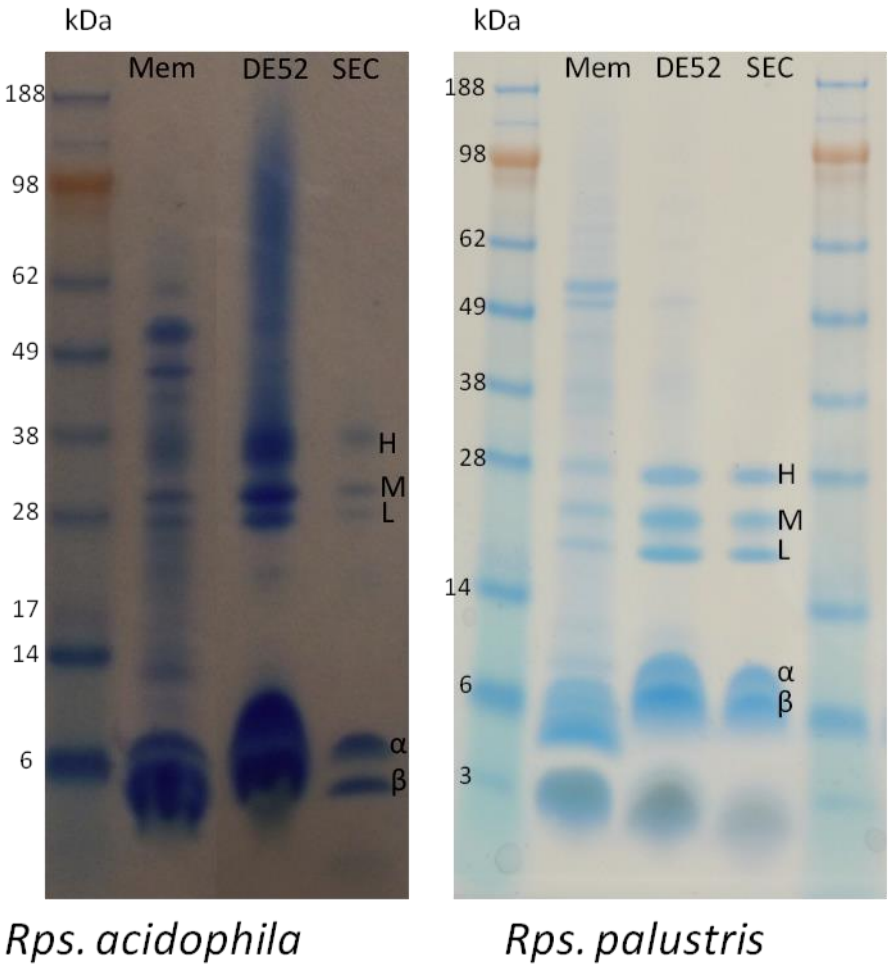


Figure 3-5 SDS-polyacrylamide gels, stained with SimplyBlue stain, showing the purification of *Rps. acidophila* and *Rps. palustris* core complexes, respectively. Each image shows the membrane fraction, protein post anion exchange and finally, post SEC. *Rps. acidophila* core complexes contain an additional band at 38 kDa, which is the cytochrome-*c* subunit. In both gels only the subunits of the RC (L, M, H and C) and LH1 (α and β) were observed.

The concentrated protein was diluted to a final OD at 878 nm of 55 cm^{-1} in the purification buffer supplemented with 20 mM $MgCl_2$ and 16 to

25% (v/v) MME-PEG2000. 5 µl of the protein was mixed with 5 µl of the reservoir solution in the sitting drop. The reservoir solution contained 2% (w/v) spermidine, 100 mM Tris-HCl pH 8.5 and 16% (w/v) MME-PEG2000. The detergent, sucrose monocholate, is no longer manufactured although a small amount had been located for these experiments. Similar conditions with a different precipitant and additive present were also attempted. These experiments, again, used purified core complexes at an OD at 878 nm of 55cm^{-1}, were mixed in a 1:1 ratio with the reservoir solution, containing: 18% (v/v) Jeffamine ED2001 as the precipitant and 0.6 to 1.1% (w/v) benzamidine-HCl as an additional additive. In these experiments sucrose monocholate was used as an additive at concentrations ranging from 0.2 to 0.8% (w/v). The use of this detergent as an additive instead of detergent exchanging was due to the limited supply of sucrose monocholate. It had also previously been shown by Dr Chris Law that the use of sucrose monocholate as an additive was also a successful crystallisation strategy. It had previously been observed that benzamidine-HCl improved the crystallisation of LH2 complexes from *Rps. acidophila* and the LH1:RC from *Blc. viridis* (personal communication with Dr Alastair Gardiner). During these experiments, only granular precipitation and microcrystals were obtained. However, it was not possible to further optimise or test these conditions due to the limited supply of sucrose monocholate.

In 2008, the sparse matrix crystallisation screen MemGold (Molecular Dimensions), designed for use with α-helical membrane proteins, was manufactured [106]. This allowed a new approach for the crystallisation trials to be undertaken. Other screens, shown in Table 3-1, were also tested to ensure that there was no bias, however, these conditions have not produced promising crystals and in some cases appeared to denature the protein as a colour change from pink to black was observed, and the protein precipitated. Crystals grew in MemGold within

2 weeks. The crystals obtained were small (approximately 20 μm long) and flat, as shown in Figure 3-7.

Screen	Manufacturer
MemGold	Molecular Dimensions
MemGold 2	Molecular Dimensions
MemSys/Start	Molecular Dimensions
Cryo 1 & 2	Emerald Biosystems
Morpheus	Molecular Dimesnions
PegIon	Emerald Biosystems
Hampton 1 & 2	Hampton

Table 3-1 Sparse matrix crystallisation screens used in crystal trials of *Rps. palustris* and *Rps. acidophila* core complexes.

Crystallisation experiments were completed at 16 and 20°C. During the optimisation of initial hit conditions, protein purified in LDAO as described above was detergent exchanged into the detergents β-decylmaltoside (β-DM) and decylmaltoseneopentylglucoside (DMNG-3). It was thought that, although the protein appeared stable during solubilisation and purification, prolonged exposure to LDAO may be detrimental to crystal formation as the protein may denature. The milder detergents, β-DM and DMNG-3, were used as they have larger detergent micelles and could improve protein stability. As *Rps. palustris* core complexes do not contain the periplasmic cytochrome-*c* subunit, there are less polar exposed subunits present to make crystal contacts. This would mean that the use of β-DDM, which has a larger micelle size

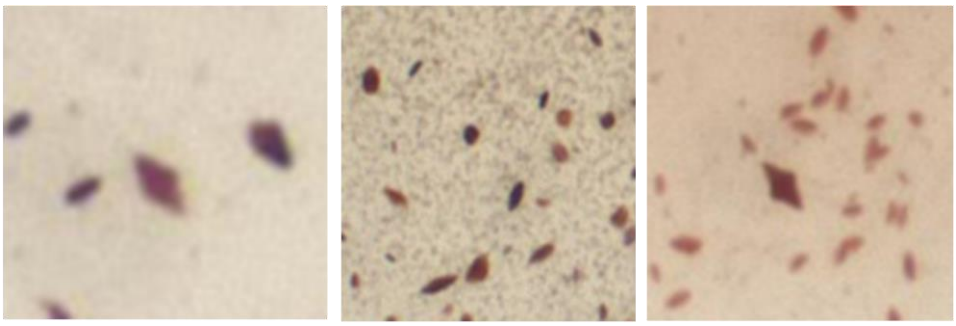
than β-DM and DMNG-3, could mask the polar exposed subunits which would render crystallisation less favourable.

As the core complexes from *Rps. acidophila* are inherently less stable and did not produce a monodisperse peak, the membranes were solubilised and purified in β-DDM. As previously mentioned, the core complexes in solution with β-DDM produced a monodisperse peak. As *Rps. acidophila* strain 10050 contains the additional cytochrome-*c* subunit, it was thought that the polar subunits would protrude outwith the β-DDM micelle, unlike the core complexes from *Rps. palustris.*

Crystallisation trials of *Rps. acidophila* core complexes were approached in the same way as those for the *Rps. palustris* core complexes, and the initial experiments that were set up are listed in Table 3-1. Again, it was observed that the only screen to produce crystals was MemGold. Figure 3-6 shows examples of crystals obtained from purified core complexes of *Rps. acidophila*.

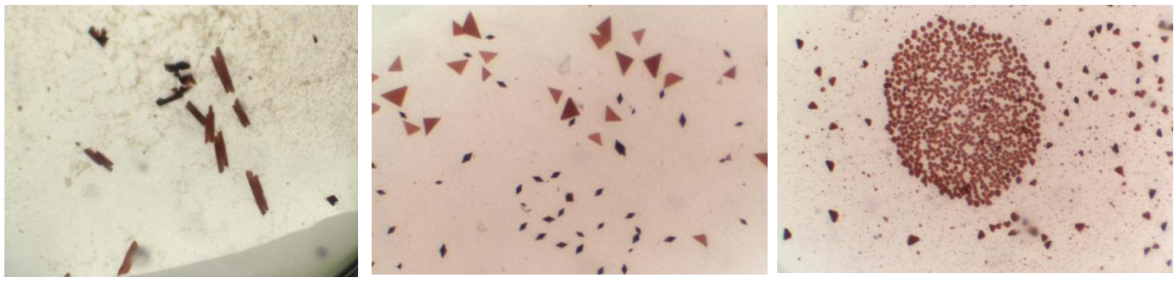
The initial crystal hit conditions of both *Rps. palustris* and *Rps. acidophila* core complexes were optimised in 24-well crystallisation plates. In total, 672 and 984 different crystallisation conditions were analysed for *Rps. acidophila* and *Rps. palustris* core complexes, respectively. Optimisations were set up with the original crystallisation condition placed in the middle of the crystallisation plate; whilst the surrounding wells varied in precipitant concentration or precipitant type (PEG-400 19 to 42 %, MME-PEG 2000 16 TO 25 % and Jeffamine ED 2001, 16 - 20 %), salt concentration/type (0.05 to 0.1 M NaCl and 0.01 to 0.02 $MgCl_2$), buffer (Tris-HCl, HEPES and MES) pH (6.5 to 8.5), detergent (LDAO, β-DDM, β-DM, NM, DMNG-3) and buffer concentration (0.02 to 0.1 M), with only one variable at a time. All optimisations were set up in duplicate to allow different protein concentrations to be tested and were incubated at 16°C. Unfortunately, these experiments did not produce crystals with improved size, quality, or diffraction properties. In some cases crystals did not grow in the conditions that were replicates

of the initial hit conditions from the sparse matrix crystallisation screen. To test that the original crystal hits were reproducible additional MemGold trays were set up. Although crystals grew in the secondary MemGold trays, they did not grow in the same conditions. This suggests batch-to-batch differences in the purified protein. It is also possible that minor changes in the crystallisation robot, such as the temperature and humidity control were not the same for each experiment or that the incubator where the plates were stored did not maintain the temperature correctly, or for periods of time the incubator was open resulting in variable temperatures. It is also possible that the small micelle size of the detergent LDAO allowed important annular lipids surrounding the complexes to be stripped during SEC. Certain lipids could be required for the function and even structural integrity of the complexes, and their removal during purification could be detrimental. Another, quite likely, explanation for the optimisations lack of success is the volume difference of the experiments from sparse matrix crystallisation screens in 96-well plates to the 24-well plates. In the initial screens, the volume of the protein drop was a 1:1 mixture of protein:reservoir solution with a final volume of 1 µl, with 500 µl of reservoir solution. The optimisations, however, had 1 ml of reservoir solution surrounding the sitting drop plinth containing 1 to 15 µl of protein mixed with 1 to 1.5 µl of reservoir solution. It has also been suggested that for reproducible crystallisation trials all chemicals should be purchased from the company who manufactured the sparse matrix crystallisation screen that was used initially. This is due to the unknown preparation of the solutions, which means they cannot be reproduced in the laboratory.



7.7 % PEG4000	24 % PEG 400	6.6 % PEG 4000
0.1 M NaCl	0.05 M Magnesium acetate	0.1 M NaCl
0.02 M TRIS pH 7.0	0.05 M sodium acetate pH 5.4	0.02 M KMES pH 6.7

Figure 3-6 Examples of crystals of *Rps. acidophila* core complexes. Images taken automatically by a Rhombix crystal imager, therefore as the crystals were ≤ 20 µm the images are not as sharp as they could be.



0.5 M lithium sulphate 10 % (w/v) PEG 8000 No buffer, no pH	0.1 M NaCl 0.325 M Na 0.1 M TRIS pH 8.0 33 % (v/v) PEG 400	0.1 M NaCl 0.12 M TRIS pH 9.4 20 % (v/v) PEG 400

Figure 3-7 Examples of crystals of *Rps. palustris* core complexes. Images taken automatically by a Rhombix crystal imager. Crystals with a forked rod morphology measured 0.1 mm in length.

Although a range of crystal morphologies was obtained most of the crystals were small, flat, and diamond shaped. A general feature for the crystallisation of these complexes, as observed previously by Dr Chris Law, is that small molecular weight PEGs and the presence of a salt are favoured. The crystals obtained from these experiments were tested in-house and at Diamond Light Source (DLS), Oxford, UK. These crystals did not diffract well. The best diffraction obtained showed high anisotropy with a high level of mosaic spread and only a few spots visible to 20 Å from a crystal grown in 0.1 M NaCl, 20 mM Tris-HCl pH 8.5 and 33 % (v/v) PEG-400 of core complexes from *Rps. palustris*. Figure 3-8 shows an example of the diffraction that was observed from

the *Rps. palustris* core complex crystals when exposed to synchrotron radiation. As optimisations and reproduction of the initial sparse matrix crystal screens were irreproducible, and the detergent which had previously shown promise, sucrose monocholate, is no longer manufactured, any further attempts to crystallise *Rps. palustris* core complexes were abandoned. The crystals of the core complexes from *Rps. acidophila* did not diffract and so these crystallisation trials were also abandoned. As previously mentioned, it is possible that essential annular lipids were removed. It would be interesting to purify the core complexes from *Rps. palustris* in lipidic bicelles to see if this improves the crystallisability and reproducibility of both the purification and crystallisation experiments. It would also be prudent to attempt the optimisations using the same micro-volumes as the initial screens, as it is possible that the change in volumes within the experiment changed the rates of evaporation and therefore shifted the phase diagram. If this did occur, then the optimisations at variable volumes would mean that the crystallisation experiments were conducted in an unknown area of the phase diagram, which makes logical optimization impossible.

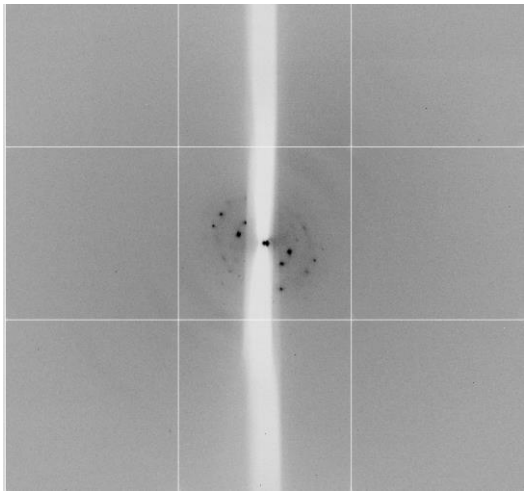


Figure 3-8 Example of diffraction from an *Rps. palustris* crystal shot at DLS I04. The diffraction extended to 20 Å therefore these crystallisation conditions were considered interesting and an optimisation was attempted. Optimisation trials however, yielded no further improvement.

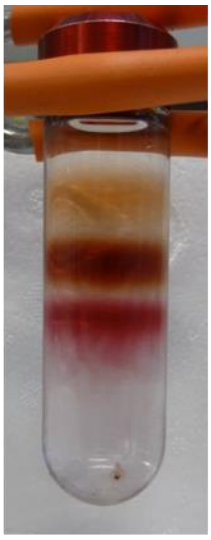
3.2. *Alc. vinosum* purification optimisation

A common technique when trying to obtain high-resolution diffracting crystals of a target protein is to test homologues from different species. *Alc. vinosum*, a purple sulphur bacterium, was available in our laboratory. It is known from the recently sequenced genome that it contains the periplasmic cytochrome-*c* subunit desirable for making crystal contacts, but also contains 3 different *pufBA* gene pairs that are all transcribed on the same mRNA strand, [69]. This generated the possibility of heterogeneity in the LH1 complex. Despite the risk of heterogeneity, it was speculated that the LH1 complex would still be composed of very similar transmembrane αβ-heterodimer subunits, and so although it could be a possible problem for crystallisation, it may not affect the crystal formation or diffraction and hence should be attempted.

Previous work on this species within the laboratory had revealed that the core complex becomes denatured when the membranes are solubilised with LDAO, as only LH2 complexes were present in the absorption spectra and SDGs of the solubilised membranes. As outlined by Simon Newstead [106] β-DDM is generally used for membrane proteins that are less stable and is a popular choice as the first detergent to try when assessing a new membrane protein crystal target [101,104,106]. Due to the change in detergent and working with a species less well characterised than *Rps. acidophila* and *Rps. palustris,* it was necessary for the SDGs to be optimised. Therefore the concentrations and volumes of sucrose solutions were varied in order to find the concentrations with which the LH2 and core complexes equilibrated with enough separation between the two bands. The SDGs of *Alc. vinosum* were optimised on a small scale testing different volumes and different concentrations of sucrose ranging from 1.8 to 0.2 M in the purification buffer. During the optimisation process it became clear that the separation between the LH2 and the core complex was

less defined with respect to *Rps. acidophila* and *Rps. palustris*. It is speculated that this is due to β-DDM being less effective at solubilising the membranes, which can be seen in the absorption spectra of the core complex from the SDGs (Figure 3-11). This can mean that in the solubilised protein solution, the LH2 and core complexes are still bound to each other in the detergent micelle. This can make purification difficult and reduces the yield of pure protein per gram of cell pellet. It could also be possible that this is due to the difference in detergents being used, as β-DDM forms a much larger micelle than LDAO, thereby affecting the densities of the two complexes, or simply a difference in the lipids present. The efficiency of solubilisation and separation of the LH2 and the core complexes could be improved by solubilising the membranes with a mixed micelle, potentially by adding a small amount of LDAO. Personal communication with Dr Elizabeth Carpenter from the structural genomics consortium, Oxford, suggested that the efficiency of solubilisation was not the most important feature to optimise as it was far more important to obtain good quality protein, regardless of the yield. The SDGs were optimised by loading lower concentrations of solubilised membranes on to the SDGs to maximise the separation. It was observed that when 2 ml of $OD_{850\ nm}$ = 50 cm^{-1} was loaded, no visible separation between the LH2 and core complexes was observed. This was optimised by adding 1 ml of $OD_{850\ nm}$ = 25 cm^{-1} to the SDGs (Figure 3-9).

From the SDGs 3 bands were visible. The identities of the bands were assigned by measuring the absorbance of the bands within a range of 250-950 nm (Figure 3-10). This showed that the upper orange pigmented band contained free carotenoid molecules, some Bchl molecules and protein. The middle, second band, was the LH2 complex and the bottom band was the core complex.



Alc. Vinosum - DDM

Figure 3-9 *Alc. vinosum* SDGs for membranes solubilised in the detergent DDM. These SDGs had to be optimised to ensure there was enough separation between the LH2 (middle band) and core complexes (bottom band). The top band corresponds to free carotenoids.

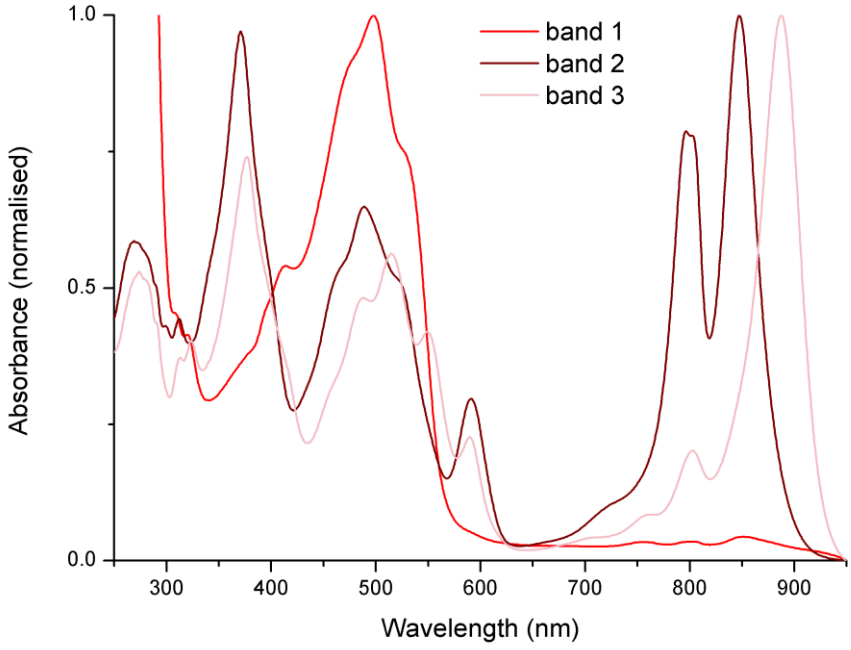


Figure 3-10 Absorbance spectra of SDG bands from solubilised *Alc. vinosum* membranes. The spectra were normalised to show a maximal absorbance at 1, except for band 1, which was normalised to 1 at the main carotenoid peak at 500 nm. The absorbance spectra allowed the identification of band 1 as free carotenoid and protein contaminants (red), band 2 as the LH2 complex (brown), which, in *Alc. vinsoum* has a double 800 nm peak, and the 3rd band as the core (pink). The optimised SDGs allowed good quality core complexes to be isolated with little LH2 contamination.

Once the core complex had been successfully isolated from the SDGs, it was purified as described in section 2.4 of the Materials & Methods, by anion exchange and SEC (Figure 3-11). The chromatogram from the SEC step showed symmetrical monodisperse peaks (Figure 3-12).

The purity and quality of the purified core complexes were initially analysed by absorbance spectroscopy. As previously discussed samples with a ratio of maxima at 888 nm (LH1) to 280 nm (protein) of 2.4 or higher were pooled and concentrated to an $OD_{888\ nm}$ of 50, 100 and 150 cm^{-1} for crystallisation trials.

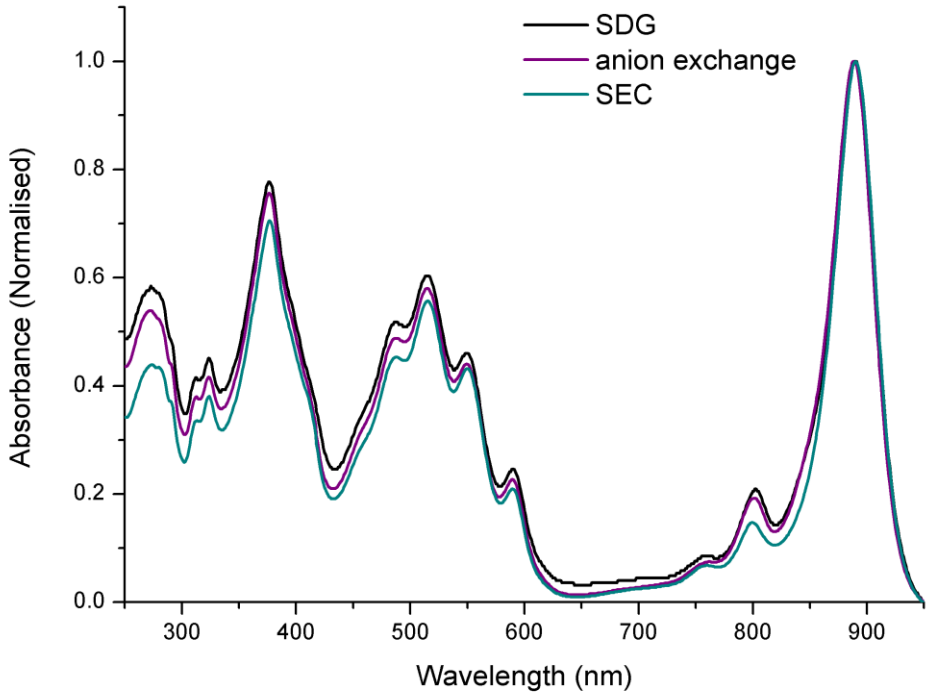


Figure 3-11 Absorbance spectra of *Alc. vinosum* core complexes at different stages during purification. In black the spectrum of the core complexes isolated from SDGs shows a relatively high absorbance at 280 nm, from protein contaminants. The peak at 800 nm is higher due to a minor LH2 contamination. This also accounts for a slightly broader 888 nm LH1 maximum. During purification by anion exchange (purple) and SEC (cyan) the 280 nm and 800 nm maximas decrease and the 888 nm peak becomes sharper as any LH2 contamination had been reduced.

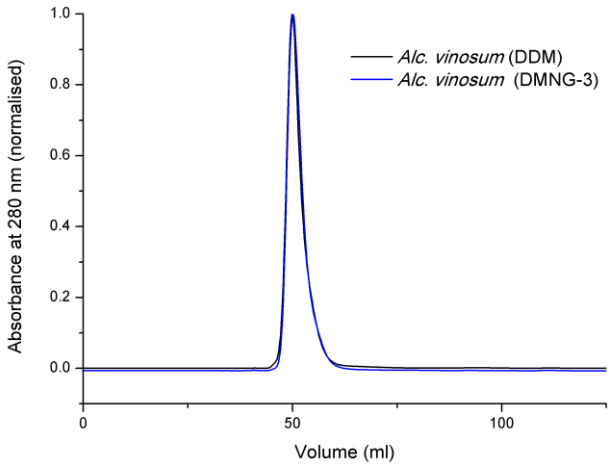


Figure 3-12 Chromatogram of core complexes post SEC on an S200 column. Both chromatograms are from *Alc. vinosum* core complexes but they are in different detergents, DDM and DMNG-3. Both chromatograms show sharp monodisperse peaks indicative of folded and uniform protein.

Initially, *Alc. vinosum* was grown using thiosulphate as the electron donor in the growth medium. Crystals were obtained in one week from the MemGold sparse matrix crystallisation screen and appeared to grow from phase separation (Figure 3-13 A). The condition that yielded these initial crystals was: 24% (v/v) PEG-400, 0.05 M sodium acetate, 0.05 M magnesium acetate, pH 5.4. The crystals were large rods, of 400 µm length, that grew into each other forming 'X' shape morphology. During harvesting of the crystals, the 'X' split into 4 individual rods approximately 175 µm long and 85 µm wide. The crystals were cryoprotected using a solution made up of the reservoir solution supplemented with 40% (v/v) PEG-400 as the cryoprotectant and 0.02% (w/v) β-DDM to ensure the membrane protein crystal remained crystalline. The crystals were exposed to X-rays for 1 hour on the home source and the resulting image is shown in Figure 3-13 B. The diffraction shown in this image is very promising as the diffraction spots are sharp and form lunes indicating low crystal mosacity. However, the diffraction limit was relatively low (15 Å) and the diffraction appeared to perform slightly better in one direction, an indication of anisotropy. Anisotropy is common in membrane protein crystallography as the detergent micelles can limit protein contacts in one direction. Ultimately

this results in processed data sets that have to be trimmed from the highest resolution spots obtained to an overall lower resolution. These crystals were subsequently tested at DLS on the beamline I04-1 and data was collected to a resolution of 8 Å. The data collected were not the best quality due to high levels of anisotropy which reduced the resolution of the scaled data.

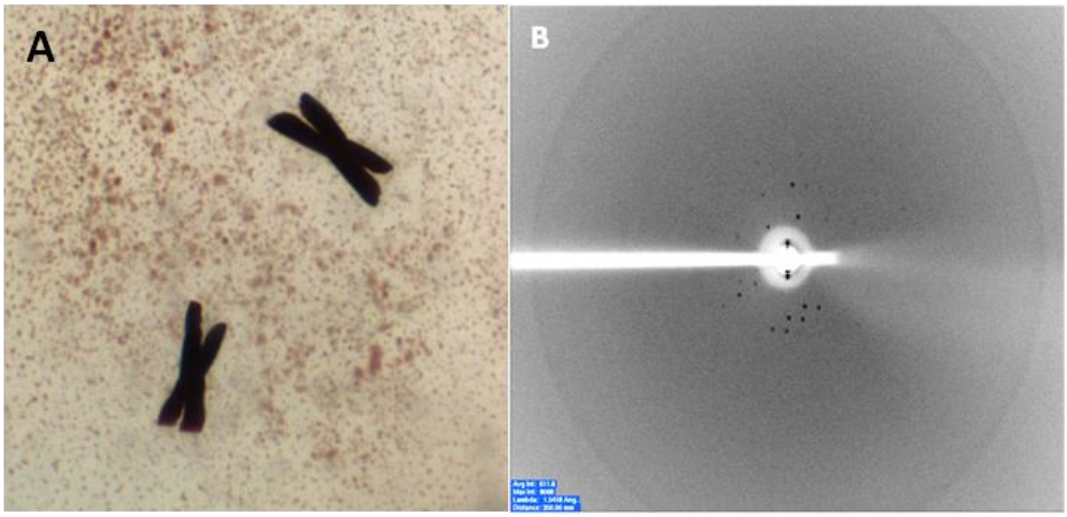


Figure 3-13 A: *Alc. vinosum* core crystals in 0.02 % (w/v) β-DDM from a 96-well sitting drop vapour diffusion tray. The reservoir contained 24 % (v/v) PEG-400, 0.05 M sodium acetate, 0.05 M magnesium acetate, pH 5.4. Crystals were 350 µm long and had a width of 170 µm. B: Diffraction pattern from the home beam when exposed for 1 hour with 1° oscillations.

Although these crystals were very promising, subsequent attempts to both optimise and directly repeat the experiments failed. It was already known that *Alc. vinosum* contained a *puf* operon with 3 copies of *pufBA* gene pairs that can all be transcribed, yet it was unknown which copies were expressed under these growth conditions. To identify the LH1 subunits present, purified protein was analysed by nLC-MS-MS at the University of Sheffield by Dr Philip Jackson. These experiments identified the presence of 2 α-apoproteins that correspond to the genes $pufA_1$ and $pufA_3$, whilst only one β-apoprotein was identified, $pufB_3$. It was therefore hypothesised that the heterogeneity in the LH1 complex did affect the crystallisation and was responsible for the irreproducibility in the core complex crystallisation from *Alc. vinosum* grown on thiosulphate media.

As the initial crystal and its diffraction was so promising it was decided to test different growth conditions to see if slight variations could alter the LH1 complex composition and therefore crystallisability. Sodium sulphide, an alternative electron donor, was used in the growth media instead of sodium thiosulphate. The light intensity and temperature of the growth conditions were maintained, so the only variable was the electron donor. The core complexes displayed the same spectral features in the absorbance spectra and were purified as described previously. Crystal trials were first attempted using MemGold, due to the previous success. These initial screens did not yield crystals in the same conditions as the previous thiosulphate grown *Alc. vinosum* core complexes, however they did produce small crystals with sharp edges in different conditions (Figure 3-14). The crystals showed a different morphology, appeared less 3-dimensional and the largest crystals were 200 µm long, with many more approximately 100 µm in length in the same well. The crystallisation conditions contained 21% (v/v) PEG-400, 0.1 M Tris-HCl pH 8.0, 0.1 M NaCl and 0.325 M sodium acetate. The cryoprotectant solution was prepared using the same conditions as the reservoir solution, with 40% (v/v) PEG-400 and 0.02% (w/v) β-DDM.

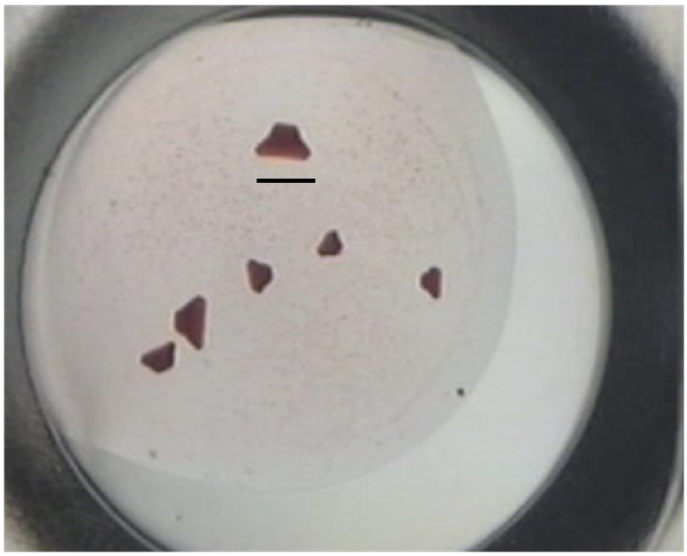


Figure 3-14 Initial crystal hit from *Alc. vinosum* core complex grown on sodium sulphide as a hydrogen donor. Line indicates 200 µm. Crystallisation conditions were 21 % (v/v) PEG-400, 0.1 M Tris-HCl pH 8.0, 0.1 M NaCl and 0.325 M sodium acetate.

These crystals were tested for diffraction at DLS-I04 in collaboration with Dr Aleksander Roszak. A data set was collected with orthorhombic symmetry belonging to space group $P2_12_12_1$ and was processed by XIA2 [125] to a resolution of 6.8 Å with spots extending to ~5 Å. The diffraction showed clean spots located in lunes, which indicated low levels of mosaic spread (estimated to be 0.15 by iMosfilm), but also showed signs of anisotropy. To correct for this latter effect the data were re-processed to 5 Å , using Mosfilm [126], and then corrected for anisotropy using the Diffraction Anisotropy Server [127]. This allowed the data to be extended to 5.5, 5.7 & 6.3 Å in directions a, b & c, respectively. Alternatively, the data were scaled by SCALA [128] to an overall resolution of 5.9 Å.

Optimisations of this crystallisation condition were carried out both on small scale 96-well MRC plates and large scale 24-well MRC plates using solutions purchased from Molecular Dimensions to ensure reproducibility. If solutions weren't available >99% grade chemicals were dissolved in ultrapure water and filter sterilised before use. In total, 6192 conditions were tested. Initial optimisations were generated in nanovolumes in 96-well trays using the robot Cartesian HoneyBee 8+1. This allowed many conditions to be tested in a short amount of time with less purified protein required (see Appendices). These experiments were reproducible and the rates of evaporation were comparable between the initial screens and optimisations as the same volumes were being tested. Optimisations tested the effect of changing the concentration of the precipitant (13 to 40%), type of precipitant (PEG-300, PEG-400, PEG-4000, PEG-2000 MME, Jeffamine ED2001), concentration of both first and second salt (0.01 to 0.5 M) salt type (NaCl, $MgCl_2$, $LiSO_4$, $ZnCl_2$, KCl), buffer concentration (0.01 to 0.1 M), buffer type (Tris-HCl, MES, ammonium acetate, sodium citrate) as well as buffer pH (pH 6 to 10), protein concentration ($OD_{888\ nm}$ = 25, 50, 100, 150 cm^{-1}) and temperature of experiments (4, 16 and 20 °C). The main improvements, shown in Figure 3-15, were due to changes in

concentration of PEG-400. Ideally, single crystals should be obtained in a well. It was shown that by changing the PEG-400 concentration from 27% to 24% (v/v) the size of the crystals increased, whilst the number of crystals decreased.

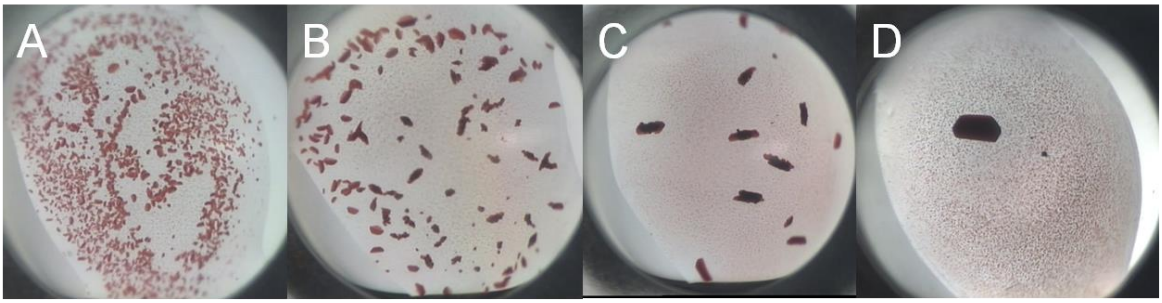


Figure 3-15 Optimising PEG-400 concentration in the reservoir. A: 27 % PEG-400 - many small crystals formed. B: 26 % PEG-400 - fewer, larger crystals formed. C: 25 % PEG-400 – still fewer and larger crystals formed, approximately 100 µm long. D: 24 % PEG-400 - a single crystal measuring 500 µm in length.

The Hampton additive screen and experiments using a seed solution were also attempted, however, neither significantly improved the crystal quality or diffraction resolution. Experiments were duplicated and incubated at 4, 16 and 20°C to analyse how changes in temperature affected the crystallisation. It was observed that at 4 and 20°C protein precipitated and denatured. Crystals were only obtained when crystallisation experiments were incubated at 16°C.

Optimisations were set up in 24-well trays in an attempt to grow larger crystals. The optimisations were based around the best conditions from previous small scale optimisations, 23-30% (v/v) PEG-400, 0.05-0.1 M NaCl, 0.1-0.325 M NaOAc, 0.1 M Tris-HCl pH 8.0, with either 0.5 or 1 ml of reservoir solution surrounding the sitting drop. 1, 2 and 4 µl of purified core complexes at $OD_{888\ nm}$ = 50, 100 and 150 cm^{-1} were added to the sitting drop and mixed in 1:1 and 2:1 ratios of protein to reservoir solution. Crystals were obtained with larger volumes; however, it became clear that higher concentrations of precipitant were required for crystal formation. These crystals measured approximately 300 µm in length, only marginally more than some of the crystals obtained from

the small scale optimisations. Although the crystals diffracted and produced the same space group ($P2_12_12_1$) data, the diffraction and resolution were only around 8 Å. This was not as good as the smaller crystals obtained from the small scale optimisations. A possible explanation for this could be that as the crystals grow more protein is recruited into the growing crystal resulting in less order within the crystal. Another way the size of the crystal could affect the diffraction quality is that the larger the crystal, the more difficult it is to perform cryoprotection. A caveat of producing larger crystals is that cryo-cooling can cause damage that may affect the diffraction properties, as the liquid within the crystal and the protein can freeze at different rates; this is more commonly called the 'lollypop effect'. The need for a higher concentration of precipitant for crystal formation would have affected the rate of evaporation. If this occurs too fast, crystals form too quickly resulting in a less regular array of protein. As more protein was required for these experiments and the crystals were not as promising, it was decided that scaling up the crystallisations was not the best way to proceed given the time restrictions of the project.

In order to improve the resolution of the data that had been collected, smaller carbon chain detergents, β-DM and DMNG-3, were used to solubilise and purify the core complexes. β-DM was chosen as it is from the same family of detergents as β-DDM, and so the protein should behave in the same way in this detergent, and it was already known to be stable, monodisperse and crystallisable in a maltoside detergent. DMNG-3 was chosen as it has been suggested to improve the stability of the dimeric core complex from *Rba. capsulatus* [129]. The SDGs that had been optimised for β-DDM were sufficient to obtain separation between the LH2 and core complexes and so required no further optimisation. The complexes were purified as previously described. Absorbance spectroscopy and SEC chromatograms suggested that the complexes were stable and monodisperse in the detergent solutions. Crystallisation trials were set up using MemGold and MemGold2 and

core complexes at an $OD_{888\ nm}$ = 50, 100 and 150 cm^{-1}. No crystals were obtained from protein purified in β-DM. Crystallisation experiments of core complexes purified in DMNG-3 did produce crystals, however, they were very large 2D-sheets and did not diffract. These crystals were optimised on a small scale, but the crystals could not be improved and showed no signs of diffraction using either the home-source or synchrotron radiation. It was also observed that the protein in the crystallisation experiments were still pink after 2 months of incubation at 16°C, suggesting that the complexes were largely intact. It is possible that by decreasing the micelle size, the complexes are more likely to have key lipids removed during the purification process, similar to the core complexes from *Rps. palustris* and *Rps. acidophila* in LDAO, however, if this were the case it would be expected that the protein would precipitate or denature. It would be, none-the-less interesting to purify these complexes in lipidic bicelles, or adding lipids as an additive to the crystallisation experiments. As these core complexes contain the large cytrochrome-*c* subunit which provides exposed soluble protein, it is possible that it extends beyond the β-DDM micelle. If this is the case, then there may be little point in decreasing the micelle size, especially if it compromises the structural integrity of the protein.

As these growth conditions provided protein that reproducibly crystallised, some of the purified core complexes were analysed by nLC-MS-MS by the proteomics facility at the University of Dundee. It was observed that the LH1 apoproteins present corresponded to $pufBA_3$ only. This suggests that the reproducibility here is possibly due to homogeneity in the LH1 ring.

As the crystals could not be further improved from the optimisations, the purification of the core complex was revisited. As the purity had been largely analysed by absorbance spectroscopy, samples were then analysed by SDS polyacrylamide gel electrophoresis (Figure 3-16). Analysis of the denatured subunits of the core complex from *Alc.*

vinosum showed that despite the protein appearing to be pure, there were 3 clear bands between 38 kDa and 49 kDa. These bands were excised separately and analysed by nLC-MS-MS at the University of Dundee Proteomics Facility to identify the proteins present.

It was shown that all these bands were fragments of the *Alc. vinosum* cytochrome-*c* subunit from the RC. It was concluded that the core complex from this species was proteolytically sensitive. To address this problem, broad-spectrum protease inhibitor cocktail tablets were used during the harvesting and breaking of cells as well as throughout protein purification. EDTA was also added to inhibit any metalloproteases (5 mM). The

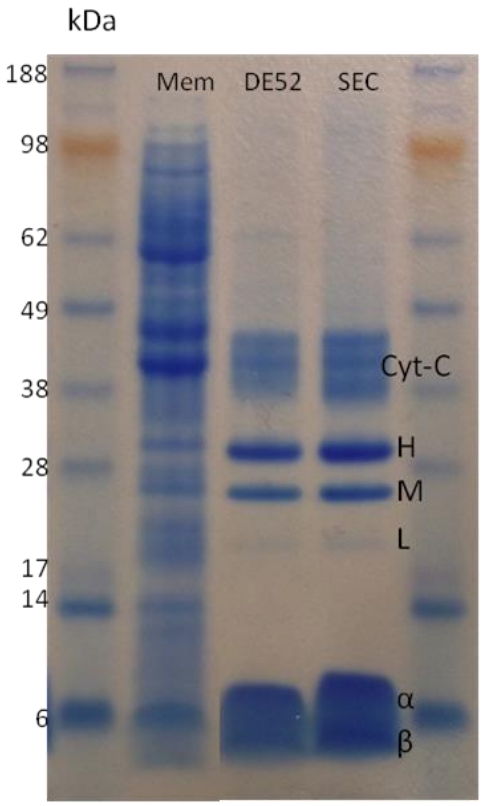


Figure 3-16 SDS polyacrylamide gels comparing denatured subunits of core complexes from *Alc. vinosum* purification experiments. This gel compares protein samples from the membranes, core complexes post anion exchange and post SEC purification. Despite appearing pure when stained with SimplyBlue stain, the cytochrome-*c* subunit had been proteolysed.

purification procedure was repeated with no changes and the protein analysed for proteolysis and purity. SDS polyacrylamide gels showed

that the presence of EDTA and broad-spectrum protease inhibitors had eliminated the proteolysis of the cytochrome-*c* subunit, however, contaminants within the protein became more apparent (Figure 3-17). This contaminant, at approximately 60 kDa, was gel excised and analysed by nLC-MS-MS at Dundee University. It was shown to be the *Alc. vinosum* GroEL chaperonin.

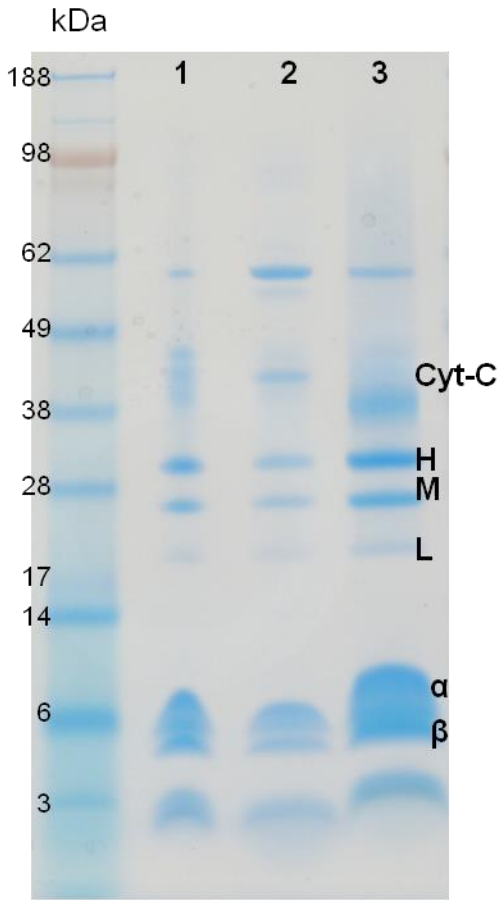


Figure 3-17 SDS polyacrylamide gels comparing denatured subunits of core complexes from *Alc. vinosum* purification experiments. This gel shows a comparison of protein purified with no protease inhibitors (1), with broad spectrum protease inhibitors (2), and resuspended crystals (3). This gel shows that protein purification with protease inhibitors eliminates proteolysis of the cytochrome-*c* subunit. In all these cases, however, a new band appeared at approximately 60 kDa. nLC-MS-MS identified this band as the chaperone GroEL/ES.

GroEL contamination has been previously well-documented in recombinant protein work [130,131]. It has been shown that an effective method to remove GroEL is to add ATP, $MgCl_2$ and KCl to the purification buffer. The addition of ATP causes the chaperone to release any protein that may be bound. This procedure was carried out on

protein immobilised on an anion exchange column and during the SEC. Neither experiments managed to remove the GroEL contamination. Comparison of the theoretical isoelectric points (pI) between the core complex and GroEL, using an online service; ExPASy (http://web.expasy.org/compute_pi/), suggested that the core complex has a pI of 6.16, whilst GroEL has a pI of 5.01. This suggests, that if GroEL was not bound to the core complex, then both protein complexes would be negatively charged at pH 8.0 and bind the anion exchange resin. It was also speculated that the GroEL heptamer and the core complex in a β-DDM micelle would be similar sizes (approximately 400 kDa). As only one sharp peak was observed from the SEC chromatogram, this seems very likely. Methods to remove this contaminant would be to apply a His tag to the core complex. This would allow the core complex to be selectively purified over the GroEL. However, this was not possible to complete with the time restraints of this project. As there is a small difference in theoretical pI values of GroEL and the core complex, it was theoretically possible to separate them by equilibrating an anion exchange column at pH 5.8, and buffer exchanging the protein from 20 mM Tris-HCl pH 8.0 to K_2HPO_4 pH 5.8, resulting in negatively charged GroEL, which will bind to the resin, and a positively charged core complex that shouldn't bind to the resin. Although this should remove the contaminant, the effect of low pH on the core complex is unknown. However, this was attempted to prove that the contamination could be removed. This reverse phase anion exchange step was implemented after the initial anion exchange step with ATP treatment. The core complex did bind to the column, although only weakly, and was eluted with 50 mM NaCl in the purification buffer. It is probable that the theoretical pI value calculated by SwissProt is less reliable for photosynthetic complexes as the cofactors could affect the final pI. The eluant was collected in 0.5 ml fractions and swiftly detergent exchanged in to the Tris-HCl purification buffer at pH 8.0 and further purified by SEC. The protein was compared with samples from membranes and post reverse phase DE52 on SDS polyacrylamide gel

shown in Figure 3-18. This gel shows that there is no observable presence of GroEL in the samples post the second anion exchange step (Figure 3-18).

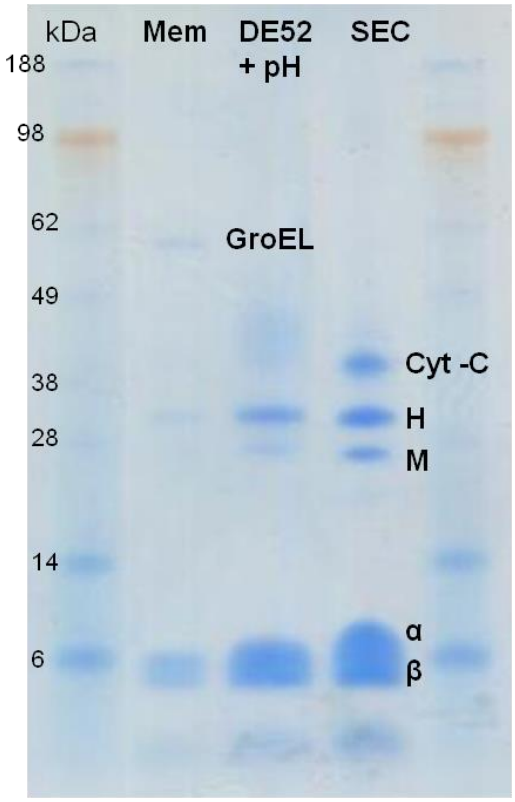


Figure 3-18 SDS polyacrylamide gels comparing denatured subunits of core complexes from *Alc. vinosum* purification experiments. Gel comparing samples from the membranes, core complexes post anion exchange and post SEC purification with an optimised purification protocol to remove GroEL/ES. GroEL/ES is clear in the membrane fractions, however, is no longer visible in the purified protein fractions.

Crystallisation trials of core complexes purified using this additional purification step have been attempted and diffracting crystals have been obtained. Crystals grew within 2 weeks, however, were more fragile and prone to breaking. The morphologies of the crystals also changed from trapezoid to clusters of pyramidal crystals. Crystals were tested on the home-source and, although some did not perform well, most of the crystals performed in a similar way to the crystals obtained from previous experiments with GroEL present. These crystals have not been tested at the synchrotron yet. Once a data set has been collected using synchrotron radiation the data will be processed and the crystals can be further optimised.

3.3. Crystallographic model of *Alc. vinosum* core complex

Crystals from the optimized conditions, described previously and grown from purified core complex in DDM at a OD_{888nm} = 100 cm^{-1} with a reservoir solution containing 26% (v/v) PEG-400, 0.1 M Tris-HCl pH 8.0, 0.1 M NaCl, 0.325 M sodium acetate. The crystals were grown in 96-well crystallization plates at 16 °C. The data set was collected and processed to 5.15 Å, with some diffraction spots observed up to 4.5 Å. The data was collected at DLS on the beam-line I02. The data showed that the crystal was not twinned. Table 3-2 outlines the data collection statistics. Data collection, processing, phasing and refinement were done in collaboration with Dr Aleksander Roszak.

Beamline	I02
Wavelength (Å)	0.97950
Resolution	5.15 (23.04)
Multiplicity	4.3 (3.1)
Space group	$P2_12_12_1$
Unit cell parameters	a = 133.61 b = 184.55 c = 190.16
Completeness (%)	99.6 (86.1)
R_{merge}	0.060 (0.540)
$\langle I/\sigma(I) \rangle$	12.6 (2.3)

Table 3-2 Data collection statistics for the core complex. Values in parentheses are for the outer shell.

Figure 3-19 shows an example of the recorded diffraction images. The quality of the diffraction pattern and diffraction spots (for example good separation between the observed lunes) indicates that the crystal had good order and a low mosaic spread. Some diffraction spots were observed to a resolution limit of 4.5 Å, but the data at the processing stage were trimmed to 5.15 Å to fulfill the condition of $<I/\sigma(I)> \geq 2$. Molecular replacement by the program Phaser [132] was then carried out using models generated from the previously solved *Thermochromatium (Tch.) tepidum* RC [133-135] *Blc. viridis* RC [21], the core complex from *Rps. palustris* from the *Rps. palustris* 4.8 Å RC-LH1 model [18]. In order to generate the *Tch. tepidum* and *Blc. viridis* core complex models the LH1 complex coordinates from *Rps. palustris* were used. The RCs from *Tch. tepidum* and *Blc. viridis* were tested as they also contain the periplasmic cytochrome-*c* subunit.

A range of MR strategies were tested and are outlined in Table 3-3. The solutions produce figures of merit which allows the experimenter to assess the quality of the solution. The statistics given for MR solutions are Z-scores for rotational function (RFZ), translational function (TFZ) and log-likelihood gain (LLG). The LLG value estimates how well the model fits the data. According to the literature a TFZ value > 8 suggests that the structure is solved [132]. The figures of merit for the different MRs attempted showed that the RC with the best fit (LLG) RFZ, TFZ scores and $R_{factors}$ was the *Tch. tepidum* RC. Using this solution (*Tch. tepidum* core model) it was clear that, although the statistics suggested the solution was correct, the electron density maps did not fit well to the LH1 helices. This suggested that the LH1 ellipse from *Rps. palustris* was not the best fit for the data collected from the *Alc. vinosum* core complex. Comparison of the different MR solutions shows that using just the RC of *Tch. tepidum* as a search model the statistics were good, but not as good as the *Tch. tepidum* core model in terms of the model fit; however, the RFZ score improved.

MR model	RFZ	TFZ	LLG	R_{factor}
Blc. viridis core	6.6	16.3	568	56.7
Tch. tepidum core	7.3	19	830	55.4
Rps. palustris core	4.2	12.2	260	61.2
Tch. tepidum core - LH1β	8.3	19.1	792	55.9
Tch. tepidum core - LH1α	8	19.2	763	55.8
Tch. tepidum RC	9.2	19.3	731	55.9
JB solution*	10.8	18.7	5342	43.7
JB solution* - LH1	12.6	20.4	3237	47.7
JB solution* - LH1α	11.6	18.1	4055	45.9
JB solution* - LH1β	11.9	22.2	4268	45.7

Table 3-3 Statistics from MR for the 5.15A data using various models.*Tch. tepidum* core model, built using the *Tch. tepidum* RC and the LH1 model from the *Rps. palustris* core complex structure. JB solution* MR model used the solution from Jellybody REFMAC refinement using the *Tch. tepidum* core model with Protein W in the LH1 ellipse moved to be parallel to the adjacent α-apoproteins.

To analyse what affect the different apoproteins of the LH1 complex had on the solution, new search models of the *Tch. tepidum* core were generated without either the LH1-β or the LH1-α apoproteins. These results showed an improvement in the LLG score when LH1-β was not present, with respect to the LH1-α apoprotein. As the β-apoprotein ring lies on the outside of the complex, it is possible that they are in a less rigid environment and hence are less well defined in the electron density maps. If there are any structural differences between the *Alc.*

vinosum and *Rps. palustris* LH1 complexes, it is probable that they are due to species variation. As there are no other structures available for any LH1 complex no other models can be generated. In order to identify what is going on, the crystals need to be further optimised in order to collect higher resolution data. Experimental phasing could then be completed by changing the wavelength of the X-ray beam to hit the Fe-edge. This is possible as the cytochrome-*c* subunits contain 4 Fe atoms and the RC contains 1. It would also be possible to do experimental phasing by soaking heavy metals, such as mercury, into the crystals. This would allow the amplitudes of the electron density to be matched, solving the phase problem.

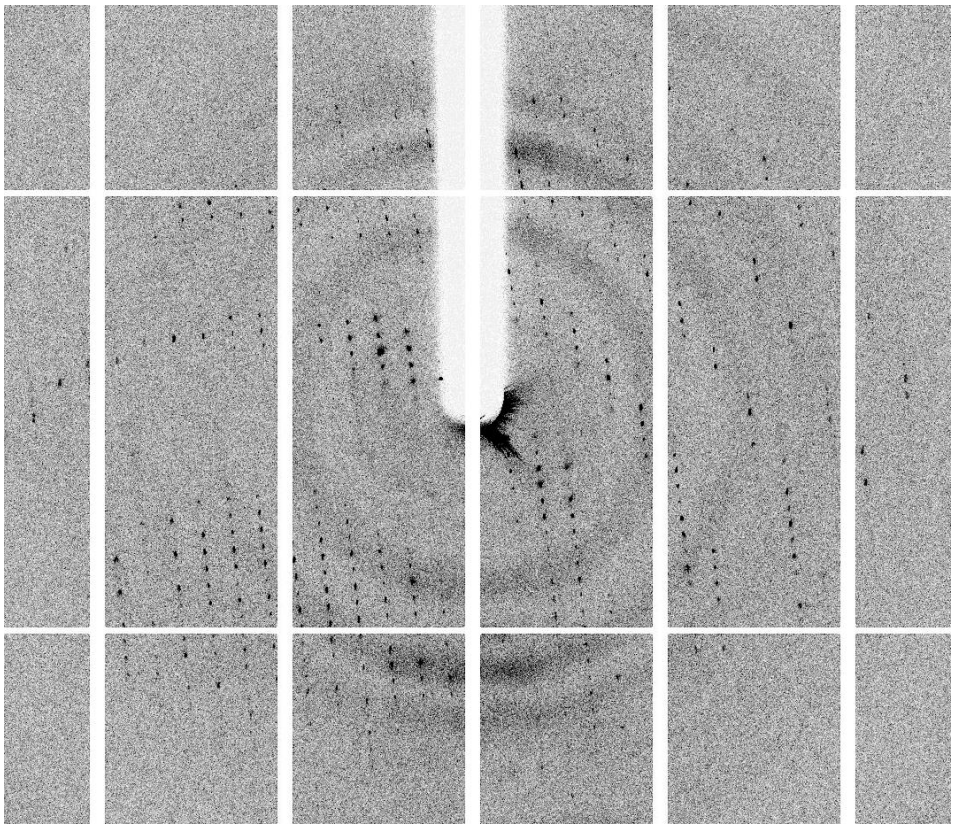


Figure 3-19 Diffraction pattern from an *Alc. vinosum* core complex crystal. The diffraction shows nice, well separated lunes with diffraction spots extending to 4.5 Å. Processing of all images produced a data set trimmed to 5.15 Å, where the average $\langle I/\sigma(I) \rangle$ is 2.3 for the data at the highest resolution shell .

Previously, the *Rps. palustris* core complex structure was phased using the RC from *Rba. sphaeroides* RC coordinates as a MR model. At 4.8 Å this solution produced a good fit to the RC model, and the electron

density map showed clear density for the LH1 complex. This density was good enough to fit the basic secondary structures and observe the disjointed Protein W. With the *Alc. vinosum* data at 5.15 Å, phasing the data with just the *Tch. tepidum* RC did not yield the same result and density for the LH1 complex was not observed.

Jelly-body (JB) REFMAC5 [136] was used to refine the MR solution. This technique improves the fit of the model to the electron density maps by increasing the signal, which aids in further model building. JB refinement is ideal for low-resolution data as it allows the atoms, or secondary structures to move if there is evidence from the electron density or external restraints to suggest that they could be better placed. The refinement process introduces restraints of geometry and B-factors (thermal factors). Refinement using the *Tch. tepidum* core model with Protein W was not successful due to the orientation of Protein W. When Protein W was moved in line with the α-apoproteins from the LH1 complex refinement was successful and the $R\text{-}_{factors}$ were improved after 300 iterations from 55.5 to 43.7%.

The electron density map of the solution was improved post refinement (Figure 3-20 A and B) and additional density became apparent opposite the helix in the position of Protein W (Figure 3-20 B) which would suggest the presence of a β-apoprotein. The view through the membrane (Figure 3-20 A) shows that the density around some of the LH1 helices is incomplete. This, again, suggests that the model used for MR was not ideal for the *Alc. vinosum* core complex. It is probable that the LH1 from *Alc. vinsoum* is a closed LH1 complex with no PufX orthologue.

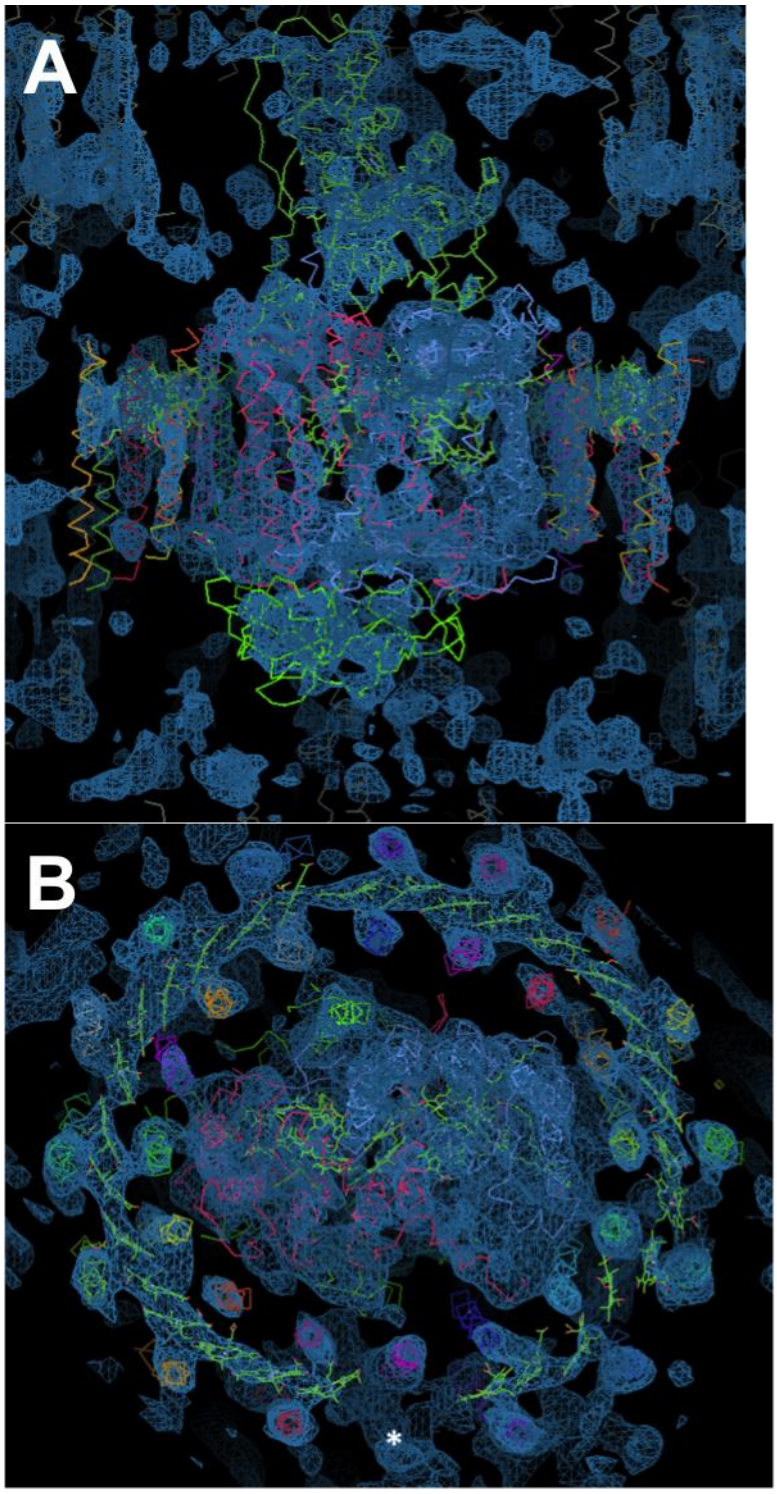


Figure 3-20 Model and electron density map (σ level = 1.38) of *Alc. vinosum* core complex after jelly-body REFMAC5 refinement. The larger periplasmic cytochrome-*c* subunit is at the top of the image. This image shows that although statistically the solution appears correct for the low resolution structure, the electron density does not fit the model perfectly.For the refinement to work protein W had to be moved. This can be observed in B (*). Protein W was moved more in line with the rest of the LH1 helices. Once this was completed additional density opposite Protein W appeared. It is possible that this density opposite could correspond to an additional helix.

The best solution generated by MR (*Tch. tepidum* core model) showed 1 molecule per asymmetric unit with a solvent content of 65.37%.

Although this percentage is high, it is quite normal for a membrane protein. On the other hand, it has been shown that a good technique for improving crystal contacts and ultimately resolution of diffraction has been crystal dehydration. This technique has not yet been tried for these crystals, but would be an additional technique to be tested in the future. Figure 3-21 shows the crystal packing for the *Alc. vinosum* core complex based on the obtained and refined MR solution.

The packing of all the *Alc. vinosum* core complex crystals shows extensive protein contacts between the cytochrome-*c*, H - and LH1 subunits. This is an improvement on the crystal contacts observed for the *Rps. palustris* core complex shown in Figure 3-21. However, the crystal packing is very similar as the molecules are stacked on top of each other, head (periplasmic side; cytochrome-c subunit) to toe (cytoplasmic side; H-subunit).

Personal communication with Professor Wang-Otomo, Ibaraki University, during the 2013 PS16 conference in St. Louis, MO, USA indicated that after 15 years characterising the core complex from *Tch. tepidum* they had finally optimised the crystallisation conditions and obtained crystals that diffract to approximately 3 Å. Although the refinement process is still ongoing, they have shown that the LH1 complex is composed of 16 homogeneous αβ-heterodimer subunits that form a closed LH1 complex. This corresponds to the additional density that was observed from the *Alc. vinosum* core complex data; however, at the resolution presented in this thesis, it is not conclusive.

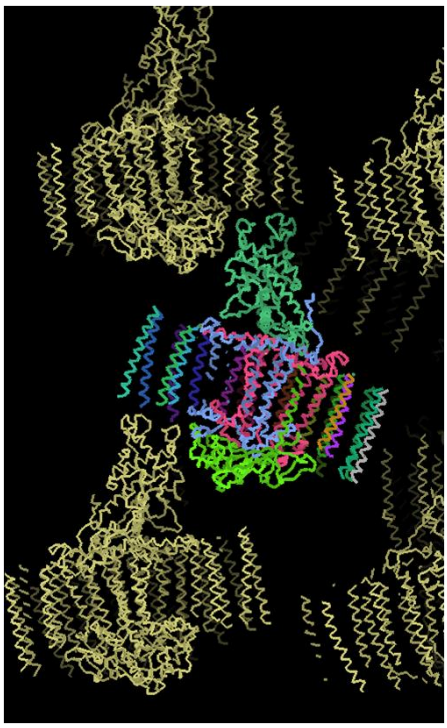


Figure 3-21 Crystal packing for *Alc. vinosum* core complex in the orthorhombic space group $P2_12_12_1$. The molecules are stacked on top of each other making protein contacts between the cytochrome-*c* of one molecule and the H-subunit and the LH1 helices of an adjacent molecule.

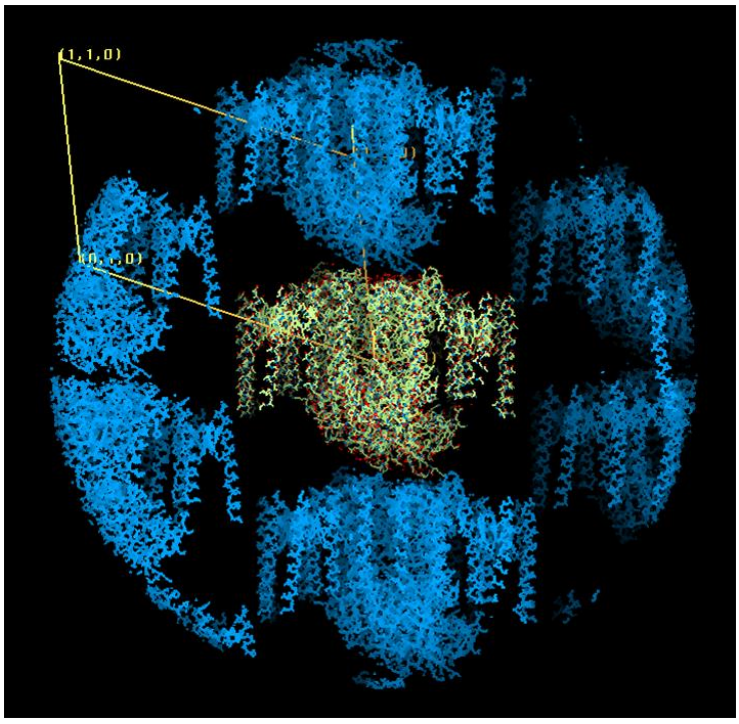


Figure 3-22 Crystal packing for *Rps. palustris* core complexes. Molecules are stacked on top of one another head to toe. Protein contacts are limited and made between the polar sections of the L and M subunits of RC to the H-subunit of an adjacent molecule. It is also possible that there are some crystal contacts between parts of the LH1 helices.

The structure also shows that there is a gap between the α and β apoproteins that is large enough for a quinol or quinone molecule to diffuse through. The optimisation process was discussed with Professor Wang-Otamo. He had concluded that testing the crystallisation conditions based upon the pH, buffer, protein concentration, preciptant, salts and additives did not yield any improvement in his crystallisation. For *Tch. tepidum* it was found that the most critical factor was detergent choice. The maltoside family of detergents were not optimal. A detergent screen indicated Fos-choline-10 and Fos-choline-12 to be good detergents. It was this change that helped him obtain well diffracting crystals.

3.4. Conclusion

Core complexes from *Rps. palustris*, *Rps. acidophila* and *Alc. vinosum* were purified to homogeneity in a range of different detergents. The most successful crystals were obtained from *Alc. vinosum*. The purification of these core complexes could still be optimised by the addition of a His-tag to the H-subunit of the RC in order to selectively purify the core complex over the chaperonin GroEL and the LH2 complexes. Sparse matrix crystallisation screens were tested initially, and showed that MemGold [106], a screen specifically designed for crystallisation of α-helical membrane proteins produced the most promising crystal hits. Core complex crystals from *Rps. acidophila* did not diffract, *Rps. palustris* crystals diffracted to 20 Å and were not optimisable. Crystals of the core complex from *Alc. vinosum,* however, were both reproducible and optimisable. Within a year crystals had been optimised that diffracted to a resolution limit of 15 Å to 4.5 Å by testing over 6000 different conditions. A clear improvement in the crystal contacts within the unit cell for these data was observed when compared to the previous data from *Rps. palustris*. This appears to be due to the presence of the additional, periplasmic cytochrome-*c* subunit. Crystallisation trials could be further optimised by dehydration

of crystals prior to data collection and further screening of additives, detergents and lipids.

A data set processed and scaled to 5.15 Å of the core complex from *Alc. vinosum* was phased using MR. The MR model generated was composed of the RC coordinates from *Tch. tepidum* [134] and the LH1 complex coordinates from *Rps. palustris* core complex structure [18]. The solution produced good statistics that suggested the solution was correct. However, a MR solution produced using just the coordinates of the RC as the search model did not produce a solution with electron density for the LH1 complex.

JB refinement improved the $R\text{-}_{factors}$ from 55.7 to 43%. For the refinement to work the orientation of the Protein W from the LH1 complex of *Rps. plaustris* had to be moved into the plane of the α-apoproteins of the LH1 complex. From the refinement it appears that in *Alc. vinosum* the LH1 ellipse is complete and composed of 16-αβ subunits. However, the density for the LH1 complex is very weak at other areas of the complex, and so this species of bacteria may have evolved a different mechanism for Q/QH_2 exchange. At this resolution the data presented here are not conclusive; however, the data presented for the *Tch. tepidum* core complex by Professor Wang-Otamo conclusively shows a closed LH1 complex with no PufX orthologue. In order to improve the data presented here, the crystals need to be further optimised via further additive screens, dehydration and testing different detergents with smaller micelles. It would also be prudent to attempt experimental phasing using heavy atoms, or utilising the Fe atoms that are present within the cytochrome-*c* subunit and RC. Since Professor Wang-Otamo and co-workers have had such success with the detergents Fos-choline-12 and Fos-choline-10, purification and crystallisation will also be attempted in these new detergents using both sparse matrix crystallisation screens and the conditions used by Professor Wang-Otamo.

4. Locating Protein W in *Rps. palustris* strain 2.1.6

A low-resolution structure of the LH1:RC 'core' complex isolated from *Rps. palustris* indicated the presence of an additional peptide in the LH1 ellipse (see section 1.3.2.3 of Introduction) [18]. This peptide, denoted 'Protein W', was hypothesised to be orthologous to the PufX protein found in the *Rhodobacter* genus of purple photosynthetic bacteria due to its close proximity to the terminal quinone binding site (Q_B) of the RC and its disjointed orientation relative to the α-apoproteins. It was hypothesised that Protein W could be involved in Q/QH_2 exchange, and therefore cyclic electron transfer, by incorporating a gap in the LH1 complex large enough for a QH_2 molecule to diffuse through.

Despite identifying the presence of Protein W in the structure of the core complex, the resolution of the data was only sufficient to provide secondary structural information. Hence, with no protein or gene sequence, it has not been possible to locate in the genome sequence. PufX proteins between different species within the same genera share approximately 20% sequence homology. This makes it impossible to identify possible putative Protein W proteins using BLAST software. Additionally, Protein W has proved difficult to isolate purify and identify by MS [18]. It is thought that this is due to its hydrophobicity. In the *Rhodobacter* genera, *pufX* is always located within the *puf* operon downstream of *pufM*. This is not the case for the *Rhodopseudomonas* genera. All that is known about Protein W is that it is a membrane inserted α-helix that lies in a disjointed orientation within the α-apoprotein ring of the LH1 complex, opposite the H-subunit. Previously open reading frames (ORF) have been analysed for the possibility of forming membrane inserted α-helices that contain the GXXXGXXXG motif, however, it was recently shown that this is a pseudo-motif [77].

The aim of this chapter was to identify the gene responsible for transcribing Protein W, and to experimentally determine the function of

the protein by constructing a Protein W knock out (KO) strain of *Rps. palustris*.

4.1. Purification of the core complex

To isolate Protein W core complexes from *Rps. palustris* were purified to homogeneity using sucrose density gradient (SDG) centrifugation, anion exchange chromatography and size exclusion chromatography (SEC) as described in section 2.4 of the Materials & Methods and Chapter 3. These techniques select for the core complexes on the basis of density, charge and size. Throughout the purification process the stability and purity of the complexes was monitored using UV-visible absorbance spectroscopy (Figure 4-1) and SDS-PAGE (Figure 4-2).

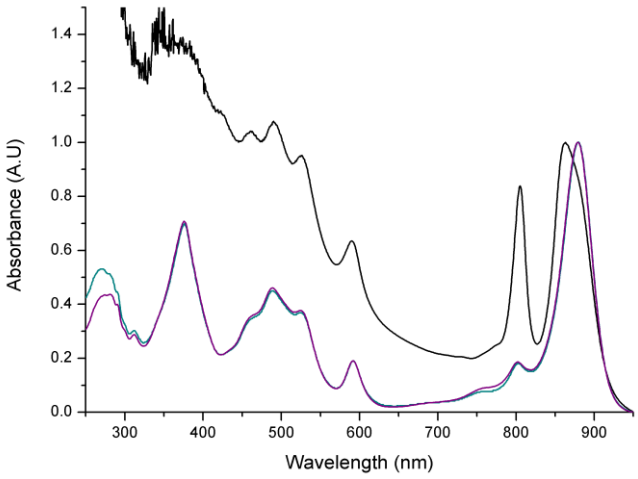


Figure 4-1 Absorbance spectra of the core complex from *Rps. palustris* strain 2.1.6 solubilised in LDAO. Black line shows the absorbance spectra of solubilised membranes. Blue line shows the isolated core complex post SDG centrifugation. The purple line is the core complex post SEC on an S200 column. The peak at 880 nm is the Q_Y absorption of the Bchl molecules. The small peak at 800 nm is the RC, whilst the large 800 nm peak in the membranes is due to Q_Y absorption of the B800 Bchl molecules from the LH2 complexes. 590 nm is where the Q_X peak of Bchl absorb. The three peaks at 450 - 550 nm are the carotenoids within the complexes. The peak at 390 nm is the soret peak. Finally at 280 nm is the protein absorbance. Throughout the purification process this peak should decrease until an optical ratio (880:280) is < 2.4

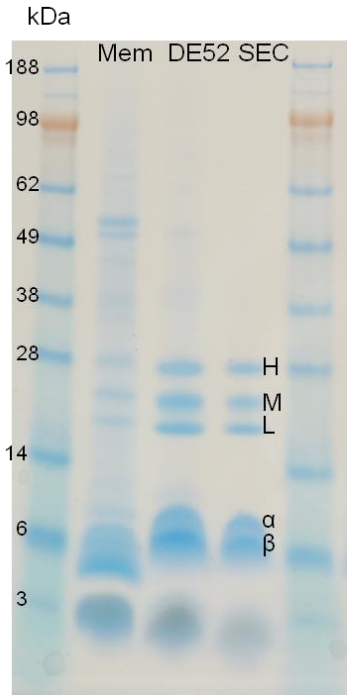
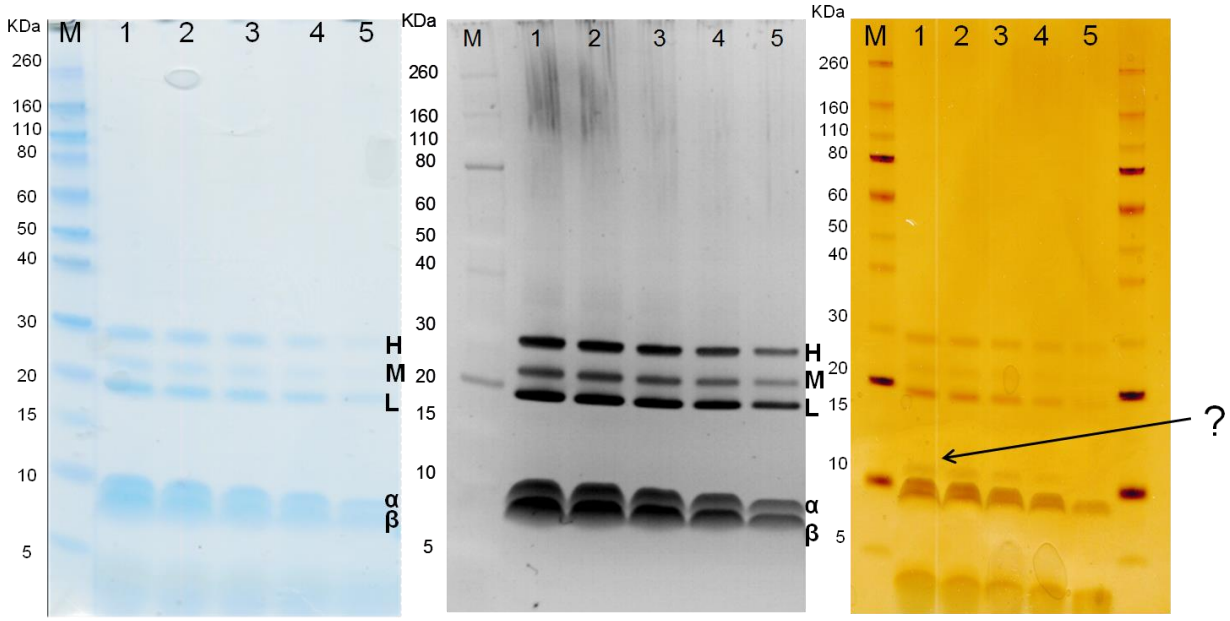


Figure 4-2 SDS polyacrylamide gel stained with SimplyBlue showing the stages of purification of the core complex from *Rps. palustris* strain 2.1.6 solubilised in LDAO. The lanes contain samples from the solubilised membranes, post anion exchange, and post SEC from left to right. The purified complex contains just the subunits from the RC, H, M & L and LH1; α and β subunits. No contaminants are observable in this stain.



1:OD_{880} = 10 cm^{-1}; **2**:OD_{880} = 8 cm^{-1}; **3**:OD_{880} = 6 cm^{-1}; **4**:OD_{880} = 4 cm^{-1}; **5**:OD_{880} = 2 cm^{-1}

Figure 4-3 SDS polyacrylamide gels of purified core complexes from *Rps. palustris* stained with SimplyBlue, SyproRuby and Silver stain. An additional band at approximately 11 kDa was only observed in the gel stained with Silver stain. The bands of the RC were identified corresponding to the work published by Clayton and Haselkorn, 1972 [137].

Light harvesting proteins are protein scaffolds that hold and maintain the cofactors in the appropriate protein environments and distances for efficient energy and electron transfer to occur. This causes the absorbance properties of the cofactors to become shifted, which allows absorbance spectroscopy to distinguish between correctly folded, partially unfolded, and unfolded protein states. Figure 4-1 shows the absorbance spectra of the core complex from solubilised membranes and purified core complexes. Purification of the core complex results in a reduction of absorbance at 280 nm, as protein contaminants have been removed. The absorption properties of the cofactors within the protein complexes remain the same suggesting that the protein was conformationally stable and the cofactors were maintained within the native protein environment.

The purity of the protein was also analysed using SDS-polyacrylamide gel electrophoresis. This separates denatured protein subunits, based on size and can subsequently be analysed by mass spectrometry. Figure 4-2 shows the protein subunits throughout different stages of the purification process; membranes, post anion exchange and post SEC, stained with SimplyBlue stain. The number of contaminants visibly decreased during the purification process until the protein was rendered pure with no visible contaminants post SEC. However, SimplyBlue is not the most sensitive stain and may not stain very hydrophobic subunits well. The pure sample contains the RC subunits; H, M & L - subunits, and the LH1 α and β apoproteins. There was no additional band that could be assigned to Protein W. It was hypothesised that if this peptide is orthologous to PufX it should be approximately 13 kDa in size. Different stains have different sensitivities to lower concentrations of protein, but also to more hydrophobic peptides. Therefore denatured core complexes were subsequently analysed using SyproRuby and silver stain (Figure 4-3). All gels show that the core complexes were very pure with no observable contaminants, and an additional band at approximately 11 kDa was observed when stained with silver stain. This

band was excised and trypsin digested. The tryptic fragments were analysed by nLC MS-MS at the University of Sheffield by Dr Philip Jackson. This analysis highlighted the possibility that Protein W was one of the following three candidate proteins; RPA0067, RPA1495 and RPA2969.

4.2. Candidate Protein W proteins

The three candidate Protein W proteins described above, are all hypothetical proteins located in the *Rps. plaustris* strain 2.1.6 genome. The candidate proteins were analysed using BioEdit (Thomas Hall, 2007 http://www.mbio.ncsu.edu/BioEdit/bioedit.html) to compare the protein sequences, protein BLAST searches (NCBI) to find similar proteins and the Meiler lab online server JUFO [138], which predicts protein secondary structure and hydrophobicity. None of the candidate proteins share more than 10% sequence homology at the protein level with any PufX protein from the *Rhodobacter* genera. This is not unexpected as PufX proteins from different species within the *Rhodobacter* genera share only 23% sequence homology. The location of these genes within the *Rps. palustris* genome CGA009 is shown in Figure 4-4. Protein BLAST searches of RPA0067, RPA1495 and RPA2969 showed similarities to other hypothetical proteins from organisms in the α-proteobacteria family.

RPA1495 is located closest to the *puf* operon, the operon which encodes most of the genes of the core complex. However, the RPA1495 protein was predicted by JUFO to be soluble. Previously, Dr Tatas Brotosudarmo studied spectral variants of LH2 complexes from *Rps. palustris*. The purified complexes were analysed by nLC MS-MS at the University of Dundee and the hypothetical protein RPA1495 was observed. STRING 9.05 [139] an online server that hosts information about known and predicted protein interactions, based on previous experimental work and genomic information, suggests that RPA1495 interacts with PucC (the putative Bchl transporter involved in the expression of LH2

complexes), as well as the LH2 α and β-apoproteins (Figure 4-5). It is therefore hypothesised that RPA1495 is a contaminant that co-purifies with the LH2 complex, as does PucC, and in this case could have been a contaminant in the core complex. It is possible that RPA1495 might be a terminal domain of PucC, hence the co-purification with LH2 complexes. However, its presence in the purified core complex suggests that there was a minor LH2 contamination. LH2 contamination is difficult to assess by denatured protein gels as the α and β apoproteins of both LH2 and LH1 are similar sizes. However, the absorbance spectra of purified complexes do not show any apparent spectral contributions at 850 nm. If there was LH2 contamination present, 2 strategies could be applied. First the LH2 genes could be knocked out. However, as there are 6 *pucBA* gene pairs spread throughout the genome, construction of an $LH2^-$ strain would be a large project. Secondly, a His-tag could be applied to the H-subunit of the RC. This would allow the core complex to be purified on a nickel affinity column, (immobilised metal affinity chromatography (IMAC)). As only the core complex would contain the His-tag, all other protein contaminants should be removed.

Analyses of RPA2969 predicted that this protein is also soluble. The location within the genome is close to *pucBAc* (3,408,431 - 3,408,798). STRING9.0 suggests that RPA2969 interacts with other hypothetical proteins and badE, badD and badF, subunits A, B and C of the benzyl-CoA reductase.

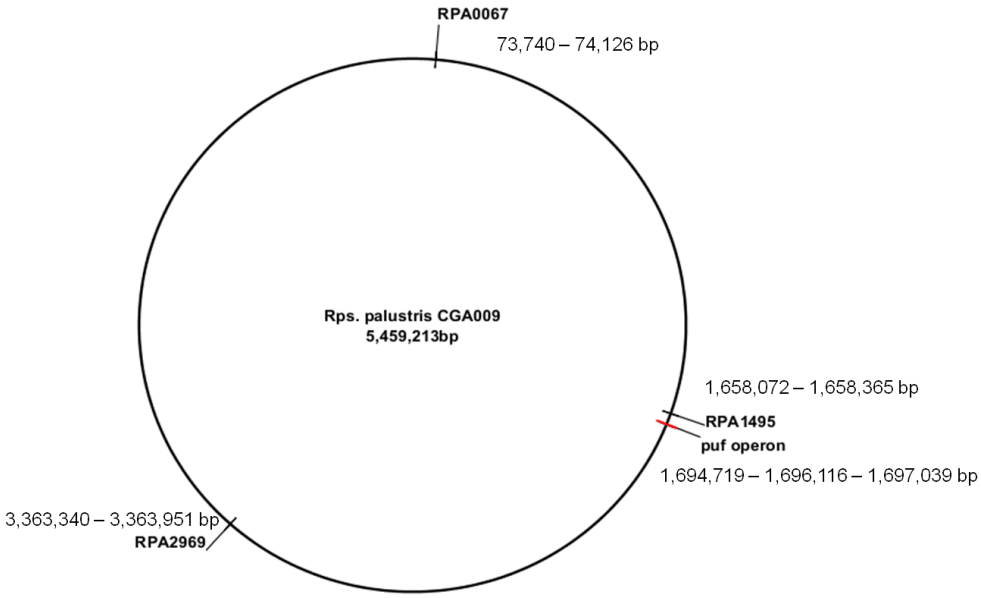


Figure 4-4 Candidate Protein W proteins in the *Rps. palustris* strain 2.1.6 genome. The location of the *puf* operon is highlighted to show where the candidate proteins are in relation.

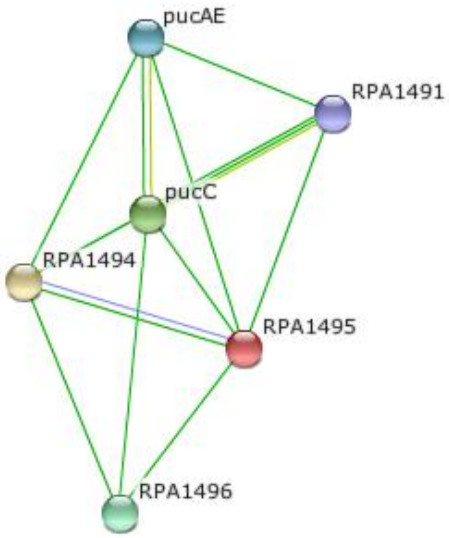


Figure 4-5 Output from STRING 9.0 [139] using RPA1495 as a query. It is predicted that RPA1495 interacts with PucC and the protein product of *pucA*$_E$ (α-apoprotein of the LH2 complex). This suggests that RPA1495 may interact and co-purify with LH2 complexes. Its presence in the core complex indicates that there was an LH2 contamination.

RPA0067 was the only candidate protein predicted to be membrane inserted and mainly α-helical. Homology modelling using the Phyre2 fold recognition server [140] presented a single transmembrane α-helix (Figure 4-6). The model was produced using the solution structure of

porcine delta-sleep-inducing2 immunoreactive peptide with 65.8% confidence. Although this confidence level is not very high the secondary structure prediction of the whole protein (Figure 4-7) shows high confidence levels in the regions predicted to be α-helical. As RPA0067 was the only candidate protein predicted to be membrane inserted and α-helical, it was hypothesised to be the most likely candidate for ProteinW.

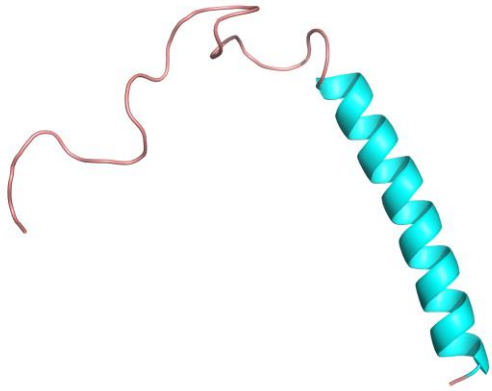


Figure 4-6 Homology model produced in Phyre2 fold recognition server [140] generated with 65.8% confidence. The model suggests that RPA0067 consists of one transmembrane α-helix and a large cytosolic loop at the C-terminus of the protein.

Alignments of RPA0067 and PufX from *Rba. capsulatus* show no sequence homology (Figure 4-9). Figure 4-8 shows an alignment of PufX protein sequences from *Rba. sphaeroides, Rba. capsulatus, Rba. veldkampii, Rba. blasticus* and *Rba. azotoformans.* The main similarities between these proteins is that they are single α-helical transmembrane proteins. They are also encoded by the *pufX* gene that in all the *Rhodobacter* species is located immediately 3' of *pufM* in the *puf* operon (see section 5.3 of chapter 5). The annotated genomes of other purple bacteria, such as *Rps. palustris* and *Alc. vinosum*, do not contain any gene or ORF similar to that of *pufX*. In *Alc. vinosum* the region of DNA downstream of *pufM* contains *pufC*, which encodes the periplasmic

cytochrome-*c* subunit, and 2 additional *pufBA* gene pairs, which encode the LH1 β and α apoproteins respectively.

Figure 4-7 Secondary structure prediction of RPA0067 produced by the Phyre2 fold recognition server [140]. The regions of RPA0067 that were predicted to be α-helical were high confidence, so although the model had a confidence level of 65.8%, the α-helical region formed by residues 30-100 have a high confidence.

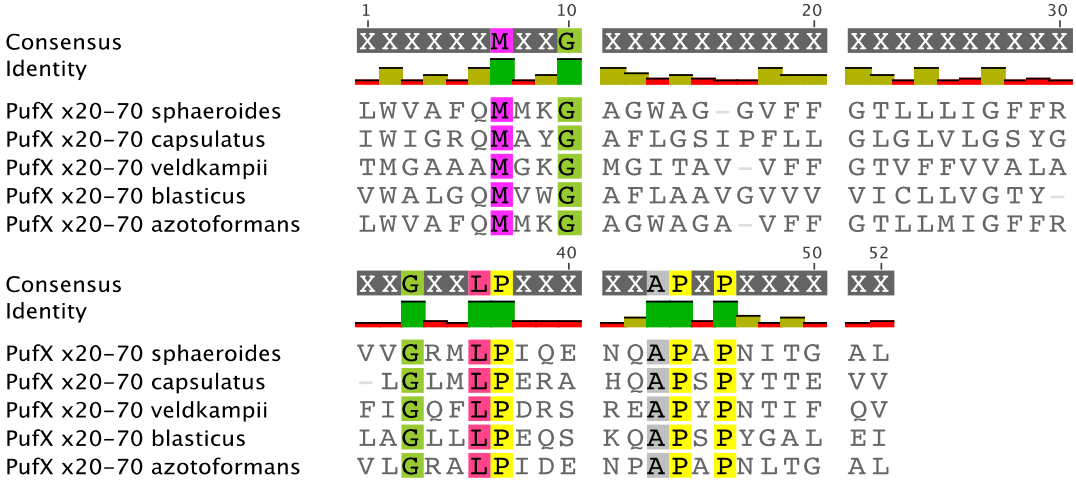


Figure 4-8 Alignment of PufX proteins from *Rba. sphaeroides, Rba. capsulatus, Rba. veldkampii, Rba. balsticus* and *Rba. azotoformans.* Fragments of the protein have been aligned, from residues 20 to 70. Despite these proteins having the same hypothesised function in all the species, there is very low homology.

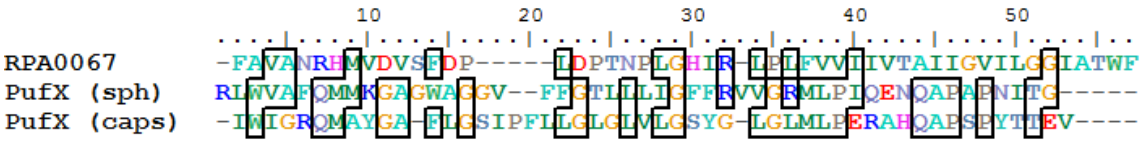


Figure 4-9 Alignment of RPA0067 and PufX from *Rba. capsulatus* (PufX (caps)) and *Rba. sphaeroides* (PufX (sph)). It is clear from this alignment that there is no sequence homology between these proteins. This would explain why analyses of genomic DNA sequences of other genera of purple photosynthetic bacteria did not identify any genes or protein products orthologous to PufX.

4.3. Construction of RPA0067KO in *Rps. palustris*

From the structure of the core complex from *Rps. palustris*, Protein W appears to have a key role in creating a gap in the LH1 complex [18]. It was previously suggested that an incomplete LH1 complex is required for cyclic electron transfer and photosynthetic growth to occur [67,73]. In the *Rhodobacter* genera the removal of PufX results in a bacterial mutant strain that fails to grow photosynthetically. Hence, if Protein W is orthologous to PufX, a mutant lacking Protein W should present the same PS^- phenotype. To test this hypothesis a RPA0067KO construct was designed and generated using the splice overlap extension (SOE) protocol as described in section 2.8 of the Materials & Methods and is outlined in Figure 2-6. Flanking primers upstream (US) and downstream (DS) were designed containing the restriction enzymes sites EcoRI and SphI respectively. Initially, these primers were used in addition to internal primers to amplify the US and DS fragments (Figure 4-10 and Table 2.1 in Materials & Methods), either side of the gene of interest. The internal primers contain complementary sequence overlaps composed of 3 codons of the US and DS DNA. This overlap region includes the start and stop codons of the gene of interest to ensure any genes DS of the KO remain in-frame. These DNA overlap sections anneal when exposed to low annealing temperatures due to the relatively unspecific nature of the overlap. The full length KO construct was amplified using the two flanking primers.

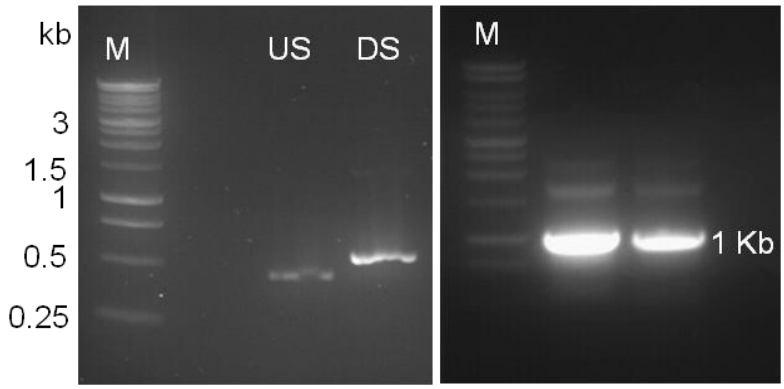


Figure 4-10 1 % (w/v) agarose gels of the US, DS and full length construct (1 Kb) of RPA0067 KO construct in *Rps. palustris* strain 2.1.6.

The full length RPA0067KO DNA fragment was cloned into the mobilisable pk18mobsacB vector [141]. Both the full length construct and the vector had been digested with the restriction enzymes EcoRI and SphI to create the sticky ends required for ligation. The modified vector was transformed into the *E. coli* strain JM109. The plasmid DNA was isolated (Figure 4-11) and sequenced by MWGeurofins. Once sequencing had confirmed the correct sequence of RPA0067KO_pk18mobsacB (Figure 4-12) the vector was transformed into *E. coli* λpir.17 cells [118]. This strain of *E. coli* was used as it contains the RP4 (IncPα) conjugative transfer apparatus that contains a pilus allowing for cell-to-cell contact between the host and recipient strains of gram negative bacteria. The mobilisable vector pk18mobsacB was used as it confers kanamycin resistance as a selectable marker and the gene *sacB*, which confers sensitivity to sucrose. When the recipient bacteria containing the plasmid are grown on agar supplemented with 10% (w/v) sucrose the plasmid 'dies' and the mutated DNA fragment is integrated into the host chromosome via allelic recombination [118,141].

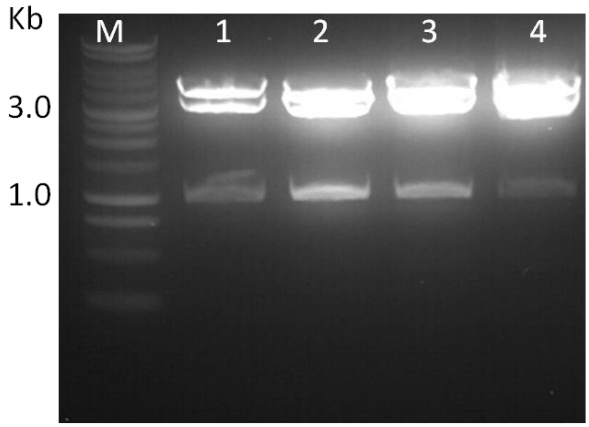


Figure 4-11 EcoRI and SphI double digest of RPA0067KO_pk18mobsacB plasmid DNA. The band at 1 kb is the RPA0067KO fragment.

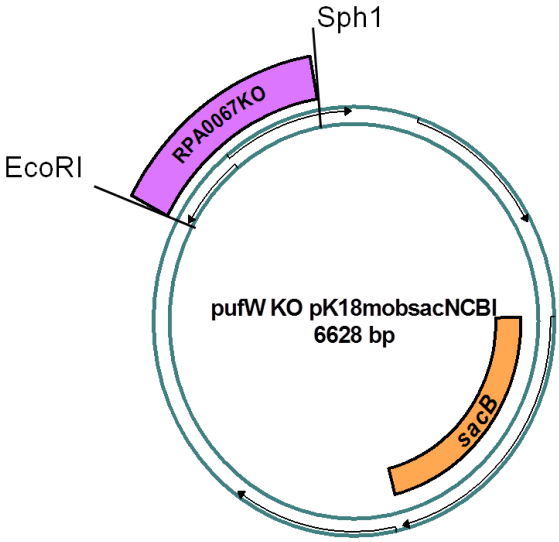


Figure 4-12 plasmid map of RPA0067KO-pk18mobsacB. The RPA0067KO insert was inserted into the multiple cloning site (MCS) between the EcoRI and Sph1 restriction enzyme sites. The pk18mobsacB vector [118] contains a Kanamycin antibiotic resistance cassette, which, was used as a selectable marker. The vector also contains the *sacB* gene, which confers sucrose sensitivity. This allows the vector to commit suicide when grown in the presence of sucrose, allowing allelic recombination to occur.

Donor *E.coli* containing the modified mobilisable vector were mixed in different ratios with WT *Rps. palustris* and grown on agar plates supplemented with Kanamycin as outlined in section 2.8.4 of the Materials & Methods. Initially a lawn of colonies grew (Figure 4-13 A). These colonies were re-grown in order to obtain single colonies (Figure 4-13 B) which were then grown in liquid culture supplemented with

Kanamycin in anaerobic conditions under illumination. Once the liquid cultures had entered the logarithmic phase of growth allelic recombination was completed by growing the cultures on agar supplemented with 10% (w/v) sucrose, allowing the RPA0067KO fragment to recombine with the host chromosome.

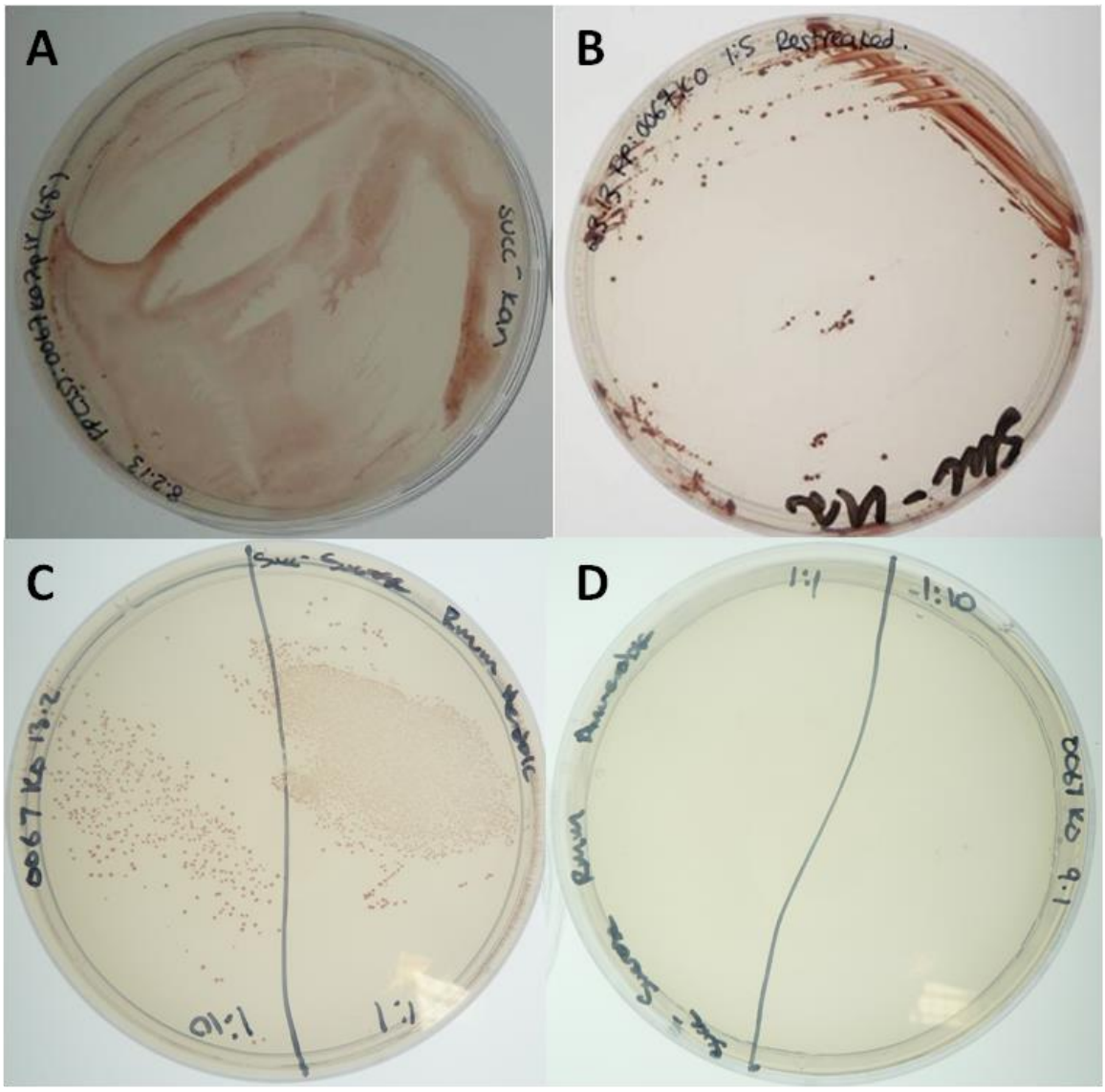


Figure 4-13 Conjugation and allelic recombination of RPA0067KO *Rps. palustris*. A: Initial conjugants of RPA0067KO_pk18mobsacB:*Rps. palustris* WT. The initial conjugants grew as a lawn and were re-grown to obtain single colonies. Colonies were grown in the presence of Kan, photosynthetically. B: secondary conjugants and single colonies of the RPA0067KO_pk18mobsacB. Colonies were grown in the presence of Kan, photosynthetically. C: Serial dilutions of the mutant culture grown on agar supplemented with 10% (w/v) sucrose grown chemoheterotrophically. D: Serial dilutions of the mutant culture grown on agar supplemented with 10 % (w/v) sucrose, grown photosynthetically.

As the modified vector is no longer present these plates were duplicated and grown both aerobically and anaerobically in the light Figure 4-13 C

and D. WT *Rps. palustris* are so metabolically versatile they can grow photosynthetically and chemoheterotrophically, however, the RPA0067KO strain failed to grow photosynthetically but did grow chemoheterotrophically. This PS^- phenotype suggests that the removal of RPA0067 is lethal for photosynthetic growth, and hence may indeed be Protein W and orthologous to PufX.

To ensure that the pk18mobsacB vector had been removed single colonies from the sucrose plates were grown in duplicate on agar plates supplemented with and without kanamycin (Figure 4-14). This selected for any colonies which still contained the pk18mobsacB vector, and therefore, cultures where allelic recombination had been unsuccessful. These cultures were all grown chemoheterotrophically to ensure that the colonies would grow.

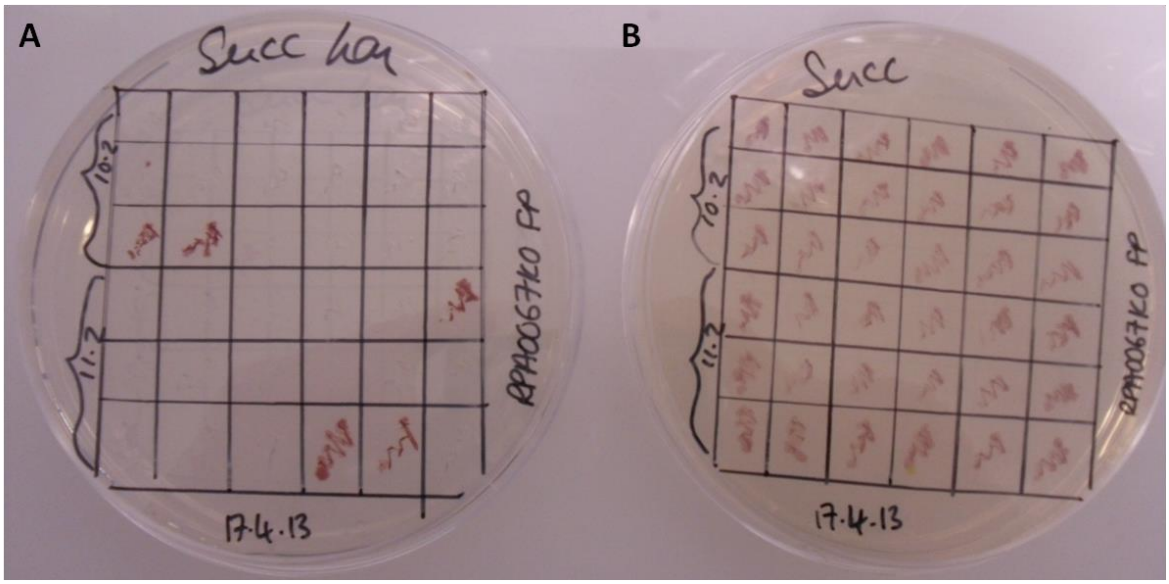


Figure 4-14 Grid plates to test for the presence of the vector pk18mobsacB. Single colonies were picked and streaked onto the two agar plates, one supplemented with kanamycin as a negative control. Colonies that grew only on the agar plate with no kanamycin, and therefore did not contain pk18mobsacB, were tested by colony PCR to ensure the RPA0067KO fragment had been inserted.

Colonies that did not contain the pk18mobsacB vector were tested by colony PCR using primers designed further US and DS (Table 2.1 in Chapter 2 Materials & Methods) of the flanking primers used to generate the KO fragment. By analysing the size of the DNA amplified it

was possible to see in which colonies the *RPA0067* gene had been removed (Figure 4-15). *Rps. palustris* WT genomic DNA was also amplified as a negative control. The fragment size of the amplified DNA if RPA0067 is present is 1.5 kb, whilst the RPA0067KO strain of *Rps. palustris* should produce a band at 1 kb. The colony PCR showed that colonies 1,2 and 3 all produced fragments at 1 kb. These colonies were subsequently grown in liquid culture chemoheterotrophically and genomic DNA was extracted as outlined in section 2.10 of the Material & Methods.

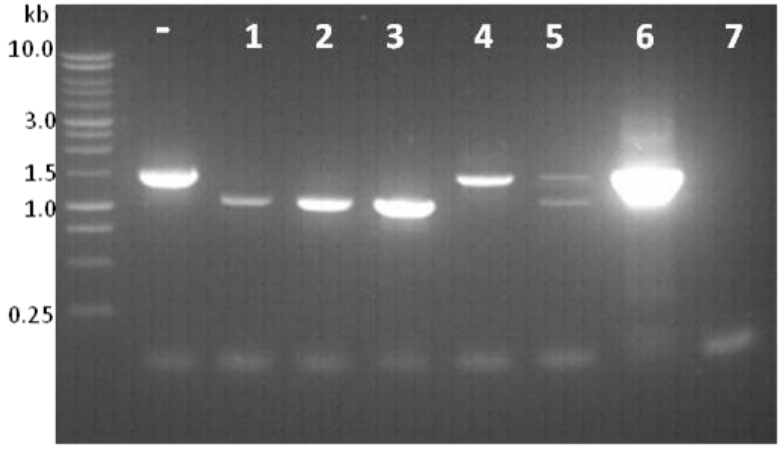


Figure 4-15 Colony PCR of the potential RPA0067KO strain, analysed on a 1% (w/v) agarose gel. The negative control was *Rps. palustris* WT genomic DNA. The band at 1.5 kb corresponds to the fragment size if RPA0067 is present in the genome. From the gel it is clear that colonies 1,2 and 3 contain the correct size (1 kb) for the RPA0067KO strain of *Rps. palustris*.

The RPA0067KO fragment was amplified from the genomic DNA, cloned into the pGEM vector and transformed into DH5α competent cells. The extracted plasmid DNA was sequenced by MWGeurofins and confirmed that the homologous recombination had been successful.

4.4. Conclusion

The additional peptide observed in the low resolution crystal structure of the *Rps. palustris* monomeric core complex, Protein W, has remained elusive since its discovery in 2003 [18]. It was hypothesised that Protein W was orthologous to PufX, found in the *Rhodobacter* genera.

Here, purified core complexes from *Rps. palustris* were analysed using denatured protein gels. An additional band was observed at approximately 11 kDa. This band was excised, trypsin digested and analysed by nLC-MS-MS to reveal 3 candidate Protein W proteins; RPA0067, RPA1495 and RPA2969. Basic bioinformatical anlaysis suggested that RPA0067 was the most likely candidate for Protein W, as it was the only protein predicted to be membrane inserted and α-helical. As PufX had previously been shown to be essential for photosynthetic growth and cyclic electron transfer, if Protein W is orthologous to PufX, a KO strain of *Rps. palustris* should produce a similar PS^- phenotype. To test this theory an RPA0067KO strain of *Rps. palustris* was constructed using SOE PCR, conjugative transfer and allelic recombination. The RPA0067KO strain on first analysis presented a PS^- phenotype suggesting that RPA0067 does affect photosynthetic growth. The work is still in its infancy, and as such growth curves comparing WT and KO strains of *Rps. palustris* grown both photoysnthetically and chemoheterotrophically should be obtained. The current data does not conclusively prove that PRA0067 is Protein W, however the absence of RPA0067 does affect photosynthesis. Purified core complexes from both strains could also be analysed on denaturing gels to see if the band at 11 kDa is not present in the RPA0067 KO strain. The purified core complexes could also be analysed by using single molecule spectroscopy [142]. Using this technique it will be possible to analyse if the core complex from the RPA0067KO are composed of a complete LH1 ring. If this is the case, then RPA0067 is likely to be Protein W.

5. Genome sequence and annotation of *Rps. acidophila* 10050

It has previously been shown that different growth conditions can cause spectral variants of LH2 complexes to be expressed [52]. Different strains of the bacterium *Rps. acidophila* have shown interesting differences. Strains 7050 and 7750 both express the B800-820 LH2 complex at low light (LL, 20 μmol m^{-2}) conditions and the B800-850 LH2 complexes at high light (HL, 60 μmol m^{-2}) conditions. It has also been observed that strain 7050 undergoes a gradual change from predominantly expressing B800-850 at HL to B800-820 at LL, but there always appears to be some B800-850 component. Whereas, temperature, as well as light, regulates LH2 expression in strain 7750. When this strain is grown at LL and lower temperatures (20-24°C) only the B800-820 LH2 complexes are expressed [52].

In contrast to strains 7050 and 7750, strain 10050 only expresses B800-850 LH2 complexes (see section 1.3.1 of the Introduction). When grown at LL intensities it adapts by increasing the size of the PSU containing B800-850 LH2 complexes, and increases the amount of ICM present [44,143]. It is possible that the lack of additional variants of LH2 complexes being expressed in 10050 may have resulted in a homogeneous sample and, therefore, aided the successful crystallisation and resulting X-ray crystal structural determination of the LH2 complex [43].

The B800-820 LH2 complex has been shown to be advantageous at LL intensities as it is more efficient at funnelling the energy of a photon to the core complex in the forward reaction only, in comparison with the B800-850 LH2 complexes [44,52,143].

Previously, analysis of the *puc* operon of *Rps. acidophila* strains 7050, 7750 and 10050 showed that as a minimal estimate of *pucBA* gene pairs present there are 8 gene pairs in 7050, 3 gene pairs in 7750 and

5 gene pairs in 10050 [44]. It is surprising that *Rps. acidophila* strain 10050 contains multiple *pucBA* gene pairs as only the B800-850 LH2 complex has ever been expressed under laboratory conditions. It is possible that these additional genes are silent and the result of a recent gene duplication [44]. However, if this is the case, bearing in mind how efficient bacteria are at removing superfluous DNA from the genome it is surprising that these silent genes have not been deleted. Another possibility is that there are different regulatory pathways in *Rps. acidophila* strain 10050 that control the expression of the individual *pucBA* gene pairs outwith the regulatory pathways defined and controlled by laboratory growth conditions and standardised media. In other words, these excess genes are required by the bacteria under the conditions prevalent in the natural environment of the cells and not under laboratory conditions.

In this chapter genomic DNA of *Rps. acidophila* strain 10050 has been isolated and sequenced, using Illumina technology, in collaboration with Dr Graham Hamilton and Dr Pawal Herzyk at the University of Glasgow. Strain 10050 was chosen for this project as the LH2 complexes are structurally well-defined. This may allow mutations in the LH2 complexes to be generated which can be studied both structurally and functionally. The aim of this work was to determine exactly how many *pucBA* gene pairs are present in the 10050 genome, their operon organisation and to establish if any of these *pucBA* genes are able to encode for any putative B800-820 LH2 complexes. The architecture of the *puf* operon, which encodes some of the genes involved in expressing the core complex, was also analysed.

5.1. Genomic isolation & annotation

Genomic DNA was isolated from *Rps. acidophila* strain 10050 cells in the logarithmic stage of growth, as described in section 2.9 of the Materials & Methods. The genomic DNA was sequenced using Illumina technology by Dr Graham Hamilton and Dr Pawel Herzyk at the

University of Glasgow, as described in section 2.9 of the Materials & Methods. Illumina technology is a next generation sequencing (NGS) technique that is quick and inexpensive relative to the first generation automated Sanger sequencing techniques used to sequence the human genome [144]. This technique utilises solid-phase amplification on a glass slide and sequences during the amplification cycles. Unfortunately, the NGS methods are error prone. As such the read length of each fragment was trimmed from 101 bp to 85 bp to ensure that the sections of sequence data that were likely to have base errors and higher noise were removed from the database.

From the sequencing 45,000,000 reads were obtained with 85 bp read length. Assuming that the genome is approximately 5.5 million bp, the expected coverage per base is 600. Currently there are 37 contigs, which have been scaffolded into 9 supercontigs (Figure 5-1). Supercontigs (or scaffolds) are contigs that have been ordered but are joined by a string of N's as there is not enough information to define the sequence between the contigs. As there are no similar genomes currently sequenced, the sequence data were assembled *de novo* by Dr Graham Hamilton, comparing the sequence data to the published *Rps. palustris* strain 2.1.6 genome [50]. As there are gaps between the contigs and within the supercontigs, the construction of a mate-pair library is required to gain enough sequence information to enable the full genome assembly. Mate-pair sequencing is very useful for the assembly of genomes *de novo* and in genome finishing. This technique fragments the genomic DNA forming sections of DNA between 2-5 Kb. These fragments are labelled with biotin labelled dNTPs, and circularised. The circular DNA is then fragmented and affinity purified using a streptavidin affinity column. These fragments are ligated to Illumina Paired End Adapters and amplified. The final library that is constructed contains short fragments of DNA segments that were originally separated by several Kbp, hence providing clues as to the

order of the sequence data obtained in the primary sequencing experiments. This work is currently on-going.

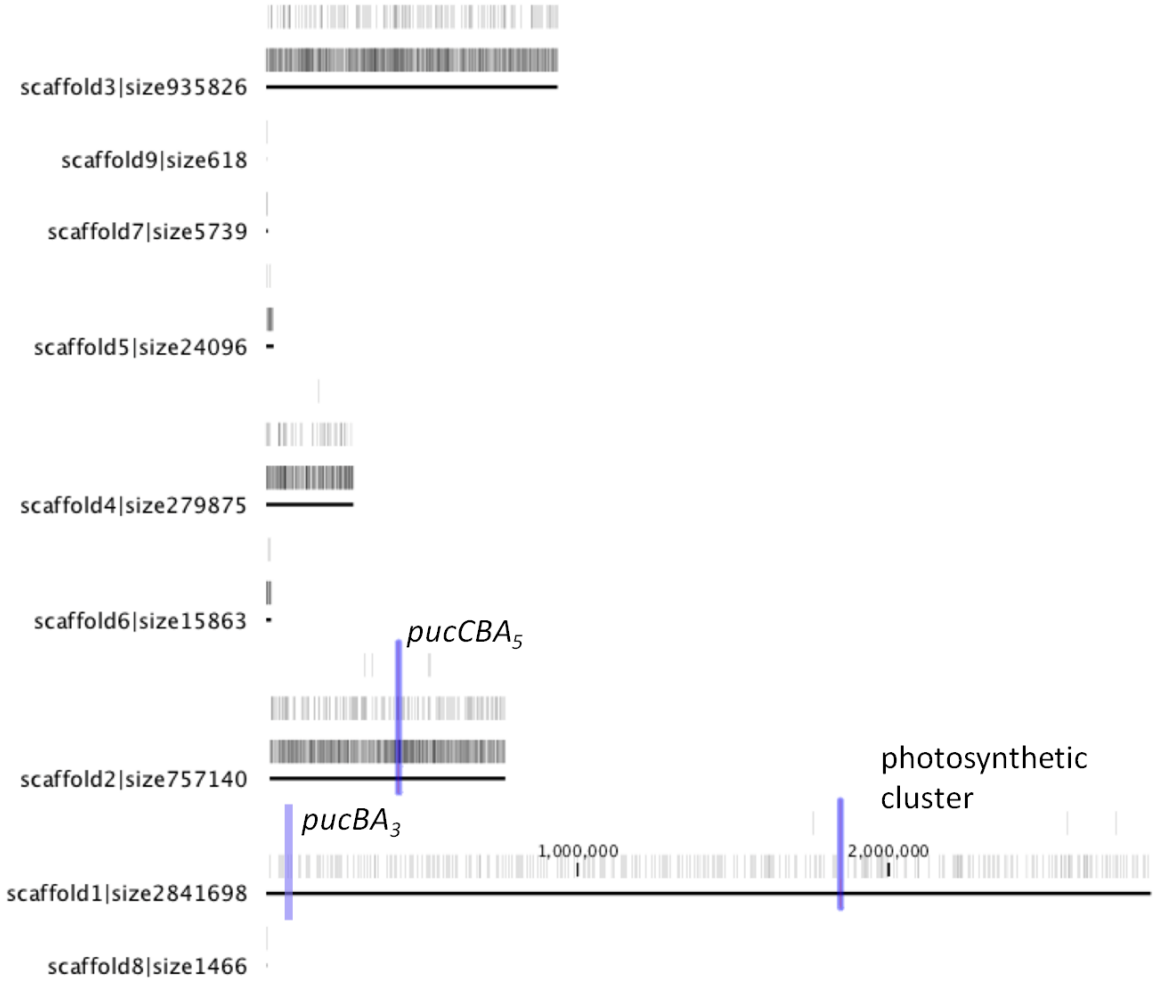


Figure 5-1 Comparison of the 9 supercontigs from the *Rps. acidophila* strain 10050 genomic sequence data. The genes involved in light harvesting were found on scaffolds 1 and 2 only.

The 9 supercontigs have been annotated and the photosynthetic gene cluster was located (Figure 5-1). This contains the *puf* operon (which encodes the LH1 ring, L, M and cytochrome-*c* subunits of the RC), genes involved in Bchl, heme and Car synthesis and is finished with *puhA*, the gene that encodes the H-subunit of the RC. The photosynthetic gene cluster is located from 1,841,340 to 1,864,460 bp on scaffold 1. There is also a *pucBA*$_3$ operon (which will be discussed in detail in section 5.2) on scaffold 1, although this is located from position 16,000 to 18,000 bp. Scaffold 2 contains an additional *pucCBA*$_5$ operon at position 413,500 to 417,850 bp.

5.2. *puc* operon

From the sequence data available 8 *pucBA* gene pairs were identified (Figure 5-2), which appear to be arranged into 2 distinct operons. As previously observed, the *pucB* genes are all preceded by a Shine-Dalgarno ribosome binding sequence, and the initiator codon GTG 4 bp US of the *pucB* start codon. The *pucBA* gene pairs are separated by 11-15 bp and contain the absolutely conserved AXXXH (β) and AXXXHXXXL (α) motifs. There is, however, one exception. The *pucA5* gene contains AXXXHXXXI on the α-apoprotein. This is a conservative change from L to I. To ensure that this is not a sequencing error, the area should be amplified by PCR, cloned and sequenced; however, this could not be done here due to the time restraints of this project.

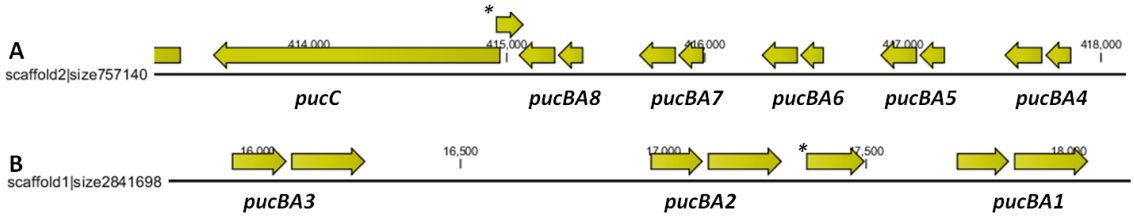


Figure 5-2 *puc* operons from *Rps. acidophila* strain 10050 located from the partially assembled genome sequence. A: *pucBA4-8*C operon, located on scaffold 2, previously identified by [145].. The gene products of the *pucA* genes all encode B800-850 type α-apoproteins. B: secondary *puc* operon, located on scaffold 1, that contains *pucBA3-1*. There is no *pucC* present in this operon. The genes in this secondary operon all encode B800-820 α-apoproteins, and so it is hypothesised that this operon is not expressed. It is unknown if that is because the right growth conditions haven't been found, or if the genes are silent and are lacking the regulatory genes or essential gene products, such as PucC. * highlights the presence of hypothetical proteins.

Figure 5-2 A displays a schematic diagram of the *puc* operon with 5 *pucBA* gene pairs followed by *pucC*. This gene encodes a putative Bchl transporter that has been shown to be essential for high level transcription of the *pucBA* genes and, therefore in the expression of the LH2 complexes [146]. Amino acid analyses of these α-apoproteins (*pucA4-8*) suggest that these genes all encode the B800-850 LH2 complexes, as the Bchl binding site (see section 1.3.1 of Introduction) contains α-Y44 and α-W45 (Figure 5-4). It has been observed that if the residues at these positions are mutated from aromatic residues to small

hydrophobic residues, such as methionine, the absorbance maxima of the Bchl-*a* molecules shifts from 850 nm to 820 nm as described in section 1.3.1 of the Introduction. This is due to a change in the site energy of the Bchl molecules [44,46,147]. Interestingly, *pucBA1-3* gene pairs (Figure 5-2 B) do not have a *pucC* gene downstream of *pucA*, suggesting that these genes may not be expressed due to the absence of PucC. Analyses of the protein products of these *pucBA* gene pairs suggests that they encode the B800-820 LH2 complexes, as the α-apoprotein sequences all contain α-F/Y44 and α-M45 (Figure 5-3).

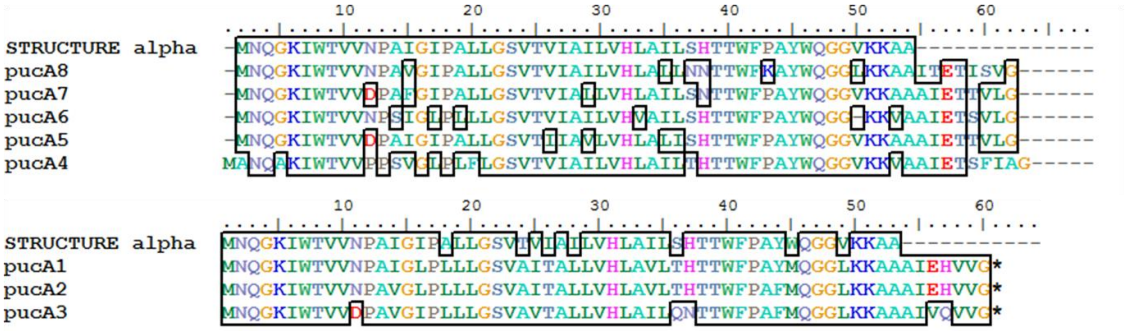


Figure 5-3 *Rps. acidophila* strain 10050 α-apoprotein sequence alignments compared to the α- apoprotein sequence used to solve the LH2 structure [43]. None of the identified α-apoprotein sequences are a perfect match for the LH2 α-apoprotein sequence that was obtained from the LH2 structure. The closest match was α6 (PucA6) with only 5 changes A14S, I16L, A18L, L32V, A52V) however, all these changes are conservative. *pucA1-3* encode the B800-820 type LH2 complexes.

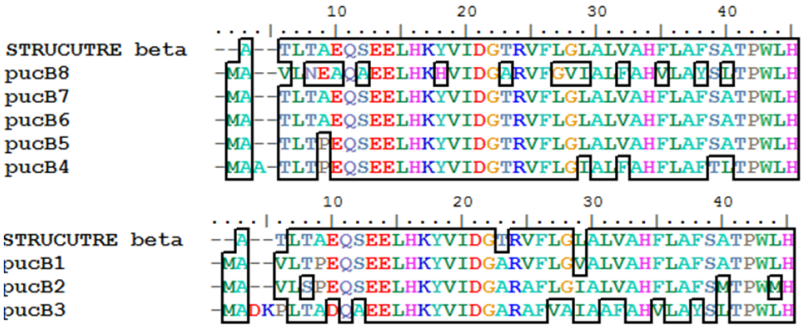


Figure 5-4 *Rps. acidophila* strain 10050 β-apoprotein sequence alignments compared to the β-apoprotein sequence used to solve the LH2 structure. Here it is clear that only β6/7-apoprotein from the genomic sequence has 100 % sequence identity to the sequence of the β-apoprotein used to solve the structure. *pucB1-3* encode the B800-820 type LH2 complexes.

It would be interesting to analyse this secondary *puc* operon and establish if the addition of the *pucC* gene would generate a mutant

strain of *Rps. acidophila* strain 10050 that could express the B800-820 LH2 complexes. It may also be possible to compare the 2 *puc* operons within the same species to analyse the differences in genes and ORFs present. As one operon appears to express LH2 complexes and the other does not, this may allow novel regulatory genes to be observed, if any are present.

BLAST analysis of the protein products of the *puc* operons confirmed that *pucBA5*-8 and *pucC* were identical to the previously observed *puc* operon published in Simmons *et al.,* 2000 [145]. This work however, did not identify *pucBA4* or the second *puc* operon, which have been observed in this work. BLAST searches using the translated *pucBA1-3* sequences showed that *pucA3* from strain 10050 was 100% identical to *pucA1* from strain 7050 [44], *pucA2* from strain 10050 was 100% identical to *pucA2* from strain 7050 [44], and *pucA1* from strain 10050 was 100% identical to *pucA3* from strain 7050 [44]. A search of the translated *pucB1-3* genes showed that both *pucB1* and *pucB3* from strain 10050 were 100% identical to *pucB3* from strain 7050 and *pucB1* from strain 7750 respectively. The translated sequence of *pucB2* showed 98% identity to *pucB2* from strain 7050 [44]. Unfortunately, the genomes of *Rps. acdiophila* strains 7050 and 7750 have not been sequenced, and so it is unknown how many *puc* operons there are and how many *pucC* genes are present.

If all the *pucBA* gene pairs have been found, it should be possible to identify which gene pairs encode the B800-850 LH2 complex in strain 10050. It is thought that only one *pucBA* gene pair is transcribed as the B800-850 LH2 complex is homogeneous and was sequenced by Zuber, 1993 [148]. Comparison of the α and β- apoprotein sequences obtained in this project to the sequence used to solve the structure of the LH2 complex [148] suggests that either the β6 or β7-apoproteins (*pucB6/7*) are transcribed as they show 100% sequence identity to the sequenced

β-apoprotein expressed. However, comparison of the α-apoprotein sequences presented no 100% match (Figure 5-3 & Table 5-1).

It would be expected that if *pucB6/7* were being transcribed, then *pucA6/7* should be co-transcribed if they are under the same control. The gene product (α7) of *pucA7* has the most sequence identity to the α-apoprotein of the structure, with only 4 residues different. Most of the differences are conservative. However, as highlighted in Table 5-1, there are also some non-conservative changes. The protein sequence of α-6 has more changes, however, they are all conservative changes.

α-apoprotein (gene)	Number of differences	Differences
α4 (pucA4)	10	G4A, **N11P**, A13S, I14V, I16L, A18L, **L19F**, S36T, A52V
α5 (*pucA5*)	5	N11D, V25I, I28V, I34L, L35I
α6 (*pucA6*)	4	A13S, I16L, A18L, L28V, A51V
α7 (*pucA7*)	6	N11D, **I14F**, I28L, **H37N**
α8 (*pucA8*)	5	I14V, I34L, **S36N**, **H37N**, **P42K**, V49L

Table 5-1 Outline of the residue differences between the sequenced α-apoproteins and the sequence used to resolve the high resolution structure of the *Rps. acidophila* strain 10050 LH2 complex. The differences in blue highlight non-conservative residue changes.

It is possible that over time strain 10050 may have naturally mutated to generate these differences in the *pucA* genes. This was recently observed for the *Blc. viridis* RC [149]. It is also possible that during the

culturing for genomic DNA isolation, the selection of a single colony of *Rps. acidophila* strain 10050 may have selected a random mutant. However, as there are currently still some gaps in the sequence data, it is also possible that not all of the *pucBA* gene pairs have been found.

However, if *pucBA7* is being expressed, the non-conservative changes, such as I14F and L19F, should be readily detectable through analyses of the crystal structure of the *Rps. acidophila* that had been cultured to extract the genomic DNA. The LH2 complexes from *Rps. acidophila* strain 10050 were isolated and purified as described in section 2.4 of the Materials & Methods. Figure 5-5 compares the absorbance spectra throughout the purification process, from the SDGs, anion exchange (DE52) chromatography and SEC purification steps. Throughout the purification process, the relative intensities of the B800-850 Bchl-*a* molecules remain the same, whilst the absorbance at 280 nm, where aromatic residues absorb, decreases as protein contaminants are removed. Once the LH2 complexes were pure, they were concentrated to an $OD_{850\ nm}$ of 100 cm^{-1}. Crystallisation experiments were set up according to section 2.6 of Materials and Methods and McDermott *et al.,* 1995 [43].

LH2 crystals grew within 3 weeks, and were generally 300 μm in length. An example of an LH2 crystal is shown in Figure 5-6. These crystals were harvested and cryo-protected using saturated sucrose solutions. As the crystallisation conditions equilibrated from pH 9.35 of the reservoir solution and pH 10.3 of the protein solution, the cryo-protectant solutions also had to be equilibrated against an artificial reservoir solution for 48 hours before they could be used. Cryo-protection of these crystals was achieved by slowly increasing the concentration of sucrose from 0% to 50% of the saturated sucrose solution in the crystallisation conditions, to avoid any change in osmotic pressure that may cause the crystal to crack or dissolve.

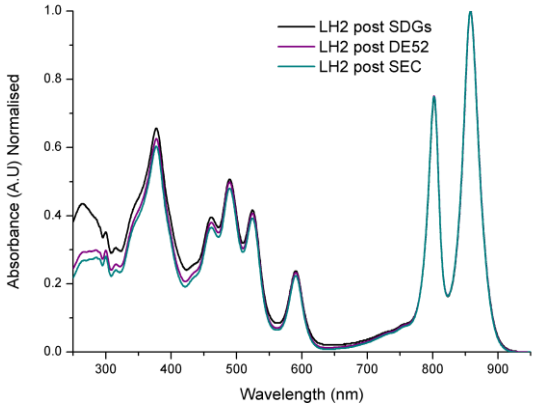


Figure 5-5 Absorbance spectrum of the LH2 complexes during the purification procedure. In Black, the absorbance spectra from LH2 complexes isolated from SDGs is shown. As this is a relatively crude separation of the LH2 from other membrane components on the basis of density, there are some protein contaminants. This explains the higher absorbance peak at 280 nm. The purple line shows the absorbance spectrum after anion exchange chromatography on a DE52 column. The cyan line shows the absorbance spectrum after SEC. It is clear that during the purification process, the intensity of absorbance at 280 nm decreases, whilst the intensity of absorbance at 800 and 850 nm, the Bchl Q_Y peaks, remains the same. This indicates that the LH2 has been specifically purified and the protein contaminants have been reduced.

Figure 5-6 An example crystal of the B800-850 LH2 complex from *Rps*. acidophila strain 10050. The protein ($OD_{850\ nm}$ = 100 cm^{-1}) was in solution with 20 mM Tris-HCl pH 8.0, 350 mM NaCl, 3.5 % (w/v) benzamidine-HCl, 0.1 % (v/v) LDAO and 1 M K_2HPO_4. The protein drop (15 µl) was surrounded by 1 ml of reservoir solution, 2.1 M Ammonium sulphate pH 9.35. Crystals were fully grown by 3 weeks and had an average length of 300 µm.

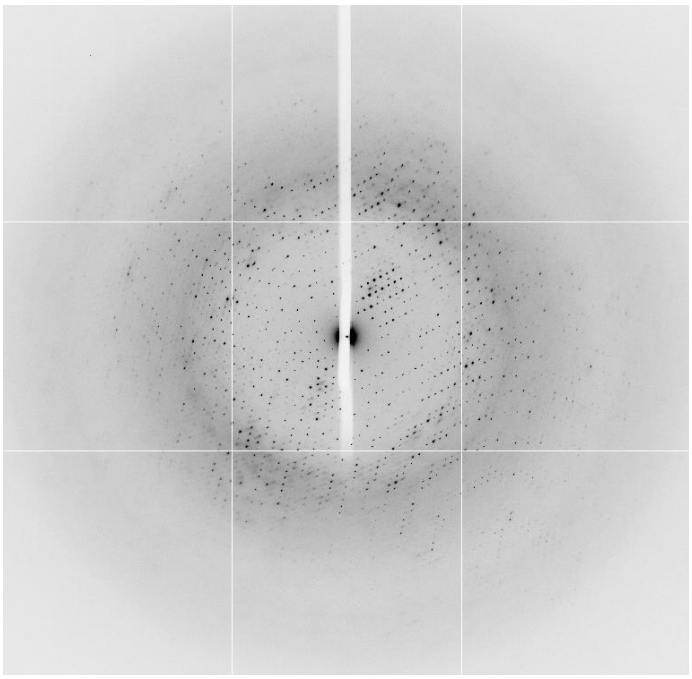


Figure 5-7 Example diffraction of the B800-850 LH2 complex from *Rps. acidophila* strain 10050 crystals tested at DLS beamline I04. The diffraction shows low anisotropy, as the diffraction extends well in all directions. The spacing between the lunes indicates low mosaic spread. The data were processed to a resolution of 2.05 Å.

The crystals were tested at DLS on beamline I04 and a data set was collected and processed to a final resolution of 2.05 Å in the hexagonal space group H32 with unit cell dimensions: *a* = 115.82 *b* = 152.82 *c* = 297.89 α = β = 90 γ = 120°. An example of the diffraction data obtained is shown in Figure 5-7. The quality of the diffraction was very good, with a low mosaic spread, indicated by the spacing between the lunes, and low anisotropy as diffraction extended well in all directions. The resolution range for the data collected was 2.05 to 59.78 Å. As the structure of the B800-850 LH2 complex from the same species of bacteria had already been solved, MR using the program PHASER [132] to gain phase information was completed. This produced a very good solution with the figure of merit: RFZ = 13.9, TFZ = 70.7 and LLG = 9021. Such good z-scores were expected as the search model used was the same protein.

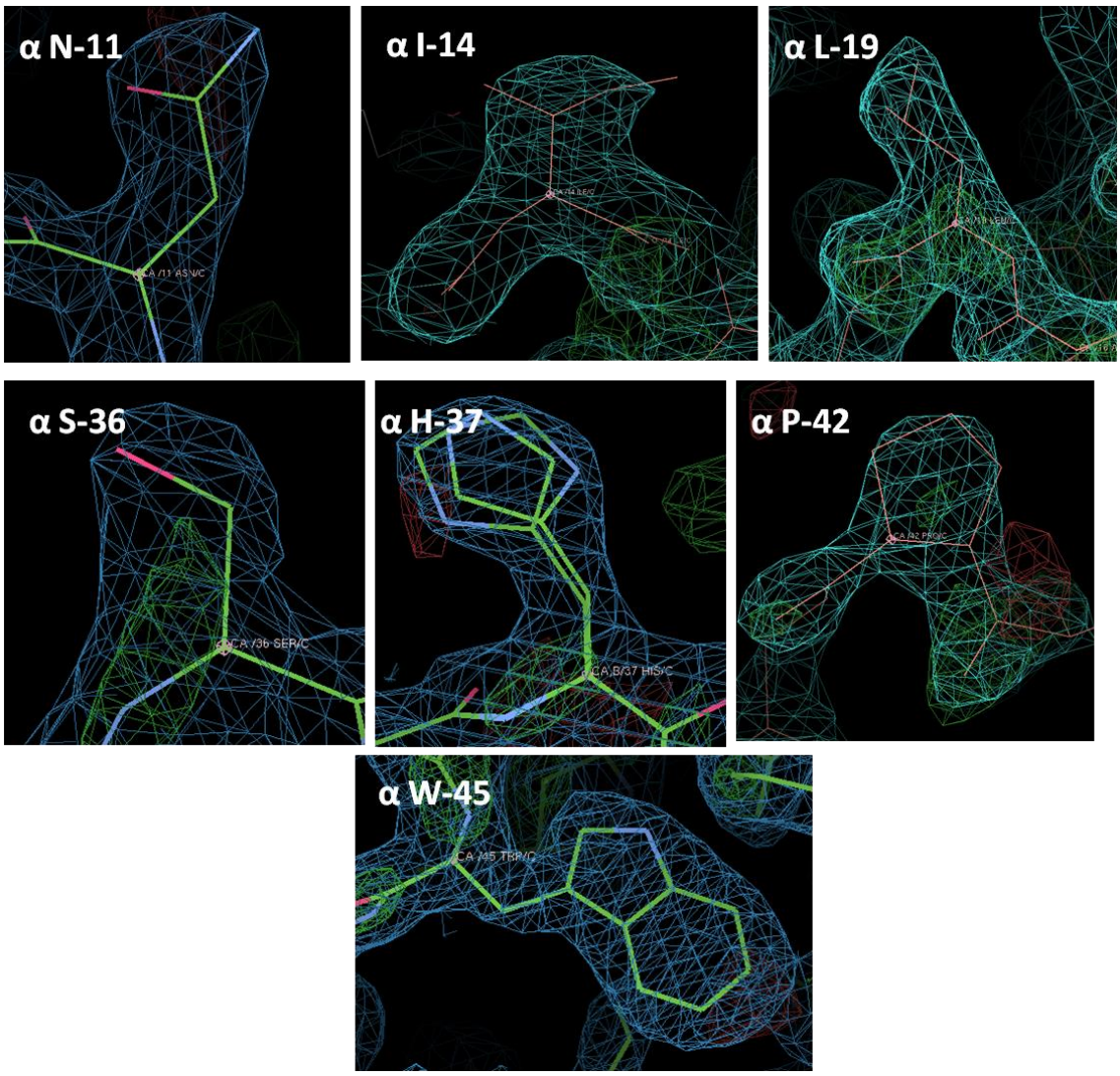


Figure 5-8 Electron density maps of the B800-850 LH2 complex solved from *Rps. acidopila* strain 10050. The residues highlighted from the α-apoprotein are the positions where non-conservative residue changes were observed from the available sequence data. It is clear from these maps that the electron density is a perfect fit for the residues present, which, suggests that there are still some *pucBA* genes that are currently undiscovered in the *Rps. acidophila* strain 10050 genome. This shows that the sequence data from the NGS Illumina technique is not completed, and it is unknown how much of the genome is currently un-sequenced.

Figure 5-8 shows the locations where the non-conservative changes in the α-apoprotein should be located. It is clear from these images that the electron density map fits the residues perfectly with respect to the sequence used previously in 1993, hence does not match any of the protein products from the *pucA* genes sequenced. This suggests that, either there are more *pucBA* gene pairs to be located within the genome sequence, or *pucBA6* is transcribed and the conservative changes can not been seen in the structure at this resolution.

5.3. puf operon

As discussed previously in Chapter 4 and section 1.3.2.1 of the Introduction, the *Rhodobacter* genera of α-proteobacteria contain an additional gene in the *puf* operon, called *pufX*. The gene product appears to be essential in the efficient cyclic electron transfer in the core complex [68,72]. It is hypothesised that PufX interrupts the LH1 ellipse, creating a gap allowing the transfer of electrons to occur [18,72]. No orthologous gene has been identified in any other genera of α-proteobacteria. Analysis of the *puf* operon in the *Rps. acidophila* strain 10050 genome, did not identify any genes that appeared orthologous to *pufX.* A PufX-like protein, Protein-W, found in *Rps. palustris* has potentially been identified (See chapter 4). A search for any similar proteins present in *Rps. acidophila* strain 10050 was attempted, however no similar proteins were identified. Given the low sequence homology between the PufX proteins, this is not -surprising.

In 10050, the *puf* operon was found to be organised of *pufBALMC*. These genes encode the β and α apoproteins of the LH1 complex and the L-, M- and cytochrome-*c* subunits of the RC. The cytochrome-*c* subunit is not universally present in these bacteria and its presence was confirmed by analysis of the purified core complexes using SDS-polyacrylamide gel electrophoresis (see Figure 3.1 of chapter 3). Figure 5-9 is a schematic diagram comparing the *puf* operons from *Rps. acidophila* obtained from this work (A), *Rps. palustris* (B), *Rba. sphaeroides* (C), *Rba. capsulatus* (D) and *Alc. vinosum* (E). There is a clear difference between the organisation of the *puf* operons between the *Rhodobacter* and *Rhodopseudomonas* generas. Both *Rba. sphaeroides* and *Rba. capsulatus* contain the additional *pufQ* and *pufX* genes, whilst *Alc. vinosum* contains multiple *pufBA* genes, 2 of which are located DS from *pufC* [69].

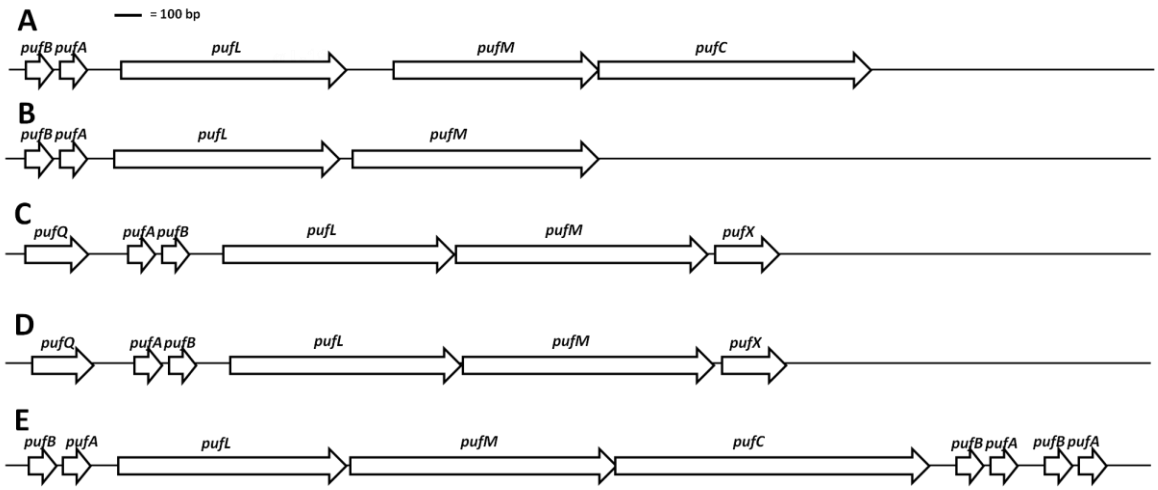


Figure 5-9 A: *Rps. acidophila puf* operon. This species operon is similar to the *puf* operon from *Rps. palustris* (B), the only difference being the presence of *pufC* in *Rps. acidophila*. *Rba. sphaeroides puf* operon (C) and *Rba. capsulatus puf* operon (D) are very similar and both contain *pufQBLAMX*. The *Alc. vinosum puf* operon is the most different and contains multiple *pufBA* gene pairs, 2 of which are located 3' of the *pufC* gene.

5.4. Conclusion

In conclusion, the sequence data obtained for the *Rps. acidophila* strain 10050 genome was incomplete. A mate pair library is being constructed to aid in the full assembly and annotation of the genomic DNA sequencing. Analysis of the available sequence data identified 8 *pucBA* gene pairs. These genes are organised into 2 distinct *puc* operons, the first, previously observed by Simmons *et al.,* 2000 [145] contained *pucBA4-8* and *pucC*. This operon is composed of *pucBA* gene pairs that encode the B800-850 LH2 complexes. The second operon contains *pucBA3-1* and no *pucC.* These gene pairs encode the B800-820 LH2 complexes only. It is possible that the lack of *pucC* in this operon means that these genes are not transcribed. This would explain the lack of expression of the B800-820 LH2 complexes under all laboratory growth conditions tested. It would be interesting to insert a *pucC* gene into this operon to see if the B800-820 LH2 complexes could be expressed, providing an unique opportunity to compare a transcribed operon to one that isn't. This may allow the understanding of the regulation and transcription of these complexes to be improved. Sequence analysis of the protein products failed to identify the *pufBA* gene pairs that encode

the B800-850 LH2 complex previously sequenced by Zuber and Brunisholz. To test if *Rps. acidophila* has mutated over time under laboratory conditions, as has previously been seen with the *Blc. viridis* RC [149], the LH2 complexes were purified and crystallised. It was hypothesised that analysis of the electron density data from a crystal structure would reveal if any of the non-conservative changes, such as N11D, L19F, I14F, S36N, H37N and P42K, had actually occurred in the expressed LH2 complex. A 2.05 Å data set was collected from DLS beamline I04. The solution showed no indication of any observable differences in the protein sequence used to solve the original 1993 McDermott LH2 structure and the sequence required to solve the crystal structure detailed in this Chapter. Two possibilities exist. Firstly that *pucBA6* encodes the B800-850 nm LH2 complex, but the conservative changes were not observable at 2.05 Å. Secondly, that there are still some *pucBA* gene pairs that have not been identified.

The *puf* operon for this species was annotated and found to contain *pufBALMC* architecture. There were no genes orthologous to *pufX* or RPA0067 (Protein W from *Rps. palustris*) located either within the *puf* operon or the available sequence data.

6. 2-dimensional electronic spectroscopy of carotenoids

When carotenoids in LH2 complexes are excited by light the energy is rapidly and efficiently transferred to the Bchl molecules as discussed in section 1.4.1 of the introduction. Carotenoids (Cars) are molecules with extended linear conjugated π-electron systems that can differ in conjugated carbon length (*N*) [87]. It has been hypothesised that in Cars with $N \geq 9$, intermediate electronic states could exist between the well-known S_2 and S_1 excited states [88]. Although the energy transfer reactions involving Car and Bchl molecules have been subject to intense investigation, the exact pathways (the exact electronic excited states of the Car involved) are still unclear [89-91,96]. The aim of this chapter was to use 2DES to look for the presence of these predicted intermediate states in Car molecules with $N = >9$ during Car to Bchl energy transfer in the LH2 complexes. Experiments were designed and carried out in collaboration with the Scholes group, University of Toronto, Canada, to try to observe the coupling between the carotenoid S_2 state and the Bchl Q_X state. The specific aim was to directly observe any intermediate carotenoid states that lie between the S_2 and Q_X/S_1 states [123].

6.1. Purification of LH2 complexes and carotenoids

6.1.1. LH2 complexes

LH2 complexes were purified from cells of *Rps. acidophila* and *Rba. sphaeroides* grown at 30°C under 60 µmol m^{-2} s^{-1} illumination in anaerobic conditions, as described in section 2.1 of the Materials & Methods. The purified complexes were diluted to a final concentration of OD_{850nm}= 0.3 mm^{-1} in the purification buffer.

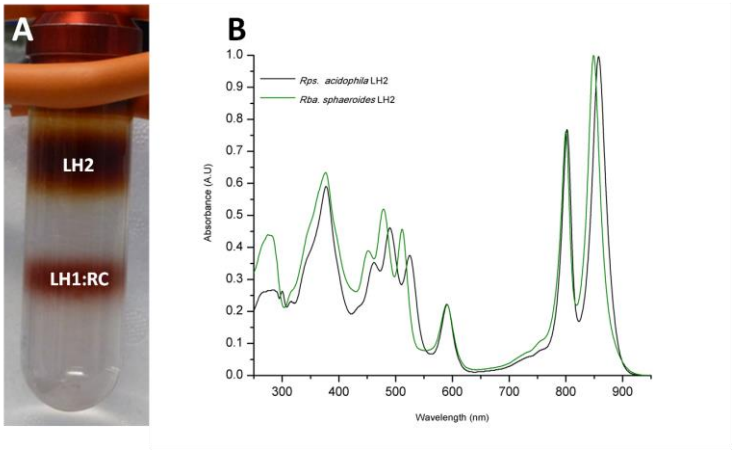


Figure 6-1 A - SDG of *Rps. acidophila* membranes. The top band contains the LH2 complexes. This band was carefully removed and purified by anion exchange and size exclusion chromatography. B - Comparison of the absorbance spectra of purified LH2 complexes from *Rps. acidophila* (black line) and *Rba. sphaeroides* (green line). The spectra highlight slight differences in the positions of the B850 maxima and the carotenoids. Spectra were normalised at the B850 maxima.

6.1.2. Purification of rhodopin glucoside and spheroidene

Carotenoids were isolated from membranes of *Rps. acidophila* and *Rba. sphaeroides* grown at 30°C under 60 µmol m^{-2} s^{-1} illumination in anaerobic conditions. The carotenoids were extracted and purified using organic solvents as described in section 2.10 of the Materials & Methods. The carotenoids were analysed by absorbance spectroscopy to ensure they were pure (Figure 6-2) and dissolved so that the maximum absorbance peak had an OD of 1 mm^{-1}.

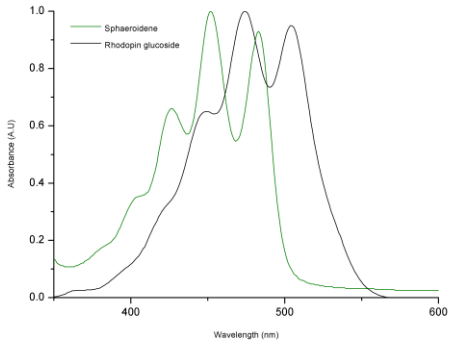


Figure 6-2 Absorbance spectra of spheroidene (green line) and rhodopin glucoside (black line) dissolved in acetone. The spectra were normalised to 1 A.U. at their absorbance maxima. The difference in absorption is due to different π-conjugation lengths (*N*), spheroidene *N* = 10, rhodopin-glucoside *N* = 11.

6.2. 2DES

Purified LH2 complexes from *Rps. acidophila* strain 10050 and *Rba. sphaeroides* strain 2.4.1 were selected for these experiments due to the small energy gap between the Car-S_2 state (490-530 nm) and Bchl-Q_X state (590 nm) as shown in Figure 6-3. This allowed the transition between the S_2 and Q_X states to be probed simultaneously using broadband laser pulses (50-60 THz) as described in section 2.11 of the Materials & Methods. The τ_1 (coherence time) was Fourier transformed to provide the excitation frequency scale, whilst T (waiting time) was Fourier transformed to provide information on the temporal evolution of the spectral response. The energy levels of the Car (red) and Bchl (green) molecules are shown in Figure 6-3 alongside the absorption spectrum of the LH2 complex from *Rps. acidophila*. The energy levels depicted by dashed lines are 'dark' states - they cannot be directly populated from the ground state in a 1 photon process as it is symmetry forbidden. The solid lines represent the populated symmetry allowed energy states.

LH2 complexes are structurally very well defined and have been the subject of extensive transient absorption and stationary absorption studies. As such, most of the peaks from the experimental data obtained in this study could be assigned, as their wavelengths and lifetimes were known from previous experimental work.

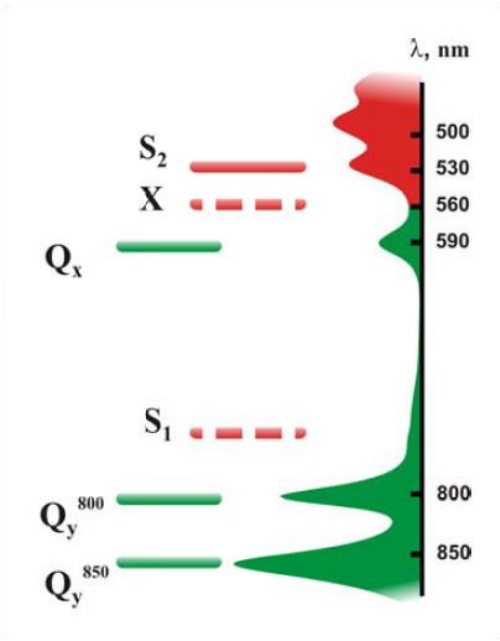


Figure 6-3 Absorption spectrum and scheme of corresponding electronic levels of LH2 complexes of *Rps. acidophila*. BChl-*a* (B800-850) is shown in green, carotenoids in red. Dashed lines correspond to the excited states inaccessible by direct optical excitation from the ground state, whilst solid lines correspond to optically allowed 'bright' states.

6.2.1. LH2 complexes

Figure 6-4 A & B shows the absorptive 2D spectra of *Rba. sphaeroides* and *Rps. acidophila* at t_2 = 215 and 158 fs, respectively. Figure 6-4 C & D show the position of the excitation pulses with respect to the S_2 and Q_X absorption peaks. 2DES, a relatively new technique, was used as it has increased time and frequency resolution and coupling between states can be sensitively detected. If the maxima of the S_2 or Q_X peaks are excited by the laser pulse, the resultant spectra would be dominated by these symmetry allowed states, masking the 'dark' states or indeed any intermediate states present. To reduce the possibility of this occurring, only the very red edge of the S_2 and blue edge of Q_X states were excited in these experiments.

Figure 6-4 A & B show 3 well resolved ground state bleach (GSB) peaks positioned along the diagonal. The S_2 and Q_X peaks were assigned in accordance with the stationary absorbance spectra at 530 and 580 nm, respectively. The 3rd peak at 560 nm has been labeled 'X'. It is a clear

peak in both spectra and cannot be assigned to either the S_2 or Q_X states.

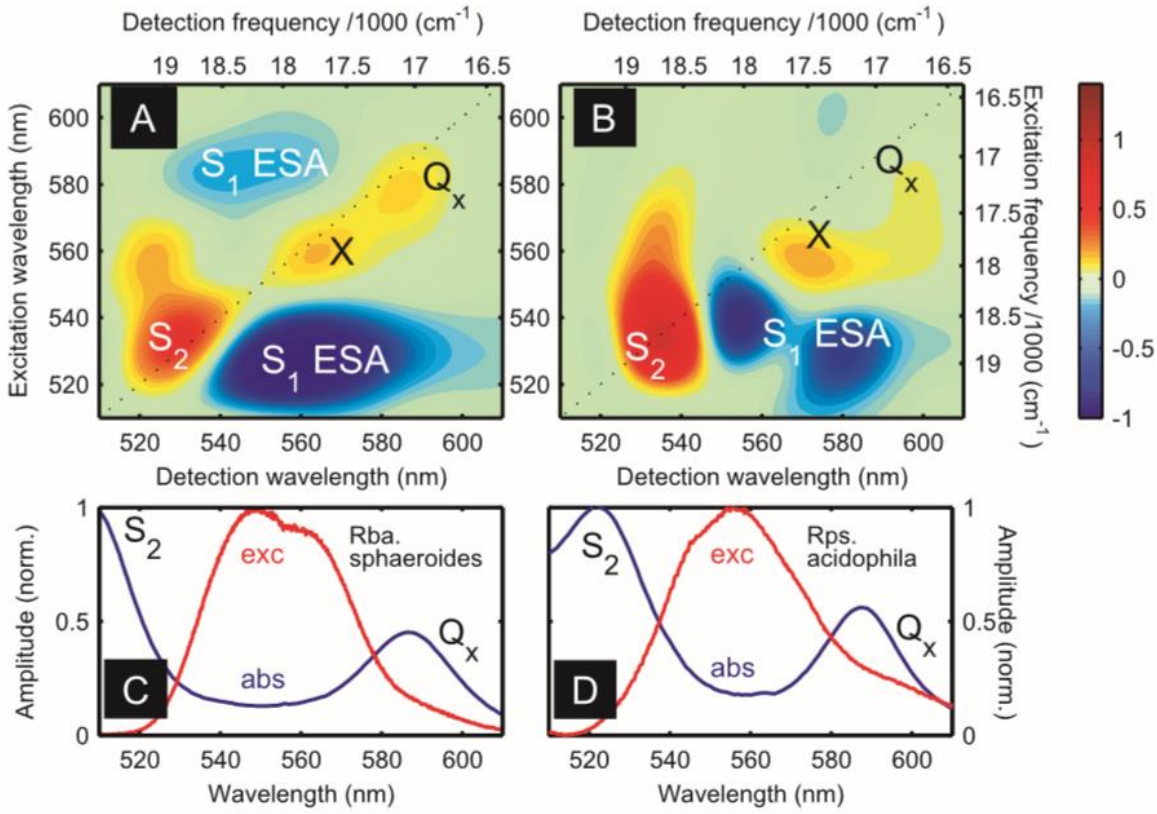


Figure 6-4 Absorptive 2D spectra of (A) *Rba. sphaeroides* taken at t_2=215 fs, and (B) *Rps. acidophila* taken at t_2 = 158 fs. Spectra of *Rba. sphaeroides* were measured in 0-400 fs range with 5 fs steps, spectra of *Rps. acidophila* were measured in 0-200 fs range with 1 fs steps. Absorption spectra (blue line) and excitation pulse spectra of (C) *Rba. sphaeroides* and (D) *Rps. acidophila*.

Rhodopin-glucoside (RG), the Car in *Rps. acidophila* LH2, contains 11 conjugated π electrons (N = 11) whilst spheroidene, the Car in *Rba. sphaeroides* LH2, has N = 10. The difference in conjugation length shifts the S_2 energy state of RG to the red relative to the S_2 state in spheroidene. Therefore, the 2D spectra from *Rps. acidophila* has a more dominant S_2 GSB with respect to *Rba. sphaeroides,* as more of the S_2 peak was excited (Figure 6-4 B) which masked some of the less favoured states. Both spectra exhibit clear negative excited state absorption (ESA) peaks that reflect the internal conversion (IC) of S_2 to S_1 within the Car. This transition is favoured and well documented in transient absorption studies [57]. Global target analysis was required to

assign the double ESA peak in Figure 6-4 B and the ESA above the diagonal in Figure 6-4 A. The ESA above the diagonal in Figure 6-4 A is a cross peak between the Car S_1 and Bchl Q_X GSB. Cross peaks in 2D spectra are very important as they contain information about ET and coupling between electronic states, hence, contain information on the history, or evolution, of the states. It was therefore hypothesised that this peak could be due to ET from the Bchl Q_X to the Car S_1 state, or IC within the Bchl, from the Q_X to the Q_Y state, but in both cases are populated by the Q_X state. This peak was a very weak feature and no analogous peak was observed in *Rps. acidophila* LH2 complexes at different t_2 delays. This is probably due to the overlap of Q_X/S_2 and Q_X/X cross-peaks.

The X peak and cross peaks at 560 nm have never been observed before, and as such the molecule responsible for producing this peak was unknown. To distinguish if this peak was associated with the Cars or Bchl molecules the t_2 data on the diagonal of the 2D spectra was Fourier transformed to provide information on the high frequency vibrations contained within each GSB feature that was observed. As Car molecules are extended conjugated carbon molecules they contain signature high frequency oscillations of the ground state; CH_3 rocking (990 cm^{-1}), C-C stretch (1195 cm^{-1}), and C=C stretch (1590 cm^{-1}). Whilst Bchl molecules have no known pronounced ground state vibrational modes in the high frequency region. Comparisons of the Fourier frequencies of S_2, X and Q_X states are shown in Figure 6-5 and Figure 6-6. The S_2 and X are almost identical and both contain all three vibrational modes characteristic of carotenoid molecules, whilst the Q_X only shows one high frequency state at approximately 1600 cm^{-1}. It is most likely that this is due to an overlap from the carotenoid in the LH2 complex as Bchl molecules contain no high-frequency vibrational states. Therefore, the X peak is associated with the carotenoids.

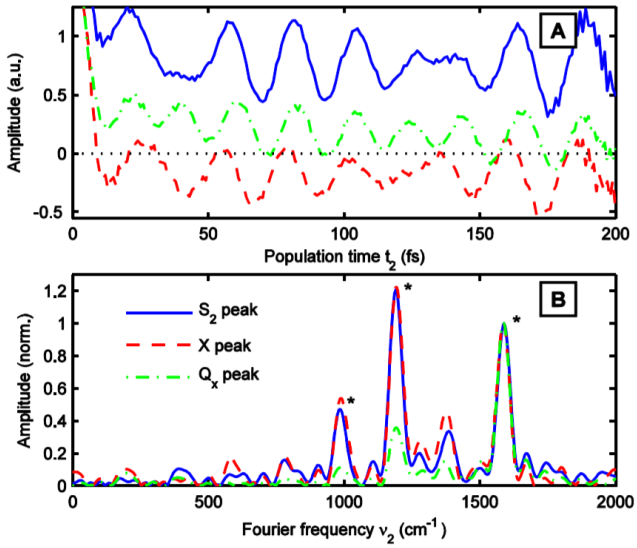


Figure 6-5 (A) Time traces of *Rps. acidophila* (measured at 0-200 fs, 1 fs steps). (B) Fourier spectra of *Rps. acidophila* corresponding to time traces in (A). Blue solid line shows the time trace at S_2 diagonal peak ($\lambda_{exc}=\lambda_{em}$=530 nm), red dashed line - the time trace at X diagonal peak ($\lambda_{exc}=\lambda_{em}$=560 nm), green dash-dotted line - the time trace at Q_x diagonal peak ($\lambda_{exc}=\lambda_{em}$=585 nm). For peak positions see Figure 6-4B. The Fourier spectra were normalised to the amplitude at 1590 cm^{-1}. The time traces represent raw data. Prior to Fourier analysis the exponentially decaying term was extracted from these time traces.

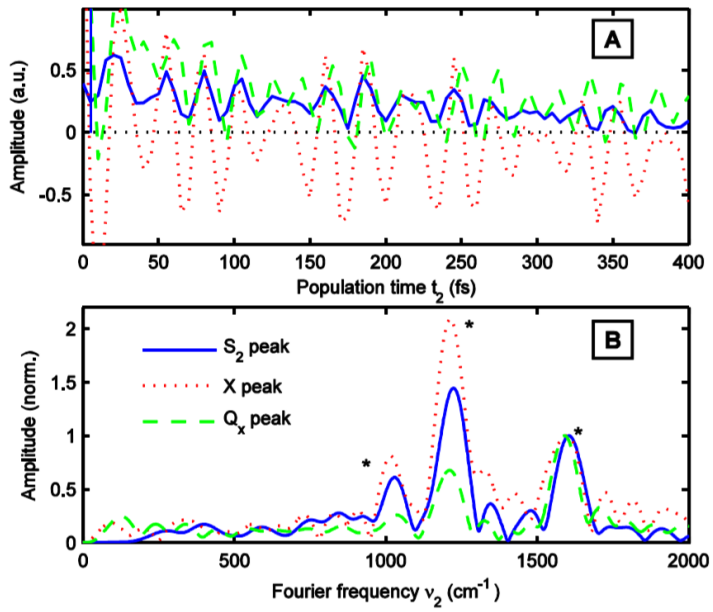


Figure 6-6 (A) Time traces of *Rba. sphaeroides* (measured at *t2* = 0-400 fs, 5 fs steps). (B) Fourier spectra of *Rba. sphaeroides* corresponding to time traces in (A). Blue solid line shows the time trace at the S2 diagonal peak, red dotted line shows the time trace at the X diagonal peak, green dashed line shows the time trace at the QX diagonal peak. For peak positions see Figure 6-4 A. The Fourier spectra are normalised to the amplitude at 1590 cm^{-1}. The time traces represent raw data. Prior to Fourier analysis the exponentially decaying term was extracted from these time traces.

6.2.2. Free carotenoids

The intermediate 'X' peak appeared to be associated with the carotenoids within the LH2 complexes. To confirm this, the isolated carotenoids, RG and spheroidene, were analysed in the same way as the LH2 complexes. The 2D spectra obtained are shown in Figure 6-7 A-B and C,D, respectively.

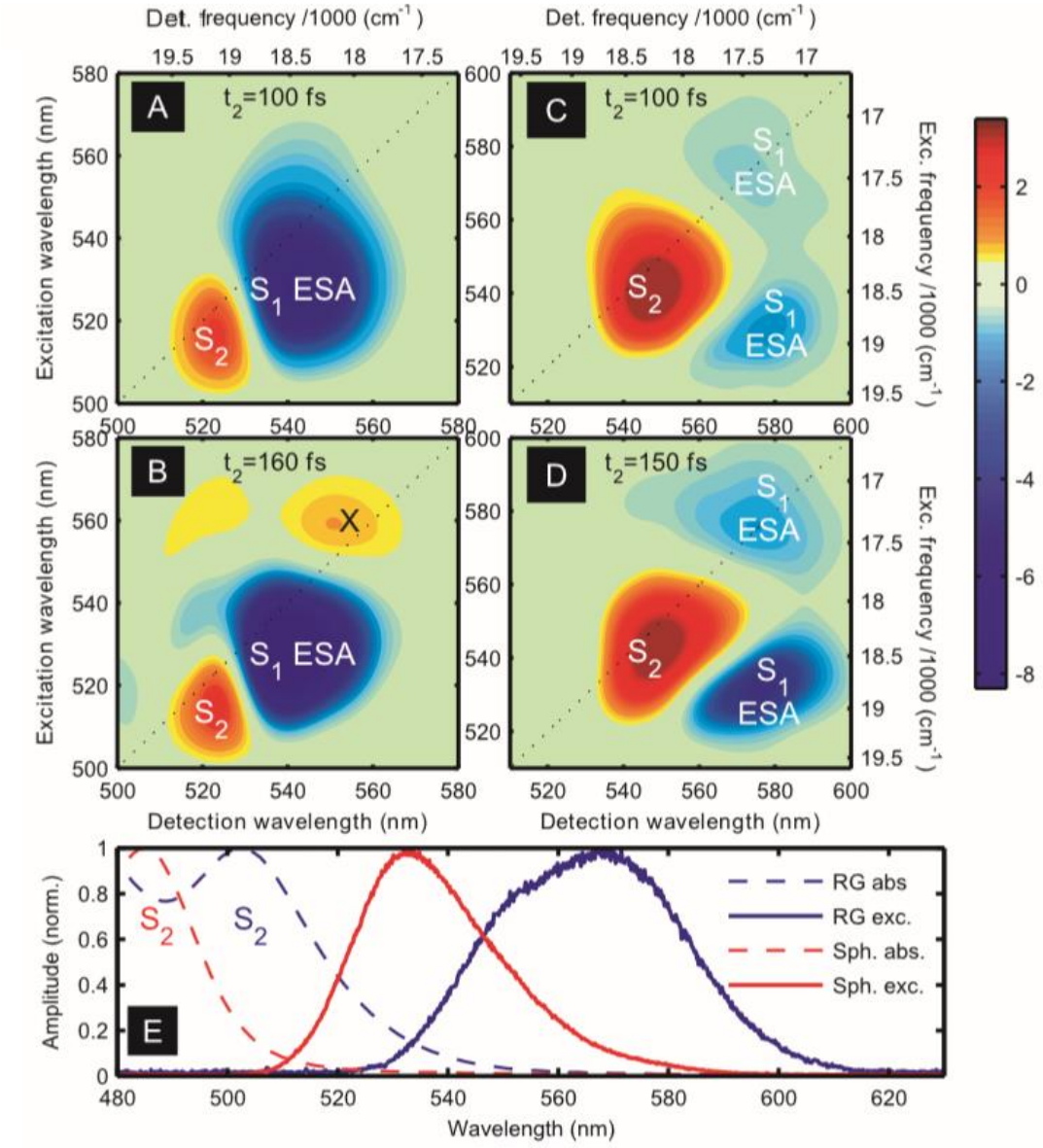


Figure 6-7 Absorptive 2D spectra of spheroidene (A-B) measured in 0-400 fs range with 2 fs step and rhodopin glucoside (C-D) measured in 0-400 fs range with 5 fs steps. Carotenoids were dissolved in acetone. The t_2 times of each spectrum are shown at the top. Absorption spectra/excitation pulse spectra (E)of spheroidene (red dashed/solid lines), and rhodopin glucoside (blue dashed/solid lines).

Figure 6-7 E shows the excitation pulse with respect to the absorbance peaks of both carotenoids. The laser set-up, built by the Scholes group, was not stable at the wavelengths required to excite RG at the appropriate wavelength to reduce the contribution of the S_2 GSB in the 2D spectra, however, is currently being re-built for further studies. The 2D spectra of RG were therefore dominated by the S_2 GSB and S_1 ESA, as shown in Figure 6-7 C and D, at all t_2 variables. The S_1 ESA, however, is broad. The cross peak below the diagonal appears to elongate towards 560 nm, potentially indicating ET from the X state to the S_1. The 2D spectra of spheroidene at different t_2 variables showed a weak peak at 560 nm corresponding to the X state. This peak was only observed at certain t_2 delays indicating that it was a short-lived species.

6.3. Global target analysis

To tease out information regarding the electronic structure of these states, and the evolution of the species, global target analysis was applied. This allowed us to observe how the spectral features evolved globally, as despite the complicated 2D spectra, there were enough clearly defined transitions (Figure 6-8 and Figure 6-9) to allow global target analysis to fit the data and reveal a kinetic model. The cross peaks acted as constraints to the analysis as they contain information on the transitions and history of the spectral features observed.

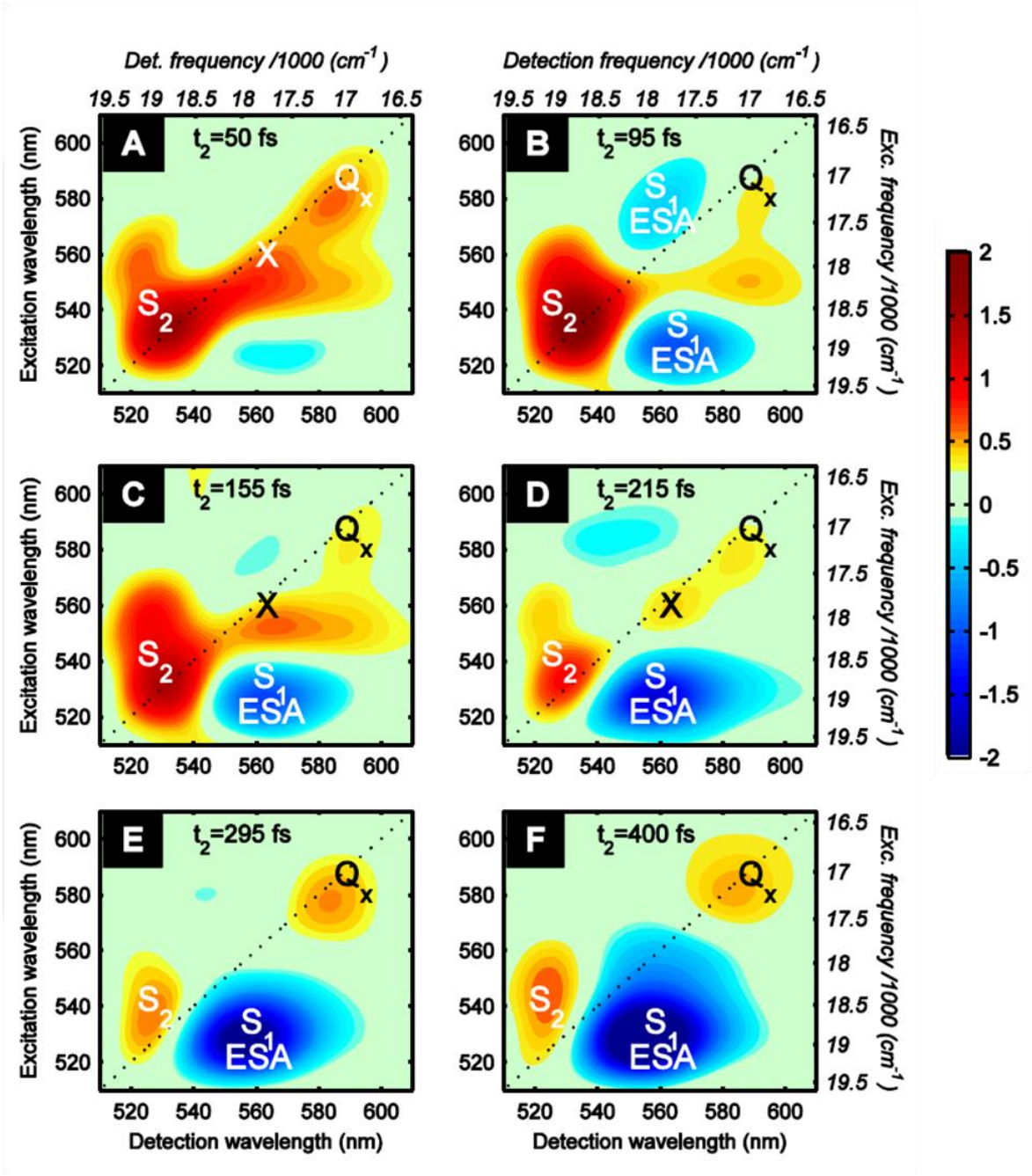


Figure 6-8 2D spectra of *Rba. sphaeroides* measured in 0-400 fs range with 5 fs steps. The t_2 times of each spectrum are shown at the top. These data are not a replica of the data plotted in Figure 6-4, but represent a separate measurement taken at identical excitation conditions.

Figure 6-8 and Figure 6-9 show 2D spectra at different t_2 delays and clearly demonstrate that the X peak is a short-lived species (approx. 50 fs in both cases) as over the time-course the X peak appears and disappears. There are also cross peaks that suggest that energy transfer between the X, S_1 and Q_X is occurring. The X peak is less well-resolved in *Rps. acidophila* (Figure 6-9) than it is in *Rba. sphaeroides* (Figure 6-8) as the S_2 GSB dominates the spectra.

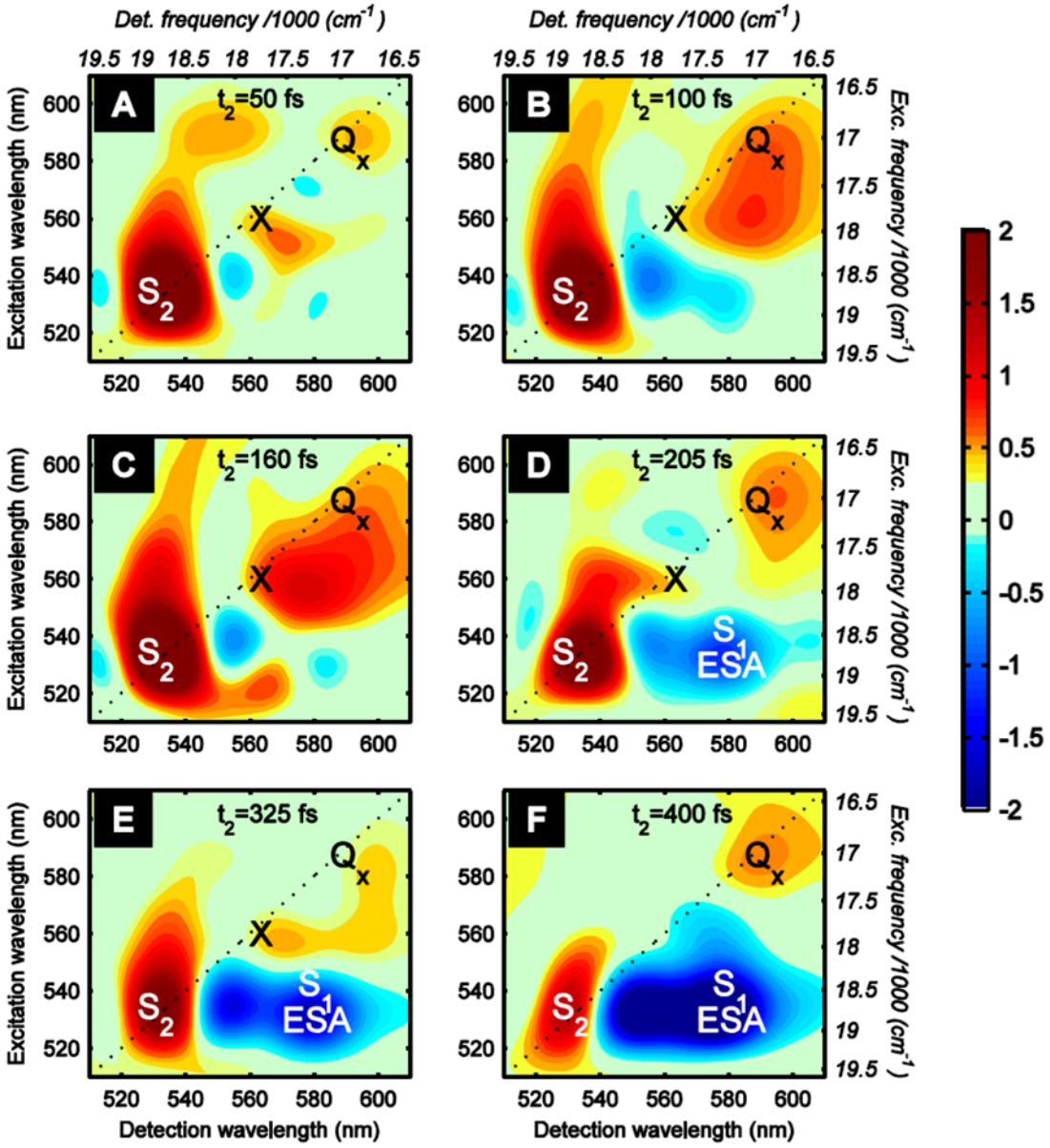


Figure 6-9 2D spectra of *Rps. acidophila* measured in 0-400 fs range with 5 fs steps. The t_2 times of each spectrum are shown in the top. These data are not a replica of the data plotted in Figure 6-4, but represent a separate measurement taken at identical excitation conditions.

Global target analysis was complicated by the spectra being composed of coherent oscillations and exponential incoherent relaxation pathways. Dr Evgeny Ostroumov designed a Matlab code that separated the incoherent and coherent signals and treated them globally. Information on the evolution of species and the kinetics of the system were obtained by analysing the incoherent decays. Kinetic schemes were used to model the incoherent decay populations. These models provided a set of time dependent populations from the kinetic model chosen, using

kinetic rates as global variables. Multiple kinetic models were tested to find the model that provided the most consistent data. These populations were then convoluted with the instrument response function.

Figure 6-10 and Figure 6-11 show the 2D evolutionary-associated spectra (2DEAS) of *Rps. acidophila* and *Rba. sphaeroides,* respectively.

6.3.1. Evolutionary associated spectra of *Rps. acidophila*

The 2DEAS presented, outline a sequential kinetic scheme for the Car and Bchl molecules of LH2 complexes from *Rps. acidophila* (Figure 6-10). These spectra allow us to determine the main features and the underlying dynamics of the system. There are 4 main features in all 2DEAS shown; S_2, S_1, X and Q_X states. The decay rate constant of 2DEAS-1 (Figure 6-10) is 48 ps^{-1}. The main features are the positive GSB/SE peak of the S_2 state at 530 nm and the positive cross-peak between 530 nm and 560 nm. This is most likely due to EET from S_2 to the X state within the Car molecule. The X and Q_X peaks on the diagonal are very weak and not readily observable (Figure 6-10) however, another cross-peak above the diagonal between 580 nm and 560 nm suggests that EET between X and Q_X has occurred.

Figure 6-10 2DEAS-2 was shown to have a decay rate constant of 16 ps^{-1}. The two positive peaks on the diagonal are S_2 and Q_X GSB at 530 nm and 589 nm respectively. There are three cross-peaks below the diagonal and one cross-peak above the diagonal. The positive cross-peaks below the diagonal show EET from the Car S_2 and X states to the Q_X of the Bchl molecule. The third cross-peak below the diagonal represents ESA of the S_2 by the S_1 state within the Car molecule. The cross-peak above the diagonal represents IC within the Car from the X state to the lower lying S_1 state.

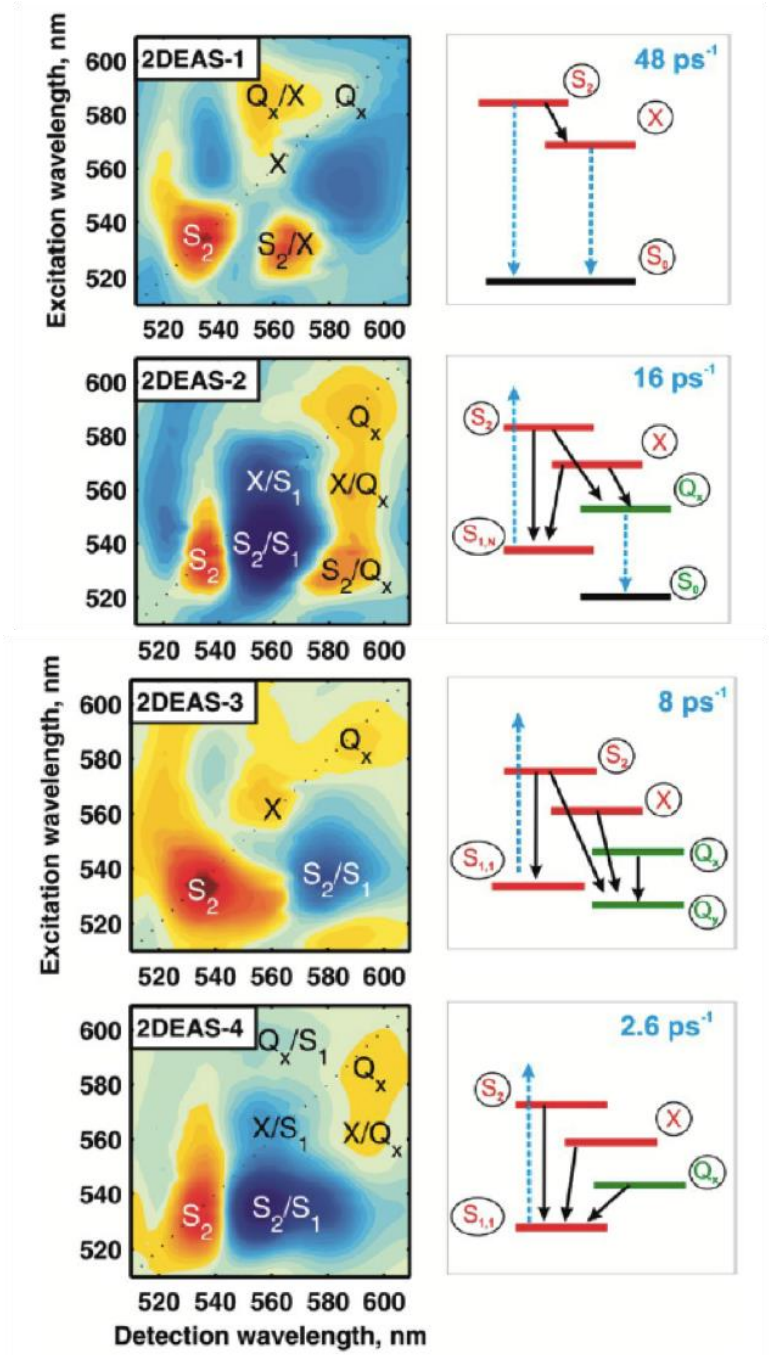


Figure 6-10 2D evolutionary-associated spectra (2DEAS) of *Rps. acidophila* of absorptive signal. Diagonal features are designated by a single symbol, off-diagonal features are marked by two symbols (indicating the locally excited state, and the emitting state, respectively).

Figure 6-10 2DEAS-3 has a decay rate constant of 8 ps^{-1}. There are three positive peaks on the diagonal, S_2 (530 nm), X (560 nm) and Q_X (580 nm) GSB and one negative cross-peak between 580 nm and 540 nm. This is the ESA and SE of the S_2 state by the S_1 state within the carotenoid.

Figure 6-10 2DEAS-4 has a decay rate constant of 2.6 ps^{-1}. There are two positive GSB peaks on the diagonal, S_2 and Q_X. The X GSB peak is obscured by the ESA of the X state by the S_1 state within the Car molecule. There are two further negative ESA cross-peaks, one above the diagonal (Q_X to S_1), and one below the diagonal (S_2 to S_1). There is also a positive GSB cross-peak between the X state of the Car and the Q_X of the Bchl molecule.

6.3.2. Evolutionary associated spectra of *Rba. sphaeroides*

The 2DEAS for *Rba. sphaeroides* reveal similar features to *Rps. acidophila* discussed in section 6.3.1. Figure 6-11 2DEAS-1 shows three positive GSB peaks belonging to S_2, S and Q_X. There are two positive cross-peaks, one above and one below the diagonal. The peaks both lie between 560 and 530 nm and show EET from the S_2 to the X state within the Car. The decay rate constant for this spectra is 30 ps^{-1}.

Figure 6-11 2DEAS-2 shows a prominent GSB peak at 580 nm on the diagonal belonging to the Q_X state of Bchl. There is a weak peak showing the GSB of the S_2 state, whilst the X state GSB is masked by ESA of the X state by the S_1 state within the Car. There are two cross-peaks, one below and one above the diagonal, which are assigned to the ESA of S_2 by S_1 and the ESA of Q_X by S_1 states, respectively. These species had a decay rate constant of 14 ps^{-1}.

Figure 6-11 2DEAS-3 shows the reappearance of a dominant S_2 GSB diagonal peak. This peak is quite broad and overlaps with the X state GSB. There is a small peak representing the Q_X GSB on the diagonal. There are three cross-peaks which were assigned to ESA of Q_X and S2 states by the S1 state, and EET from X to Q_X. The decay rate constant was 3 ps^{-1}.

Finally, Figure 6-11 2DEAS-4 is dominated by an ESA peak. This peak is very broad and spans from 520-560 nm and therefore is hypothesised to be the GSB of the S_2 and X states by the S_1 state. There is also a clear GSB peak of the Q_X state at 580 nm.

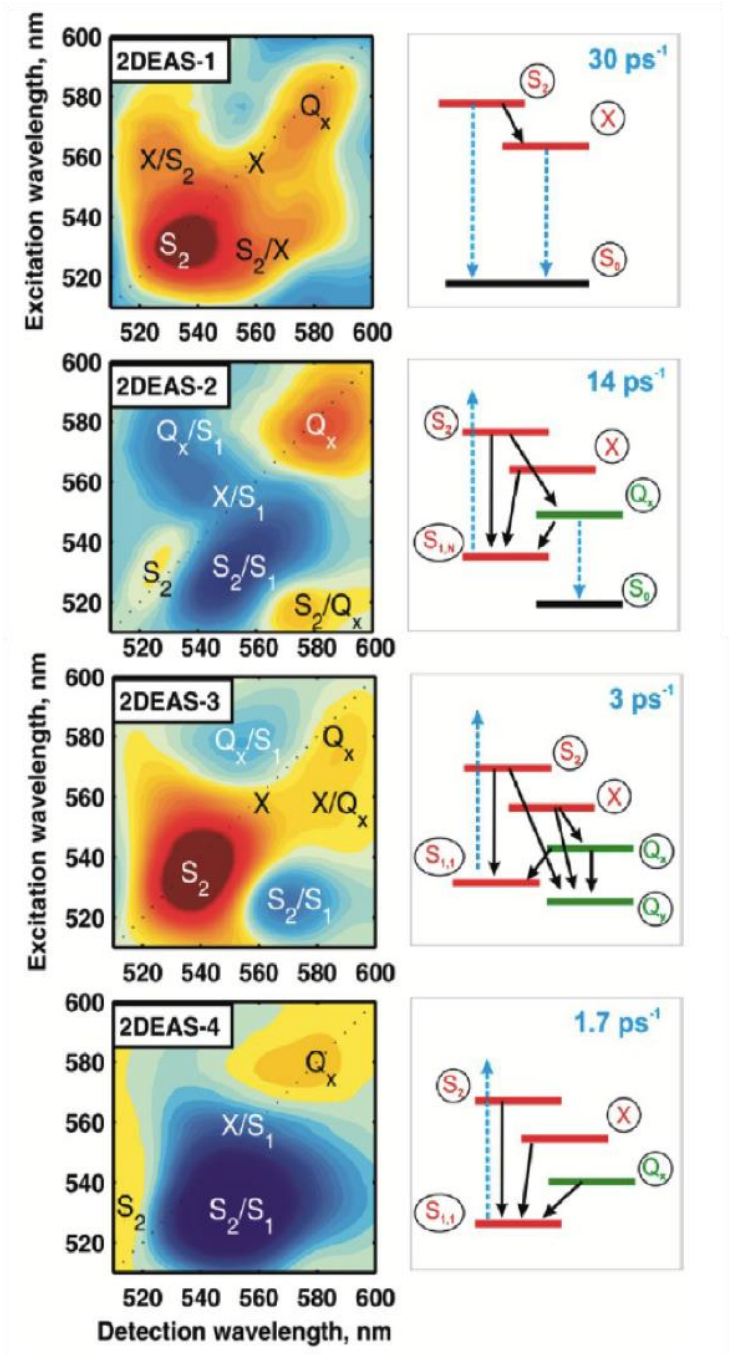


Figure 6-11 2D evolutionary-associated spectra (2DEAS) of *Rba. sphaeroides* of absorptive signal. Diagonal features are designated by a single symbol, off-diagonal features are marked by two symbols (indicating the locally excited state, and the emitting state, respectively).

6.4. Conclusions

Using broadband 2DES an intermediate carotenoid state has been directly observed for the first time at 560 nm from purified LH2 complexes (*Rps. acidophila* and *Rba. sphaeroides*) and isolated Car molecules, RG and spheroidene. Fourier transform analysis of the t_2 variables demonstrates that the X state contained the signature high frequency ground state vibrations of a Car molecule, and was postulated to be an intermediate energy level between the S_2 and S_1 states. [123]. Using global target analysis it has been postulated that the S_2 state rapidly decays into the X state, which in turn, decays into either the S_1 or Q_X state. This additional intermediate state, first hypothesised by Tavan and Schulten [88] increases the spectral overlap in Car with $N = \geq 9$. This is beneficial as Car with increasing N contain increasing energy gaps between S_2 and S_1 states, whilst the lifetime of S_2 decreases [95]. This should decrease the efficiency of EET between Car and Bchl to <20%, however experimentally determined EET efficiencies have been observed up to 90% for *Rba. sphaeroides* [90,91,123]. The higher than expected EET efficiency rates can be explained by the presence of intermediate states, such as the X state observed here, as they bridge the energy gap between the S_2 and S_1 states, acting like a stepping stone to aid EET from S_2 to Q_X and S_1 states (Figure 6-12).

As the laser could not be tuned to the wavelengths required to analyse the Car molecules with minimal S_2 excitation, the experiments will need to be repeated with a laser that is more stable at lower energies. To aid in these future experiments, the Car molecules will be analysed in the presence of different solvents [150]. The difference in solvent polarities can shift where the molecules absorb maximally. By titrating the S_2 state above and below the intermediate state at 560 nm it should be possible to observe the appearance and disappearance of the X state. The X state can only be occupied from energy relaxing from the S_2

state, and so when the S_2 energy is less than 560 nm, the X state should not be observable.

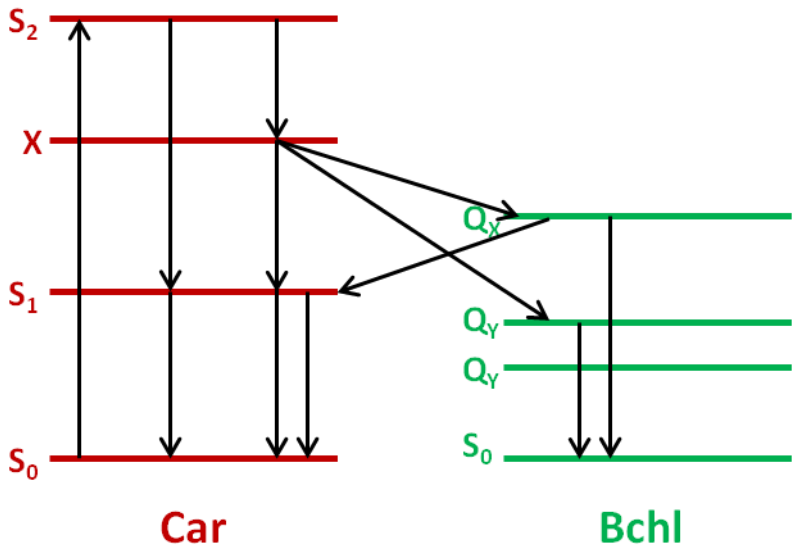


Figure 6-12 Schematic diagram outlining the energy states and possible EET pathways in LH2 complexes via the Car and Bchl molecules. An excited carotenoid promotes an electron from S_0 to the symmetry allowed S_2 state. The lifetime of this state for carotenoid molecules with $N = 11$ is approximately 130 fs. This state rapidly decays in to the X state. From this state energy is either transferred to the Q_X/Q_Y of Bchl molecules or to the S_1 within the carotenoid molecule. From both S_1 and Q_X/Q_Y the states relax to S_0.

Data from this chapter has been published in:

Ostroumov, EE., **Mulvaney RM**., Cogdell RJ., Scholes, GD "Broadband 2D Electronic Spectroscopy Reveals a Carotenoid Dark State in Purple Bacteria" 2013 *Science*. 5 April 2013: 340 (6128), 52-56. [DOI:10.1126/science.1230106]

Ostroumov, EE., **Mulvaney, RM**., Anna, JM., Cogdell, RJ., Scholes, GD., "Energy Transfer Pathways in Light-Harvesting Complexes of Purple Bacteria as Revealed by Global Kinetic Analysis of Two-Dimensional Transient Spectra"*J. Phys. Chem B* **Publication Date (Web):** July 18, 2013 (Article) **[**DOI: 10.1021/jp403028x]

7. Summary and Outlook

In this thesis some of the finer details of the structure and function of the core complex and LH2 complexes from purple photosynthetic bacteria have been analysed.

The main components of the core complexes are encoded by the *puf* operon. In the *Rhodobacter* genera, an additional gene was found at the 3' end of this operon, *pufX*. It was found that PufX was critical for efficient cyclic electron transfer and appeared to create a gap within the LH1 complex [68,72]. It was therefore postulated that this protein organizes the core complex and creates a gap allowing for the hydrophobic Q/QH_2 exchange to occur through the membrane. This gene, however, does not seem to be ubiquitous amongst all the genera of purple photosynthetic bacteria. Historically there has been much debate about whether monomeric core complexes contain an open or closed LH1 ring. Low-resolution models of core complexes from *Rs. rubrum* [58,74] and *Rps. palustris* have shown either closed and open LH1 complexes, respectively. The low-resolution crystal structure of *Rps. palustris* core complex modeled an additional peptide, Protein W, into the LH1 ellipse. This peptide is orientated in such a way that it disrupts the LH1 complex in a manner predicted to be orthologous to PufX. As the resolution of the core complex structure only sufficed to elucidate secondary structural details the identity of Protein W has remained a mystery since its discovery in 2003 [18].

Here, *Rps. palustris* core complexes have been purified to homogeneity. Analysis by SDS-PAGE identified an additional band that contained the hypothetical protein RPA0067 from the *Rps. palustris* genome. A RPA0067KO *Rps. palustris* strain was generated that fails to grow photosynthetically but appears to grow normally chemoheterotrophically. Although this work is still in its infancy it suggests that RPA0067 is Protein W. The growth rates of RPA0067KO and WT strains of *Rps. palustris* need to be compared, both

photosynthetically and chemoheterotrophically. As it has already been reported that PufX$^-$ strains revert, care must be taken when culturing the RPA0067KO strain. Additional experiments to assess the function of RPA0067 would be single molecule spectroscopy. A comparison between the WT and RPA0067KO would inform the experimenter if the absence of RPA0067 closes the LH1 complex. The presence of a gap in the LH1 ring produces a 'fingerprint' narrow line in the excitation spectrum on the low energy side of this spectrum [142]. A complete ring never shows such a narrow line. If the LH1 ring is closed by the loss of RPA0067 it would explain why these mutants cannot grow photosynthetically. But this requires more experimentation.

The highest resolution model of a monomeric core complex is currently 4.8 Å [18] and many groups are currently trying to improve upon this. Here, crystallisation trials of core complexes from *Rps. palustris, Rps. acidophila* and *Alc. vinosum* are presented. The crystallisation trials of purified core complexes from *Alc. vinosum* were reproducible and optimisable. After testing over 6000 different conditions a low-resolution model (5.15 Å) was generated using MR. This model showed that the LH1 complex from *Rps. palustris* and *Alc. vinosum* are not the same. During MR of the core complex from *Alc. vinosum* using the LH1 coordinates from *Rps. palustris* it was observed that there was no Protein W in the LH1 complex. The Protein W helix from the *Rps. palustris* model was moved in-line with the α-apoprotein ring. Once this was completed additional density opposite this helix appeared, suggesting that at this position was a normal LH1 subunit, and therefore *Alc. vinosum* contains a closed LH1 complex. At the PS16 conference in August 2013 the structure of the core complex from *Tch. tepidum* was presented at a resolution of approximately 3 Å. Even though this data set is still undergoing refinement, it was clear that the LH1 complex in the *Tch. tepidum* core complex is closed. It was also suggested that in the absence of a PufX-like protein Q/QH_2 exchange occurs through small gaps between the individual α and β heterodimer

subunits (personal communication with Professor Wang-Otoma). It remains unclear why some species have PufX-like proteins and others do not, and this requires more experimentation.

LH2 complexes have been found to exist as different spectral variants depending upon the conditions of their growth [44]. Different strains of *Rps. acidophila* have been shown to respond differently to the same growth conditions. Whilst strains 7050 and 7750 both express the B800-850 LH2 at HL conditions and B800-820 at LL conditions, strain 10050 only expresses B800-850 LH2 complexes under laboratory conditions despite containing multiple *pucBA* gene pairs. Here, genomic DNA was isolated from *Rps. acidophila* strain 10050 and sequenced by Dr Graham Hamilton and Dr Pawel Herzyk using NGS Illumina technology. Although this work is still on-going, currently 8 *pucBA* gene pairs have been annotated and appear to be organised into 2 distinct operons. The largest operon contains 5 B800-850 type *pucBA* gene pairs (*pucBA4-8*) followed by *pucC*, whilst the second operon contains 3 B800-820 type *pucBA*gene pairs (*pucBA1-3*). The presence of genes that would transcribe a spectral variant of LH2 complexes is intriguing. It is possible that the lack of *pucC* in this second operon could retard the B800-820 LH2 expression. This is suggested as PucC [145] is a putative Bchl transporter and has been shown to be required for efficient LH2 expression. It would be interesting to incorporate or move the *pucC* gene from the first operon to the second operon to see if the type of LH2 complex expressed changes. As the LH2 complexes from this species are structurally amenable, any mutations in the LH2 complex could be analysed structurally by X-ray crystallography. The new genome sequence for *Rps. acidophila* strain 10050 determined here will be an important resource for future studies using deletion mutation approaches to really try to unravel the functional consequences of LH2 multi-gene families.

LH2 complexes contain Bchl and Car molecules. The Car molecules in these complexes contain 9 or more π electrons (N=≥9). In Car molecules with N=≥9 there are theoretical intermediate states [88]. Here, for the first time an intermediate state has been directly observed in LH2 complexes from *Rps. acidophila* strain 10050 and *Rba. sphaeroides* strain 2.4.1 and their respective isolated Car molecules RG and spheroidene. In addition, it has shown that this intermediate electronic state is also involved in ET between the Car and Bchl molecules. Therefore the presence of the intermediate state below the S_2 state bridges the energy gap between S_2 and S_1 and aids ET between S_2-Q_X and S_1-Q_X. This explains the efficiency of ET in LH2 complexes, even when they contain Car molecules with N=≥9. This work was a collaboration with Dr Evgeny Ostroumov and Dr Gregory Scholes using 2DES. Further experiments have been designed that manipulate the spectral shifts of the states in different solvents. By doing so, the S_2 state will be titrated above and below the intermediate state. As the intermediate state accepts energy from the S_2, the appearance and disappearance of the intermediate state should be observed.

8. Appendices

8.1. C-Succinate media

	Per Litre
Concentrated Base	20 ml
1M K_2HPO_4	10 ml
1M KH_2PO_4	10 ml
10 % (w/v) Ammonium sulphate	5 ml
1M Na succinate pH 6.8	10 ml
Growth factors	1 ml
Casamino acids	1 g

Made up with distilled water and autoclaved. Stored at 4°C

Concentrated base

Nitrilotroacetic acid	10 g
Magnesium sulphate	14.45 g
Calcium chloride.$2H_2O$	3.4 g
Ammonium molybdate	0.00925 g
Ferrous sulphate.$7H_2O$	0.099 g
Nicotinic acid	0.05 g

Thiamine HCl	0.025 g
Biotin	0.005 g
Metos 44	50 ml

Adjust to pH 6.8 using 5 N KOH. Store at 4°C

Growth factors

Biotin	0.02 g
$NaHCO_3$	0.5 g
Nicotinic acid	1.0 g
Thiamine HCl	0.5 g
p-Amino benzoic acid	1.0 g

Store at 4°C

Metos 44

EDTA	2.5 g
Zinc (II) sulphate	10.95 g
Manganous sulphate.H_2O	1.54 g
Copper (II) sulphate.H_2O	392 mg
Cobaltous nitrate.H_2O	248 mg
Ferrous sulphate.H_2O	5.5 g

Disodium tetraborate.10H_2O	177 mg

8.2. Pfennig's media

	per litre
Potassium di-hydrogen phosphate	1.0 g
$MgSO_4.7H_2O$	0.4 g
NaCl	0.4 g
Sodium succinate	1.5 g
$CaCl_2.7H_2O$	50 mg
Ammonium chloride	0.5 g
Ferric citrate solution	5.0 ml
Trace elements solution	10.0 ml

Adjust pH to 5.2 with HCl and then autoclave.

Ferric-citrate solution

Add 100mg of ferric citrate to 100mL of boiling distilled water.

Trace elements Solution	per litre
di-Sodium EDTA	0.5 g
$FeSO_4.7H_2O$	10 mg
$MnCl_2.4H_2O$	3 mg

Boric acid	30 mg
$CaCl_2.2H_2O$	1 mg
Nickel (II) chloride.$6H_2O$	2 mg
Sodium molybdate.$2H_2O$	3 mg
Cobalt chloride.$6H_2O$	28 mg

Make up with distilled water, pH should be 3-4.

8.3. Pfennig's medium I

	per litre
$CaCl_2.2H_2O$	0.1 g
$MgCl_2.6H_2O$	0.5 g
$NaHCO_3$	0.8 g
Casamino acid	1 g
Solution A	10 ml
Solution B	1 ml
Solution C	1 ml
Solution D	1.5 ml

Note: Solution D should be added last.

Solution A	per 2 litres

$KH_2PO_4.6H_2O$	68 g
NH_4Cl	68 g
KCl	68 g
Adjust to pH 7.0	
Solution B	per litre
EDTA.2Na	3.0 g
$FeSO_3.7H_2O$	1.1 g
$CoCl_2.6H_2O$	190 mg
$MnCl_2.2H_2O$	50 mg
$ZnCl_2$	42 mg
$NiCl_2.6H_2O$	24 mg
$Na_2MoO_4.2H_2O$	18 mg
H_3BO_3	300 mg
$CuCl_2.2H_2O$	2 mg
Adjust to pH 6.0	

Solution C	per litre
Vitamin B12	0.002 g

Solution D	per litre
$Na_2S.9H_2O$	120 g
Adjust to pH 7.0	

8.4. Luria Broth

	per litre
NaCl	10 g
Yeast extract	5 g
Bactotryptone	10 g

8.5. Lysis buffer

	per ml
1M Tris-HCl pH 8.0	10 µl
0.5 M EDTA	200 µl
10 % (w/v) SDS	50 µl
DNAse and RNAse free water	740 µl

8.6. MemGold

Well	Salt	Buffer	pH	precipitant
A1	None	0.08 M sodium citrate	5.2	2.2 M ammonium sulphate
B1	None	0.01 M Tris-HCl	8.0	1.2 M tri-sodium citrate
C1	None	0.015 M tricine	8.5	24 % (w/v) PEG 4000
D1	0.36 M sodium chloride 0.1 % (w/v) sodium azide	0.015 M sodium phosphate	7.0	9.9 % (w/v) PEG 4000
E1	0.3 M sodium chloride	0.01 M Tris-HCl	8.0	27.5 % (w/v) PEG 4000
F1	None	0.225 M MES/bis-tris	6.6	6.6 % (w/v) PEG 6000
G1	0.1 M ammonium sulphate	0.1 M HEPES	7.5	12.0 % (w/v) PEG 4000/22 % (v/v) glycerol
H1	0.02 M calcium chloride 0.01 M magnesium sulphate 0.02 M sodium chloride	0.2 M MES	6.5	7.7 % (w/v) PEG 1500
A2	None	0.05 M HEPES	7.5	2.5 M ammonium sulphate
B2	None	0.0665 M HEPES	7.5	1.1 M tri-sodium citrate
C2	None	0.15 M potassium phosphate	6.5	3.3 M ammonium sulphate
D2	0.1 M magnesium acetate	0.1 M sodium citrate	5.8	14 % (w/v) PEG 5000 MME
E2	0.1 M sodium chloride	0.02 M sodium citrate	5.6	11 % (w/v) PEG 3350
F2	0.1 M sodium chloride	0.02 M sodium citrate	5.6	5.5 % (w/v) PEG 3350
G2	0.05 M calcium chloride 0.05 M barium chloride	0.1 M Tris-HCl	8.2	32 % (v/v) PEG 400
H2	0.05 M sodium chloride	0.1 M sodium phosphate	6.2	16 % (w/v) PEG 4000
A3	0.1 M magnesium chloride	0.03 M Tris-HCl	8.2	19 % (w/v) PEG 4000
B3	0.2 M sodium chloride	0.025 M HEPES	7.5	13 % (w/v) PEG 4000
C3	None	0.1 M HEPES	7.5	11 % (w/v) PEG 3350
D3	0.1 M sodium chloride	0.02 M KMES	6.7	6.6 % (w/v) PEG 4000
E3	0.1 M potassium chloride	0.02 M Tris	7.0	20 % (w/v) PEG 4000
F3	0.05 M magnesium chloride 0.1 % (w/v) sodium azide	0.1 M sodium cacodylate	6.7	6.6 % (w/v) PEG 3350
G3	0.2 M potassium chloride	0.1 M sodium citrate	5.5	37 % (v/v) pentaerythritol propoxylate
H3	None	0.1 M Tris-HCl	8.0	5.5 % (w/v) PEG 4000
A4	0.1 M sodium chloride	0.02 M Tris	7.0	7.7 % (w/v) PEG 4000
B4	0.1 M magnesium chloride	0.1 M Tris-HCl	7.5	22 % (v/v) PEG 400
C4	0.04 M sodium chloride	0.04 M Tris	8.0	27 % (v/v) PEG 350 MME
D4	0.05 M sodium chloride 0.02 M magnesium chloride	0.1 M sodium citrate	6.0	22 % (v/v) PEG 400
E4	None	0.1 M sodium acetate	5.5	8.8 % (w/v) PET 2000 MME
F4	None	0.4 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
G4	None	0.02 M bis Tris	7.0	15 % (w/v) PEG 2000
H4	0.1 M sodium chloride 0.1 M magnesium chloride	0.02 M Tris	7.5	11 % (w/v) PEG 1500

A5	0.1 M sodium chloride 0.1 M magnesium chloride	0.1 M HEPES	8.0	11 % (w/v) PEG 1500
B5	0.2 M sodium acetate 0.2 M potassium chloride	0.1 M HEPES	7.0	22 % (w/v) PEG 3000
C5	0.02 M nickel sulphate	0.1 M HEPES	7.0	33 % (v/v) Jeffamine-M600
D5	0.15 M sodium chloride	0.1 M Tris-HCl	8.0	13 % (w/v) PEG 6000
E5	0.2 M calcium chloride	0.1 M HEPES	7.5	53 % (v/v) PEG 400
F5	0.05 M magnesium acetate	0.05 M sodium acetate	5.0	28 % (v/v) PEG 400
G5	None	0.05 M HEPES	7.5	22 % (v/v) PEG 4000
H5	0.2 M calcium chloride	0.1 M Tris-HCl	8.0	44 % (v/v) PEG 400
A6	0.05 M magnesium acetate	0.05 M sodium acetate	5.4	24 % (v/v) PEG 400
B6	0.2 M calcium chloride	0.1 M MES	6.5	26 % (v/v) PEG 350 MME
C6	0.1 M potassium chloride	0.1 M Tris-HCl	8.5	39 % (v/v) PEG 400
D6	0.05 M magnesium chloride	0.1 M glycine	9.0	22 % (v/v) PEG 400
E6	0.1 M ammonium sulphate	0.1 M glycine	3.8	28 % (w/v) tri-ethylene glycol
F6	0.15 M sodium formate	0.1 M HEPES	7.2	18 % (w/v) PEG 3350
G6	None	0.2 M sodium acetate	6.8	8.8 % (w/v) PEG 6000
H6	0.2 M potassium chloride	0.1 M MES	6.5	18 % (w/v) PEG 6000
A7	0.22 M sodium citrate	0.1 M Tris	8.0	35 % (v/v) PEG 400
B7	None	0.1 M sodium acetate	4.5	17 % (v/v) PEG 400
C7	None	0.02 M Tris	8.5	1.0 M lithium sulphate/ 1.8 % (w/v) PEG 8000
D7	None	0.02 M Tris	7.5	22 % (v/v) PEG 550 MME
E7	0.05 M sodium chloride	0.02 M glycine	10.0	33 % (w/v) PEG 1000
F7	0.2 M magnesium chloride	0.1 M Tris	8.5	25 % (w/v) PEG 4000
G7	0.2 M magnesium chloride	0.1 M sodium cacodylate	6.5	31 % (w/v) PEG 2000
H7	None	0.64 M sodium acetate	4.6	18 % (w/v) PEG 3350
A8	0.1 M sodium chloride 0.1 M cadmium chloride	0.1 M Tris-HCl	8.0	33 % (v/v) PEG 400
B8	None	0.1 M Bicine	8.9	31 % (w/v) PEG 2000
C8	0.05 M sodium sulphate 0.05 M lithium sulphate	0.05 M Tris	8.5	35 % (v/v) PEG 400
D8	0.1 M sodium chloride	0.05 M glycine	9.5	33 % (v/v) PEG 300
E8	0.3 M magnesium nitrate	0.1 M Tris	8.0	23 % (w/v) PEG 2000
F8	0.12 M lithium sulphate	0.02 M Tris 0.1 M sodium citrate	7.5 5.0	20 % (v/v) PEG 300
G8	0.1 M sodium chloride	0.12 M Tris	9.4	20 % (v/v) PEG 400
H8	0.2 M sodium chloride	0.1 M HEPES	7.0	22 % (v/v) PEG 550 MME
A9	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris	8.0	21 % (v/v) PEG 400
B9	0.02 M sodium citrate	0.08 M sodium phosphate	6.2	18 % (w/v) PEG 2000
C9	0.02 M potassium nitrate	0.03 M potassium citrate	6.5	7.7 % (w/v) PEG 4000
D9	0.1 M sodium chloride 0.005 M magnesium chloride	0.1 M Tris	8.5	30 % (w/v) PEG 2000 MME
E9	0.2 M calcium chloride	0.1 M HEPES	7.0	33 % (v/v) PEG 400

F9	0.1 M calcium chloride	0.1 M Tris	6.5	13 % (w/v) PEG 2000 MME
G9	0.2 M ammonium sulphate 0.02 M sodium chloride	0.02 M sodium acetate	4.0	33 % (v/v) PEG 200
H9	0.07 M sodium chloride	0.05 M sodium citrate	4.5	22 % (v/v) PEG 400
A10	0.2 M ammonium sulphate	0.1 M sodium acetate	4.6	28 % (v/v) PEG 550 MME
B10	None	0.05 M glycine	9.0	55 % (v/v) PEG 400
C10	0.1 M magnesium chloride 0.1 M sodium chloride	0.1 M Tris	8.5	33 % (v/v) PEG 400
D10	0.1 M lithium sulphate 0.05 M di-sodium hydrogen phosphate	0.05 M citric acid	None	19 % (w/v) PEG 1000
E10	0.2 M magnesium chloride 0.1 M postassium chloride	0.025 M sodium citrate	4.0	33 % (v/v) PEG 400
F10	0.05 M zinc acetate	0.05 M MES	6.1	11 % (w/v) PEG 8000
G10	0.3 M magnesium nitrate	0.1 M Tris	8.0	22 % (w/v) PEG 8000
H10	0.1 M sodium chloride 4 % (v/v) ethylene glycol	0.1 M MES	6.0	33 % (v/v) PEG 400
A11	0.05 M sodium chloride	0.1 M sodium citrate	5.5	26 % (v/v) PEG 400
B11	0.1 M lithium sulphate	0.1 M glycine	9.3	30% (v/v) PEG 400
C11	0.15 M potassium citrate 0.05 M lithium citrate	0.1 M sodium phosphate	None	22 % (w/v) PEG 6000
D11	0.001 M zinc sulphate	0.05 M HEPES	7.8	28 % (v/v) PEG 600
E11	0.1 M sodium chloride	0.1 M sodium phosphate	7.0	33 % (v/v) PEG 300
F11	0.1 M sodium chloride	0.05 M Bicine	9.0	3 % (v/v) PEG 300
G11	0.05 M zinc acetate 6 % (v/v) ethylene glycol	0.1 M sodium cacodylate	6.0	6.6 % (w/v) PEG 8000
H11	0.2 M lithium sulphate	0.1 M sodium citrate	3.5	28 % (v/v) PEG 400
A12	0.1 M sodium chloride	0.1 M Tris	7.5	11 % (w/v) PEG 4000
B12	0.05 M lithium sulphate	0.1 M tricine	7.4	7 % (w/v) PEG 3000
C12	0.2 M calcium chloride	0.1 M MES	6.5	33 % (v/v) PEG 400
D12	1 M sodium chloride	0.1 M sodium citrate	6.0	28 % (w/v) PEG 4000
E12	None	0.1 M HEPES	7.5	11 % (w/v) PEG 4000
F12	0.002 M zinc sulphate	0.08 M HEPES	7.0	25 % (v/v) Jeffamine ED2001
G12	0.001 M cadmium chloride 0.03 M magnesium chloride	0.1 M MES	6.5	30 % (v/v) PEG 400
H12	None	0.1 M bis-tris-propane	7.0	3.0 M sodium chloride

8.7. 96-well optimisation - experiment 1

Well	Salt	Buffer	pH	precipitant
A1	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	19 % (v/v) PEG 400
A2	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	20 % (v/v) PEG 400
A3	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
A4	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
A5	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
A6	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
A7	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
A8	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v) PEG 400
A9	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
A10	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	28 % (v/v) PEG 400
A11	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	29 % (v/v)PEG 400
A12	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	30 % (v/v) PEG 400
B1	0.01 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B2	0.05 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B3	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B4	0.15 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B5	0.2 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B6	0.25 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B7	0.3 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B8	0.35 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B9	0.4 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B10	0.45 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B11	0.5 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B12	0.55 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C1	0.1 M sodium chloride 0.1 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C2	0.1 M sodium chloride 0.15 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C3	0.1 M sodium chloride 0.2 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C4	0.1 M sodium chloride 0.25 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C5	0.1 M sodium chloride 0.3 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400

C6	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C7	0.1 M sodium chloride 0.35 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C8	0.1 M sodium chloride 0.4 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C9	0.1 M sodium chloride 0.45 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C10	0.1 M sodium chloride 0.5 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C11	0.1 M sodium chloride 0.55 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C12	0.1 M sodium chloride 0.6 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
D1	0.1 M sodium chloride 0.325 M sodium acetate	0.035 M Tris-HCl	8.0	21 % (v/v) PEG 400
D2	0.1 M sodium chloride 0.325 M sodium acetate	0.04 M Tris-HCl	8.0	21 % (v/v) PEG 400
D3	0.1 M sodium chloride 0.325 M sodium acetate	0.045 M Tris-HCl	8.0	21 % (v/v) PEG 400
D4	0.1 M sodium chloride 0.325 M sodium acetate	0.05 M Tris-HCl	8.0	21 % (v/v) PEG 400
D5	0.1 M sodium chloride 0.325 M sodium acetate	0.055 M Tris-HCl	8.0	21 % (v/v) PEG 400
D6	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
D7	0.1 M sodium chloride 0.325 M sodium acetate	0.15 M Tris-HCl	8.0	21 % (v/v) PEG 400
D8	0.1 M sodium chloride 0.325 M sodium acetate	0.2 M Tris-HCl	8.0	21 % (v/v) PEG 400
D9	0.1 M sodium chloride 0.325 M sodium acetate	0.25 M Tris-HCl	8.0	21 % (v/v) PEG 400
D10	0.1 M sodium chloride 0.325 M sodium acetate	0.3 M Tris-HCl	8.0	21 % (v/v) PEG 400
D11	0.1 M sodium chloride 0.325 M sodium acetate	0.35 M Tris-HCl	8.0	21 % (v/v) PEG 400
D12	0.1 M sodium chloride 0.325 M sodium acetate	0.4 M Tris-HCl	8.0	21 % (v/v) PEG 400
E1	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	4.5	21 % (v/v) PEG 400
E2	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	5.0	21 % (v/v) PEG 400
E3	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	5.5	21 % (v/v) PEG 400
E4	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	6.0	21 % (v/v) PEG 400
E5	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	6.5	21 % (v/v) PEG 400
E6	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.0	21 % (v/v) PEG 400
E7	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.5	21 % (v/v) PEG 400
E8	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
E9	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.5	21 % (v/v) PEG 400
E10	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	9.0	21 % (v/v) PEG 400
E11	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	9.5	21 % (v/v) PEG 400
E12	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	10.0	21 % (v/v) PEG 400

F1	None	0.15M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F2	None	0.2 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F3	None	0.25 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F4	None	0.3 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F5	None	0.35 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F6	None	0.4 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F7	None	0.45 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F8	None	0.5 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F9	None	0.55 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F10	None	0.6 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F11	None	0.65 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
F12	None	0.7 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
G1	None	0.4 M ammonium acetate	4.5	13 % (w/v) PEG 2000 MME
G2	None	0.4 M ammonium acetate	5.0	13 % (w/v) PEG 2000 MME
G3	None	0.4 M ammonium acetate	5.5	13 % (w/v) PEG 2000 MME
G4	None	0.4 M ammonium acetate	6.0	13 % (w/v) PEG 2000 MME
G5	None	0.4 M ammonium acetate	6.5	13 % (w/v) PEG 2000 MME
G6	None	0.4 M ammonium acetate	7.0	13 % (w/v) PEG 2000 MME
G7	None	0.4 M ammonium acetate	7.5	13 % (w/v) PEG 2000 MME
G8	None	0.4 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
G9	None	0.4 M ammonium acetate	8.5	13 % (w/v) PEG 2000 MME
G10	None	0.4 M ammonium acetate	9.0	13 % (w/v) PEG 2000 MME
G11	None	0.4 M ammonium acetate	9.5	13 % (w/v) PEG 2000 MME
G12	None	0.4 M ammonium acetate	10.0	13 % (w/v) PEG 2000 MME
H1	None	0.4 M ammonium acetate	8.0	10 % (w/v) PEG 2000 MME
H2	None	0.4 M ammonium acetate	8.0	11 % (w/v) PEG 2000 MME
H3	None	0.4 M ammonium acetate	8.0	12 % (w/v) PEG 2000 MME
H4	None	0.4 M ammonium acetate	8.0	13 % (w/v) PEG 2000 MME
H5	None	0.4 M ammonium acetate	8.0	14 % (w/v) PEG 2000 MME
H6	None	0.4 M ammonium acetate	8.0	15 % (w/v) PEG 2000 MME
H7	None	0.4 M ammonium acetate	8.0	16 % (w/v) PEG 2000 MME
H8	None	0.4 M ammonium acetate	8.0	17 % (w/v) PEG 2000 MME
H9	None	0.4 M ammonium acetate	8.0	18 % (w/v) PEG 2000 MME
H10	None	0.4 M ammonium acetate	8.0	19 % (w/v) PEG 2000 MME
H11	None	0.4 M ammonium acetate	8.0	20 % (w/v) PEG 2000 MME
H12	None	0.4 M ammonium acetate	8.0	21 % (w/v) PEG 2000 MME

The second 96-well optimisation experiment contained the original conditions with the Hampton additive screen.

8.8. 96-well optimisation experiment-3

Well	Salt	Buffer	pH	precipitant
A1*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	19 % (v/v) PEG 400
A2*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	20 % (v/v) PEG 400
A3*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
A4*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
A5*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
A6*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
A7*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
A8*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v) PEG 400
A9*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
A10*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	28 % (v/v) PEG 400
A11*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	29 % (v/v)PEG 400
A12*	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	30 % (v/v) PEG 400
B1	0.01 M sodium chloride 0.325 M sodium acetate	0.035 M Tris-HCl	8.0	21 % (v/v) PEG 400
B2	0.05 M sodium chloride 0.325 M sodium acetate	0.04 M Tris-HCl	8.0	21 % (v/v) PEG 400
B3	0.1 M sodium chloride 0.325 M sodium acetate	0.045 M Tris-HCl	8.0	21 % (v/v) PEG 400
B4	0.15 M sodium chloride 0.325 M sodium acetate	0.05 M Tris-HCl	8.0	21 % (v/v) PEG 400
B5	0.2 M sodium chloride 0.325 M sodium acetate	0.055 M Tris-HCl	8.0	21 % (v/v) PEG 400
B6	0.25 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
B7	0.3 M sodium chloride 0.325 M sodium acetate	0.15 M Tris-HCl	8.0	21 % (v/v) PEG 400
B8	0.35 M sodium chloride 0.325 M sodium acetate	0.2 M Tris-HCl	8.0	21 % (v/v) PEG 400
B9	0.4 M sodium chloride 0.325 M sodium acetate	0.25 M Tris-HCl	8.0	21 % (v/v) PEG 400
B10	0.45 M sodium chloride 0.325 M sodium acetate	0.3 M Tris-HCl	8.0	21 % (v/v) PEG 400
B11	0.5 M sodium chloride 0.325 M sodium acetate	0.35 M Tris-HCl	8.0	21 % (v/v) PEG 400
B12	0.55 M sodium chloride 0.325 M sodium acetate	0.4 M Tris-HCl	8.0	21 % (v/v) PEG 400
C1	0.1 M sodium chloride 0.035 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C2	0.1 M sodium chloride 0.064 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C3	0.1 M sodium chloride 0.092 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C4	0.1 M sodium chloride 0.121 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C5	0.1 M sodium chloride 0.15 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400

C6	0.1 M sodium chloride 0.178 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C7	0.1 M sodium chloride 0.207 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C8	0.1 M sodium chloride 0.235 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C9	0.1 M sodium chloride 0.264 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C10	0.1 M sodium chloride 0.293 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C11	0.1 M sodium chloride 0.321 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
C12	0.1 M sodium chloride 0.35 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
D1	0.01 M sodium chloride 0.325 M sodium acetate	0.035 M Tris-HCl	8.0	21 % (v/v) PEG 400
D2	0.027 M sodium chloride 0.325 M sodium acetate	0.04 M Tris-HCl	8.0	21 % (v/v) PEG 400
D3	0.045 M sodium chloride 0.325 M sodium acetate	0.045 M Tris-HCl	8.0	21 % (v/v) PEG 400
D4	0.062 M sodium chloride 0.325 M sodium acetate	0.05 M Tris-HCl	8.0	21 % (v/v) PEG 400
D5	0.079 M sodium chloride 0.325 M sodium acetate	0.055 M Tris-HCl	8.0	21 % (v/v) PEG 400
D6	0.096 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (v/v) PEG 400
D7	0.114 M sodium chloride 0.325 M sodium acetate	0.15 M Tris-HCl	8.0	21 % (v/v) PEG 400
D8	0.131 M sodium chloride 0.325 M sodium acetate	0.2 M Tris-HCl	8.0	21 % (v/v) PEG 400
D9	0.148 M sodium chloride 0.325 M sodium acetate	0.25 M Tris-HCl	8.0	21 % (v/v) PEG 400
D10	0.165 M sodium chloride 0.325 M sodium acetate	0.3 M Tris-HCl	8.0	21 % (v/v) PEG 400
D11	0.183 M sodium chloride 0.325 M sodium acetate	0.35 M Tris-HCl	8.0	21 % (v/v) PEG 400
D12	0.2 M sodium chloride 0.325 M sodium acetate	0.4 M Tris-HCl	8.0	21 % (v/v) PEG 400
E1	0.1 M sodium chloride 0.035 M sodium acetate	0.01 M Tris-HCl	8.0	21 % (v/v) PEG 400
E2	0.1 M sodium chloride 0.064 M sodium acetate	0.027 M Tris-HCl	8.0	21 % (v/v) PEG 400
E3	0.1 M sodium chloride 0.092 M sodium acetate	0.045 M Tris-HCl	8.0	21 % (v/v) PEG 400
E4	0.1 M sodium chloride 0.121 M sodium acetate	0.061 M Tris-HCl	8.0	21 % (v/v) PEG 400
E5	0.1 M sodium chloride 0.15 M sodium acetate	0.079 M Tris-HCl	8.0	21 % (v/v) PEG 400
E6	0.1 M sodium chloride 0.178 M sodium acetate	0.096 M Tris-HCl	8.0	21 % (v/v) PEG 400
E7	0.1 M sodium chloride 0.207 M sodium acetate	0.114 M Tris-HCl	8.0	21 % (v/v) PEG 400
E8	0.1 M sodium chloride 0.235 M sodium acetate	0.131 M Tris-HCl	8.0	21 % (v/v) PEG 400
E9	0.1 M sodium chloride 0.263 M sodium acetate	0.148 M Tris-HCl	8.0	21 % (v/v) PEG 400
E10	0.1 M sodium chloride 0.293 M sodium acetate	0.165 M Tris-HCl	8.0	21 % (v/v) PEG 400
E11	0.1 M sodium chloride 0.321 M sodium acetate	0.183 M Tris-HCl	8.0	21 % (v/v) PEG 400
E12	0.1 M sodium chloride 0.35 M sodium acetate	0.2 M Tris-HCl	8.0	21 % (v/v) PEG 400
F1	0.1 M sodium chloride	0.1 M Tris-HCl	8.0	19 % (w/v) PEG 400

	0.035 M sodium acetate			
F2	0.1 M sodium chloride 0.064 M sodium acetate	0.1 M Tris-HCl	8.0	20 % (w/v) PEG 400
F3	0.1 M sodium chloride 0.092 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (w/v) PEG 400
F4	0.1 M sodium chloride 0.121 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (w/v) PEG 400
F5	0.1 M sodium chloride 0.15 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (w/v) PEG 400
F6	0.1 M sodium chloride 0.178 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (w/v) PEG 400
F7	0.1 M sodium chloride 0.207 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (w/v) PEG 400
F8	0.1 M sodium chloride 0.235 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (w/v) PEG 400
F9	0.1 M sodium chloride 0.263 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (w/v) PEG 400
F10	0.1 M sodium chloride 0.293 M sodium acetate	0.1 M Tris-HCl	8.0	28 % (w/v) PEG 400
F11	0.1 M sodium chloride 0.321 M sodium acetate	0.1 M Tris-HCl	8.0	29 % (w/v) PEG 400
F12	0.1 M sodium chloride 0.35 M sodium acetate	0.1 M Tris-HCl	8.0	30 % (w/v) PEG 400
G1	0.01 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	19 % (w/v) PEG 400
G2	0.027 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	20 % (w/v) PEG 400
G3	0.045 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	21 % (w/v) PEG 400
G4	0.061 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (w/v) PEG 400
G5	0.079 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (w/v) PEG 400
G6	0.096 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (w/v) PEG 400
G7	0.114 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (w/v) PEG 400
G8	0.131 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (w/v) PEG 400
G9	0.148 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (w/v) PEG 400
G10	0.165 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	28 % (w/v) PEG 400
G11	0.183 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	29 % (w/v) PEG 400
G12	0.2 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	30 % (w/v) PEG 400
H1	0.1 M sodium chloride 0.325 M sodium acetate	0.01 M Tris-HCl	8.0	19 % (w/v) PEG 400
H2	0.1 M sodium chloride 0.325 M sodium acetate	0.027 M Tris-HCl	8.0	20 % (w/v) PEG 400
H3	0.1 M sodium chloride 0.325 M sodium acetate	0.045 M Tris-HCl	8.0	21 % (w/v) PEG 400
H4	0.1 M sodium chloride 0.325 M sodium acetate	0.061 M Tris-HCl	8.0	22 % (w/v) PEG 400
H5	0.1 M sodium chloride 0.325 M sodium acetate	0.079 M Tris-HCl	8.0	23 % (w/v) PEG 400
H6	0.1 M sodium chloride 0.325 M sodium acetate	0.096 M Tris-HCl	8.0	24 % (w/v) PEG 400
H7	0.1 M sodium chloride 0.325 M sodium acetate	0.114 M Tris-HCl	8.0	25 % (w/v) PEG 400
H8	0.1 M sodium chloride 0.325 M sodium acetate	0.131 M Tris-HCl	8.0	26 % (w/v) PEG 400

H9	0.1 M sodium chloride 0.325 M sodium acetate	0.148 M Tris-HCl	8.0	27 % (w/v) PEG 400
H10	0.1 M sodium chloride 0.325 M sodium acetate	0.165 M Tris-HCl	8.0	28 % (w/v) PEG 400
H11	0.1 M sodium chloride 0.325 M sodium acetate	0.183 M Tris-HCl	8.0	29 % (w/v) PEG 400
H12	0.1 M sodium chloride 0.325 M sodium acetate	0.2 M Tris-HCl	8.0	30 % (w/v) PEG 400

* = only the Tris-HCl buffer was adjusted to pH 8.0. In experiment 1 all individual reservoir solutions were adjusted to the stated pH.

8.9. 96-well optimisation - experiment 4

Well	Salt	Buffer	pH	precipitant
A1	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
A2	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
A3	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
A4	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
A5	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v) PEG 400
A6	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
A7	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
A8	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
A9	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
A10	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
A11	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
A12	0.1 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
B1	0.05 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
B2	0.05 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
B3	0.05 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
B4	0.05 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
B5	0.05 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
B6	0.05 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
B7	0.1 M sodium chloride 0.3 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
B8	0.1 M sodium chloride 0.3 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
B9	0.1 M sodium chloride 0.3 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
B10	0.1 M sodium chloride 0.3 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
B11	0.1 M sodium chloride	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400

	0.3 M sodium acetate			
B12	0.1 M sodium chloride 0.3 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
C1	0.025 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
C2	0.025 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
C3	0.025 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
C4	0.025 M sodium chloride 0.325m	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
C5	0.025 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
C6	0.025 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
C7	0.1 M sodium chloride 0.2 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
C8	0.1 M sodium chloride 0.2 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
C9	0.1 M sodium chloride 0.2 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
C10	0.1 M sodium chloride 0.2 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
C11	0.1 M sodium chloride 0.2 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
C12	0.1 M sodium chloride 0.2 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
D1	0.15 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
D2	0.15 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
D3	0.15 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
D4	0.15 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
D5	0.15 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
D6	0.15 M sodium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
D7	0.1 M sodium chloride 0.1 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
D8	0.1 M sodium chloride 0.1 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
D9	0.1 M sodium chloride 0.1 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
D10	0.1 M sodium chloride 0.1 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
D11	0.1 M sodium chloride 0.1 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
D12	0.1 M sodium chloride 0.1 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
E1	0.1 M sodium chloride 0.035 M sodium acetate	0.1 M Tris-HCl	7.0	22 % (v/v) PEG 400
E2	0.1 M sodium chloride 0.035 M sodium acetate	0.1 M Tris-HCl	7.4	23 % (v/v) PEG 400
E3	0.1 M sodium chloride 0.035 M sodium acetate	0.1 M Tris-HCl	7.8	24 % (v/v) PEG 400
E4	0.1 M sodium chloride 0.035 M sodium acetate	0.1 M Tris-HCl	8.2	25 % (v/v) PEG 400
E5	0.1 M sodium chloride 0.035 M sodium acetate	0.1 M Tris-HCl	8.6	26 % (v/v)PEG 400
E6	0.1 M sodium chloride 0.035 M sodium acetate	0.1 M Tris-HCl	9.0	27 % (v/v) PEG 400

E7	0.1 M sodium chloride 0.207 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
E8	0.1 M sodium chloride 0.235 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
E9	0.1 M sodium chloride 0.263 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
E10	0.1 M sodium chloride 0.293 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
E11	0.1 M sodium chloride 0.321 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
E12	0.1 M sodium chloride 0.35 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
F1	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
F2	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.0	23 % (v/v) PEG 400
F3	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.4	24 % (v/v) PEG 400
F4	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.8	25 % (v/v) PEG 400
F5	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.2	26 % (v/v)PEG 400
F6	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.6	27 % (v/v) PEG 400
F7	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	9.0	22 % (v/v) PEG 400
F8	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
F9	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
F10	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
F11	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
F12	0.1 M zinc chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
G1	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
G2	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.0	23 % (v/v) PEG 400
G3	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.4	24 % (v/v) PEG 400
G4	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.8	25 % (v/v) PEG 400
G5	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.2	26 % (v/v)PEG 400
G6	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.6	27 % (v/v) PEG 400
G7	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	9.0	22 % (v/v) PEG 400
G8	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
G9	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
G10	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
G11	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
G12	0.1 M potassium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400
H1	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	22 % (v/v) PEG 400
H2	0.1 M magnesium chloride	0.1 M Tris-HCl	7.0	23 % (v/v) PEG 400

	0.325 M sodium acetate			
H3	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.4	24 % (v/v) PEG 400
H4	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	7.8	25 % (v/v) PEG 400
H5	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.2	26 % (v/v)PEG 400
H6	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.6	27 % (v/v) PEG 400
H7	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	9.0	22 % (v/v) PEG 400
H8	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	23 % (v/v) PEG 400
H9	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	24 % (v/v) PEG 400
H10	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	25 % (v/v) PEG 400
H11	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	26 % (v/v)PEG 400
H12	0.1 M magnesium chloride 0.325 M sodium acetate	0.1 M Tris-HCl	8.0	27 % (v/v) PEG 400

8.10. 96-well optimisation experiment-5

Well	Salt	Buffer	pH	precipitant
A1	0.3 M lithium sulphate	0.1 M MES	6.5	23 % (v/v) PEG 400
A2	0.3 M lithium sulphate	0.1 M MES	6.5	24 % (v/v) PEG 400
A3	0.3 M lithium sulphate	0.1 M MES	6.5	25 % (v/v) PEG 400
A4	0.3 M lithium sulphate	0.1 M MES	6.5	26 % (v/v) PEG 400
A5	0.3 M lithium sulphate	0.1 M MES	6.5	27 % (v/v) PEG 400
A6	0.3 M lithium sulphate	0.1 M MES	6.5	28 % (v/v) PEG 400
A7	None	0.07 M sodium citrate	4.5	20 % (v/v) PEG 300
A8	None	0.07 M sodium citrate	4.5	21 % (v/v) PEG 300
A9	None	0.07 M sodium citrate	4.5	22 % (v/v) PEG 300
A10	None	0.07 M sodium citrate	4.5	23 % (v/v) PEG 300
A11	None	0.07 M sodium citrate	4.5	24 % (v/v) PEG 300
A12	None	0.07 M sodium citrate	4.5	25 % (v/v) PEG 300
B1	0.05 M lithium sulphate	0.1 M MES	6.5	25 % (v/v) PEG 400
B2	0.1 M lithium sulphate	0.1 M MES	6.5	25 % (v/v) PEG 400
B3	0.15 M lithium sulphate	0.1 M MES	6.5	25 % (v/v) PEG 400
B4	0.2 M lithium sulphate	0.1 M MES	6.5	25 % (v/v) PEG 400
B5	0.25 M lithium sulphate	0.1 M MES	6.5	25 % (v/v) PEG 400
B6	0.3 M lithium sulphate	0.1 M MES	6.5	25 % (v/v) PEG 400
B7	None	0.05 M sodium citrate	4.5	22 % (v/v) PEG 300
B8	None	0.06 M sodium citrate	4.5	22 % (v/v) PEG 300
B9	None	0.07 M sodium citrate	4.5	22 % (v/v) PEG 300
B10	None	0.08 M sodium citrate	4.5	22 % (v/v) PEG 300
B11	None	0.09 M sodium citrate	4.5	22 % (v/v) PEG 300
B12	None	0.1 M sodium citrate	4.5	22 % (v/v) PEG 300
C1	0.3 M lithium sulphate	0.05 M MES	6.5	25 % (v/v) PEG 400
C2	0.3 M lithium sulphate	0.06 M MES	6.5	25 % (v/v) PEG 400
C3	0.3 M lithium sulphate	0.07 M MES	6.5	25 % (v/v) PEG 400
C4	0.3 M lithium sulphate	0.08 M MES	6.5	25 % (v/v) PEG 400
C5	0.3 M lithium sulphate	0.09 M MES	6.5	25 % (v/v) PEG 400
C6	0.3 M lithium sulphate	0.1 M MES	6.5	25 % (v/v) PEG 400
C7	None	0.05 M sodium citrate	4.5	20 % (v/v) PEG 300
C8	None	0.06 M sodium citrate	4.5	20 % (v/v) PEG 300
C9	None	0.07 M sodium citrate	4.5	20 % (v/v) PEG 300

C10	None	0.08 M sodium citrate	4.5	20 % (v/v) PEG 300
C11	None	0.09 M sodium citrate	4.5	20 % (v/v) PEG 300
C12	None	0.1 M sodium citrate	4.5	20 % (v/v) PEG 300
D1	0.05 M lithium sulphate	0.1 M MES	6.5	23 % (v/v) PEG 400
D2	0.1 M lithium sulphate	0.1 M MES	6.5	23 % (v/v) PEG 400
D3	0.15 M lithium sulphate	0.1 M MES	6.5	23 % (v/v) PEG 400
D4	0.2 M lithium sulphate	0.1 M MES	6.5	23 % (v/v) PEG 400
D5	0.25 M lithium sulphate	0.1 M MES	6.5	23 % (v/v) PEG 400
D6	0.3 M lithium sulphate	0.1 M MES	6.5	23 % (v/v) PEG 400
D7	None	0.05 M sodium citrate	4.5	21 % (v/v) PEG 300
D8	None	0.06 M sodium citrate	4.5	21 % (v/v) PEG 300
D9	None	0.07 M sodium citrate	4.5	21 % (v/v) PEG 300
D10	None	0.08 M sodium citrate	4.5	21 % (v/v) PEG 300
D11	None	0.09 M sodium citrate	4.5	21 % (v/v) PEG 300
D12	None	0.1 M sodium citrate	4.5	21 % (v/v) PEG 300
E1	0.05 M lithium sulphate	0.1 M MES	6.5	24 % (v/v) PEG 400
E2	0.1 M lithium sulphate	0.1 M MES	6.5	24 % (v/v) PEG 400
E3	0.15 M lithium sulphate	0.1 M MES	6.5	24 % (v/v) PEG 400
E4	0.2 M lithium sulphate	0.1 M MES	6.5	24 % (v/v) PEG 400
E5	0.25 M lithium sulphate	0.1 M MES	6.5	24 % (v/v) PEG 400
E6	0.3 M lithium sulphate	0.1 M MES	6.5	24 % (v/v) PEG 400
E7	None	0.05 M sodium citrate	4.5	23 % (v/v) PEG 300
E8	None	0.06 M sodium citrate	4.5	23 % (v/v) PEG 300
E9	None	0.07 M sodium citrate	4.5	23 % (v/v) PEG 300
E10	None	0.08 M sodium citrate	4.5	23 % (v/v) PEG 300
E11	None	0.09 M sodium citrate	4.5	23 % (v/v) PEG 300
E12	None	0.1 M sodium citrate	4.5	23 % (v/v) PEG 300
F1	0.05 M lithium sulphate	0.1 M MES	6.5	26 % (v/v) PEG 400
F2	0.1 M lithium sulphate	0.1 M MES	6.5	26 % (v/v) PEG 400
F3	0.15 M lithium sulphate	0.1 M MES	6.5	26 % (v/v) PEG 400
F4	0.2 M lithium sulphate	0.1 M MES	6.5	26 % (v/v) PEG 400
F5	0.25 M lithium sulphate	0.1 M MES	6.5	26 % (v/v) PEG 400
F6	0.3 M lithium sulphate	0.1 M MES	6.5	26 % (v/v) PEG 400
F7	None	0.05 M sodium citrate	4.5	24 % (v/v) PEG 300
F8	None	0.06 M sodium citrate	4.5	24 % (v/v) PEG 300
F9	None	0.07 M sodium citrate	4.5	24 % (v/v) PEG 300
F10	None	0.08 M sodium citrate	4.5	24 % (v/v) PEG 300
F11	None	0.09 M sodium citrate	4.5	24 % (v/v) PEG 300
F12	None	0.1 M sodium citrate	4.5	24 % (v/v) PEG 300
G1	0.05 M lithium sulphate	0.1 M MES	6.5	27 % (v/v) PEG 400
G2	0.1 M lithium sulphate	0.1 M MES	6.5	27 % (v/v) PEG 400
G3	0.15 M lithium sulphate	0.1 M MES	6.5	27 % (v/v) PEG 400
G4	0.2 M lithium sulphate	0.1 M MES	6.5	27 % (v/v) PEG 400
G5	0.25 M lithium sulphate	0.1 M MES	6.5	27 % (v/v) PEG 400
G6	0.3 M lithium sulphate	0.1 M MES	6.5	27 % (v/v) PEG 400
G7	None	0.05 M sodium citrate	4.5	25 % (v/v) PEG 300
G8	None	0.06 M sodium citrate	4.5	25 % (v/v) PEG 300
G9	None	0.07 M sodium citrate	4.5	25 % (v/v) PEG 300
G10	None	0.08 M sodium citrate	4.5	25 % (v/v) PEG 300
G11	None	0.09 M sodium citrate	4.5	25 % (v/v) PEG 300
G12	None	0.1 M sodium citrate	4.5	25 % (v/v) PEG 300
H1	0.05 M lithium sulphate	0.1 M MES	6.5	28 % (v/v) PEG 400
H2	0.1 M lithium sulphate	0.1 M MES	6.5	28 % (v/v) PEG 400
H3	0.15 M lithium sulphate	0.1 M MES	6.5	28 % (v/v) PEG 400
H4	0.2 M lithium sulphate	0.1 M MES	6.5	28 % (v/v) PEG 400
H5	0.25 M lithium sulphate	0.1 M MES	6.5	28 % (v/v) PEG 400
H6	0.3 M lithium sulphate	0.1 M MES	6.5	28 % (v/v) PEG 400
H7	None	0.05 M sodium citrate	4.5	26 % (v/v) PEG 300
H8	None	0.06 M sodium citrate	4.5	26 % (v/v) PEG 300
H9	None	0.07 M sodium citrate	4.5	26 % (v/v) PEG 300
H10	None	0.08 M sodium citrate	4.5	26 % (v/v) PEG 300

H11	None	0.09 M sodium citrate	4.5	26 % (v/v) PEG 300
H12	None	0.1 M sodium citrate	4.5	26 % (v/v) PEG 300

Bibliography

[1] Cogdell, R.J., Brotosudarmo, T.H., Gardiner, A.T., Sanchez, P.M. and Cronin, L. (2010). Artificial photosynthesis-solar fuels: current status and future prospects. *Biofuels* **1**, 861-876.

[2] Scheer, H. (2001). A solar manifesto: second edition. James and James *Science Publishers Ltd*

[3] Gust, D., Moore, T.A. and Moore, A.L. (2009). Solar Fuels via Artificial Photosynthesis. *Accounts of Chemical Research* 42, 1890-1898.

[4] Cogdell, R.J., Gardiner, A.T. and Cronin, L. (2012). Learning from photosynthesis: how to use solar energy to make fuels. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences* **370**, 3819-3826.

[5] Gust, D., Moore, T.A. and Moore, A.L. (2001). Mimicking photosynthetic solar energy transduction. *Accounts of Chemical Research* 34, 40-48.

[6] Moore, T.A., Moore, A.L. and Gust, D. (2013) Artificial photosynthesis combines biology with technology for sustainable energy transformation. In *Nobel Symposium 153: Nanoscale Energy Converters* (Linke, H., Borgstrom, M., Pullerits, T., Samuelson, L., Sundstrom, V. and Inganas, O., ed.^eds), pp. 68-72

[7] Clayton, R.K. (1980) Photosynthesis. Physical mechanisms and chemical patterns, *Cambridge University Press*

[8] Gest, H. and Blankenship, R.E. (2004). Time line of discoveries: anoxygenic bacterial photosynthesis. *Photosynthesis Research* **80**, 59-70.

[9] van Neil, C.B. (1941). The bacterial photosyntheses and their importance for the general problem of photosynthesis. Advances in Enzymology and Related Subjects of Biochemistry 1, 263-328.

[10] Drews, G. (1996). Formation of the light-harvesting complex I (B870) of anoxygenic phototrophic purple bacteria. *Archives of Microbiology* **166**, 151-159.

[11] Madigan, M. and Jung, D. (2008) An Overview of Purple Bacteria: Systematics, Physiology, and Habitats. In The Purple Phototrophic Bacteria (Hunter, C.N., Daldal, F., Thurnauer, M. and Beatty, J.T., ed.^eds), pp. 1-15. *Springer Netherlands*

[12] Stackebrandt, E., Murray, R.G.E. and Truper, H.G. (1988). Proteobacteria-classis NOV, a name for the phylogenetic taxon that includes the purple bacteria and their relatives. *International Journal of Systematic Bacteriology* **38**, 321-325.

[13] Williams, K.P., Sobral, B.W. and Dickerman, A.W. (2007). A robust species tree for the Alphaproteobacteria. *Journal of Bacteriology* **189**, 4578-4586.

[14] Cogdell, R.J., Fyfe, P.K., Barrett, S.J., Prince, S.M., Freer, A.A., Isaacs, N.W., McGlynn, P. and Hunter, C.N. (1996). The purple bacterial photosynthetic unit. *Photosynthesis Research* **48**, 55-63.

[15] Emerson, R. and Arnold, W. (1932). The photochemical reaction in photosynthesis. *Journal of General Physiology* **16**, 191-205.

[16] Aagaard, J. and Sistrom, W.R. (1972). Control of synthesis of reaction center bacteriochlorophyll in photosynthetic bacteria. *Photochemistry and Photobiology* **15**, 209-&.

[17] Kaplan, S., Fraley, R.T. and Lueking, D.R. (1979) Relationship of intra-cytoplasmic membrane assembly to the cell-division cycle in

rhodopseudomonas sphaeroides. *Journal of Biological Chemistry* **254**, 6:1980-1986.

[18] Roszak, A.W., Howard, T.D., Southall, J., Gardiner, A.T., Law, C.J., Isaacs, N.W. and Cogdell, R.J. (2003). Crystal structure of the RC-LH1 core complex from Rhodopseudomonas palustris. *Science* **302**, 1969-1972.

[19] Jones, M.R., Fowler, G.J.S., Gibson, L.C.D., Grief, G.G., Olsen, J.D., Crielaard, W. and Hunter, C.N. (1992).Mutants of rhodobacter-sphaeroides lacking one or more pigment protein complexes and complementation with reaction-center, LH1 and LH2 genes. *Molecular Microbiology* **6**, 1173-1184.

[20] Cogdell, R.J., Howard, T.D., Isaacs, N.W., McLuskey, K. and Gardiner, A.T. (2002). Structural factors which control the position of the Q(y) absorption band of bacteriochlorophyll a in purple bacterial antenna complexes. *Photosynthesis Research* **74**, 135-141.

[21] Deisenhofer, J., Epp, O., Miki, K., Huber, R. and Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction center of rhodopseudomonas viridis at 3A resolution. *Nature* **318**, 618-624.

[22] Deisenhofer, J. and Michel, H. (1989). The photosynthetic reaction center from the purple bacterium rhodopseudomonas viridis - Nobel lecture, December 8 1988. *Chemica Scripta* **29**, 205-&.

[23] DeLano, W.L. and Lam, J.W. (2005). PyMOL: A communications tool for computational models. Abstracts of Papers of the *American Chemical Society* **230**, U1371-U1372.

[24] Holten, D., Windsor, M.W., Parson, W.W. and Thornber, J.P. (1978). Primary photo-chemical processes in isolated reaction centers of rhodopseudomonas viridis. *Biochimica Et Biophysica Acta* **501**, 112-126.

[25] Dutton, P.L. and Jackson, J.B. (1972). Thermodynamic and kinetic characterisation of electron transfer components in-situ in rhodopseudomonas sphaeroides and rhodospirillum rubrum. *European Journal of Biochemistry* **30**, 495-&.

[26] Rockley, M.G., Windsor, M.W., Cogdell, R.J. and Parson, W.W. (1975). Picosecond detection of an intermediate in photochemical reaction of bacterial photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 2251-2255.

[27] Kaufmann, K.J., Dutton, P.L., Netzel, T.L., Leigh, J.S. and Rentzepis, P.M. (1975). Picosecond kinetics of events leading to reaction center bacteriochlorophyll oxidation. *Science* **188**, 1301-1304.

[28] Deisenhofer, J. and Norris, J.R. (1993) The Photosynthetic reaction center, *Academic Press*, *San Diego.*

[29] Tanaka, S. and Marcus, R.A. (1997). Electron transfer model for the electric field effect on quantum yield of charge separation in bacterial photosynthetic reaction centers. *Journal of Physical Chemistry B* **101**, 5031-5045.

[30] Moser, C.C., Page, C.C., Cogdell, R.J., Barber, J., Wraight, C.A. and Dutton, P.L. (2003). Length, time, and energy scales of photosystems. *Membrane Proteins* **63**, 71-109.

[31] Moser, C.C. and Dutton, P.L. (1988). Cytochrome-c and cytochrome-c2 binding dynamics and electron transfer with photosynthetic reaction center protein and other integral membrane redox proteins. *Biochemistry* **27**, 2450-2461.

[32] Allen, J.P., Feher, G., Yeates, T.O., Komiya, H. and Rees, D.C. (1987). Structure of the reaction center from rhodobacter shpaeroides R-26 - the

cofactors 1. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 5730-5734.

[33] Holzapfel, W., Finkele, U., Kaiser, W., Oesterhelt, D., Scheer, H., Stilz, H.U. and Zinth, W. (1990). Initial electron transfer in the reaction center from rhodobacter sphaeroides. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 5168-5172.

[34] Kirmaier, C. and Holten, D. (1987). Primary photochemistry of reaction centers from the photosynthetic purple bacteria. *Photosynthesis Research* **13**, 225-260.

[35] Stowell, M.H.B., McPhillips, T.M., Rees, D.C., Soltis, S.M., Abresch, E. and Feher, G. (1997). Light-induced structural changes in photosynthetic reaction center: Implications for mechanism of electron-proton transfer. *Science* **276**, 812-816.

[36] Michel, H., Epp, O. and Deisenhofer, J. (1986). Pigment protein interactions in the photosynthetic reaction center from rhodopseudomonas viridis. *Embo Journal* **5**, 2445-2451.

[37] Fleming, G.R., Martin, J.L. and Breton, J. (1988). Rates of primary electron transfer in photosynthetic reaction centers and their mechanistic implications. *Nature* **333**, 190-192.

[38] Jones, M.R. (2009). The petite purple photosynthetic powerpack. *Biochemical Society Transactions* **37**, 400-407.

[39] Allen, J.P., Feher, G., Yeates, T.O., Komiya, H. and Rees, D.C. (1987). Structure of the reaction center from rhodobacter sphaeroides R-26 - the protein subunits. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 6162-6166.

[40] Allen, J.P., Feher, G., Yeates, T.O., Komiya, H. and Rees, D.C. (1987). Structure of the reaction center from rhodopseudomonas sphaeroides at 3.3 A resolution. *Biophysical Journal* **51**, A377-A377.

[41] Yeates, T.O., Komiya, H., Rees, D.C., Allen, J.P. and Feher, G. (1987). Structure of the reaction center from rhodopseudomonas sphaeroides R-26 - membrane protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 6438-6442.

[42] Kereiche, S., Bourinet, L., Keegstra, W., Arteni, A.A., Verbavatz, J.M., Boekema, E.J., Robert, B. and Gall, A. (2008). The peripheral light-harvesting complexes from purple sulfur bacteria have different 'ring' sizes. *Febs Letters* **582**, 3650-3656.

[43] McDermott, G., Prince, S.M., Freer, A.A., Hawthornthwaitelawless, A.M., Papiz, M.Z., Cogdell, R.J. and Isaacs, N.W. (1995). Crystal structure of an integral membrane light harvesting complex from photosynthetic bacteria. *Nature* **374**, 517-521.

[44] Gardiner, A.T., MacKenzie, R.C., Barrett, S.J., Kaiser, K. and Cogdell, R.J. (1996). The purple photosynthetic bacterium Rhodopseudomonas acidophila contains multiple puc peripheral antenna complex (LH2) genes: Cloning and initial characterisation of four beta/alpha pairs. *Photosynthesis Research* **49**, 223-235.

[45] Law, C.J., Roszak, A.W., Southall, J., Gardiner, A.T., Isaacs, N.W. and Cogdell, R.J. (2004). The structure and function of bacterial light-harvesting complexes (Review). *Molecular Membrane Biology* **21**, 183-191.

[46] Fowler, G.J.S., Gardiner, A.T., Mackenzie, R.C., Barratt, S.J., Simmons, A.E., Westerhuis, W.H.J. and Cogdell, R.J. (1995). Heterologous expression of genes encoding bacterial light harvesting complexes in

rhodobacter sphaeroides. *Journal of Biological Chemistry* **270**, 23875-23882.

[47] Kiley, P.J. and Kaplan, S. (1987). Cloning, DNA sequence and expression of the rhodobacter sphaeroides light harvesting B800-850 alpha and B800-850 beta genes. *Journal of Bacteriology* **169**, 3268-3275.

[48] Lee, J.K., Wang, S.Q., Eraso, J.M., Gardner, J. and Kaplan, S. (1993). Transcriptional regulation of puc operon expression in rhodobacter sphaeroides - involvement of an integration host factor binding sequence. *Journal of Biological Chemistry* **268**, 24491-24497.

[49] Brotosudarmo, T.H.P., Collins, A.M., Gall, A., Roszak, A.W., Gardiner, A.T., Blankenship, R.E. and Cogdell, R.J. (2011). The light intensity under which cells are grown controls the type of peripheral light-harvesting complexes that are assembled in a purple photosynthetic bacterium. *Biochemical Journal* **440**, 51-61.

[50] Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., Tabita, F.R., Gibson, J.L., Hanson, T.E., Bobst, C., Torres, J., Peres, C., Harrison, F.H., Gibson, J., and Harwood, C.S. (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nature Biotechnology* **22**, 55-61.

[51] Weissgerber, T., Zigann, R., Bruce, D., Chang, Y.J., Detter, J.C., Han, C., Hauser, L., Jeffries, C.D., Lano, M., Munk, A.C., Tapia, R., and Dahl, C. (2011) Complete genome sequence of Allochromatium vinosum DSM 180 (T) *Standards in Genomic Sciences* **5**, 3:311-330

[52] Gardiner, A.T., Cogdell, R.J. and Takaichi, S. (1993). The effect of growth conditions on the light harvesting apparatus in rhodopseudomonas acidophila. *Photosynthesis Research* **38**, 159-167.

[53] Niedzwiedzki, D.M., Bina, D., Picken, N., Honkanen, S., Blankenship, R.E., Holten, D. and Cogdell, R.J. (2012). Spectroscopic studies of two spectral variants of light-harvesting complex 2 (LH2) from the photosynthetic purple sulfur bacterium Allochromatium vinosum. *Biochimica Et Biophysica Acta-Bioenergetics* **1817**, 1576-1587.

[54] Prince, S.M., Papiz, M.Z., Freer, A.A., McDermott, G., HawthornthwaiteLawless, A.M., Cogdell, R.J. and Isaacs, N.W. (1997). Apoprotein structure in the LH2 complex from Rhodopseudomonas acidophila strain 10050: Modular assembly and protein pigment interactions. *Journal of Molecular Biology* **268**, 412-423.

[55] McLuskey, K., Prince, S.M., Cogdell, R.J. and Isaacs, N.W. (2001). The crystallographic structure of the B800-820 LH3 light-harvesting complex from the purple bacteria Rhodopseudomonas acidophila strain 7050. *Biochemistry* **40**, 8783-8789.

[56] Cherezov, V., Clogston, J., Papiz, M.Z. and Caffrey, M. (2006). Room to move: crystallizing membrane proteins in swollen lipidic mesophases. *Journal of Molecular Biology* **357**, 1605-18.

[57] Sundstrom, V., Pullerits, T. and van Grondelle, R. (1999). Photosynthetic light-harvesting: Reconciling dynamics and structure of purple bacterial LH2 reveals function of photosynthetic unit. *Journal of Physical Chemistry B* **103**, 2327-2346.

[58] Karrasch, S., Bullough, P.A. and Ghosh, R. (1995). The 8.5-A projection map of the light harvesting complex 1 from rhodospirillum rubrum reveals a ring composed of 16 subunits. *Embo Journal* **14**, 631-638.

[59] Qian, P., Hunter, C.N. and Bullough, P.A. (2005). The 8.5 angstrom projection structure of the core RC-LH1-PufX dimer of Rhodobacter sphaeroides. *Journal of Molecular Biology* **349**, 948-960.

[60] Suzuki, H., Hirano, Y., Kimura, Y., Takaichi, S., Kobayashi, M., Miki, K. and Wang, Z.-Y. (2007). Purification, characterization and crystallization of the core complex from thermophilic purple sulfur bacterium Thermochromatium tepidum. *Biochimica Et Biophysica Acta-Bioenergetics* **1767**, 1057-1063.

[61] Saijo, S., Sato, T., Kumasaka, T., Tanaka, N., Harata, K. and Odahara, T. (2005). Crystallization and preliminary X-ray studies on the reaction center-light-harvesting 1 core complex from Rhodopseudomonas viridis. *Acta Crystallographica Section F Structural Biology Crystal Communication* **61**, 83-6.

[62] Francia, F., Wang, J., Venturoli, G., Melandri, B.A., Barz, W.P. and Oesterhelt, D. (1999). The reaction center-LH1 antenna complex of Rhodobacter sphaeroides contains one PufX molecule which is involved in dimerization of this complex. *Biochemistry* **38**, 6834-6845.

[63] Jungas, C., Ranck, J.L., Rigaud, J.L., Joliot, P. and Vermeglio, A. (1999). Supramolecular organization of the photosynthetic apparatus of Rhodobacter sphaeroides. Embo Journal 18, 534-542.

[64] Fidai, S., Kalmar, G.B., Richards, W.R. and Borgford, T.J. (1993). Recombinant expression of the pufQ gene of Rhodobacter capsulatus. *Journal of Bacteriology* **175**, 4834-4842.

[65] Farchaus, J.W., Gruenberg, H. and Oesterhelt, D. (1990). Complementation of a reaction center-deficient rhodobacter sphaeroides pufLMX delection strain in trans with pufBALM does not restore the photosynthetic positive phenotype. *Journal of Bacteriology* **172**, 977-985.

[66] Lilburn, T.G., Haith, C.E., Prince, R.C. and Beatty, J.T. (1992). Pleiotropic effects of pufX gene deletion on the structure and function of rhodobacter capsulatus. *Biochimica Et Biophysica Acta* **1100**, 160-170.

[67] Recchia, P.A., Davis, C.M., Lilburn, T.G., Beatty, J.T., Parkes-Loach, P.S., Hunter, C.N. and Loach, P.A. (1998). Isolation of the PufX protein from Rhodobacter capsulatus and Rhodobacter sphaeroides: Evidence for its interaction with the alpha-polypeptide of the core light-harvesting complex. *Biochemistry* **37**, 11055-11063.

[68] Barz, W.P., Francia, F., Venturoli, G., Melandri, B.A., Vermeglio, A. and Oesterhelt, D. (1995). Role of pufX protein in photosynthetic growth of rhodobacter sphaeroides 1. pufX is required for efficient light-driven electron transfer and photophosphorylation under anaerobic conditions. *Biochemistry* **34**, 15235-15247.

[69] Nagashima, S., Shimada, K., Matsuura, K. and Nagashima, K.V.P. (2002). Transcription of three sets of genes coding for the core light-harvesting proteins in the purple sulfur bacterium, Allochromatium vinosum. *Photosynthesis Research* **74**, 269-280.

[70] Farchaus, J.W. and Oesterhelt, D. (1989). A rhodobacter sphaeroides puf L,M and X deletion mutant and its complementation in trans with a 5.3 Kb puf operon shuttle fragment. *Embo Journal* **8**, 47-54.

[71] Klug, G. and Cohen, S.N. (1988). Pleiotropic effects of localised rhodobacter capsulatus puf operon deletions on production of light absorbing pigment-protein complexes. *Journal of Bacteriology* **170**, 5814-5821.

[72] Barz, W.P., Vermeglio, A., Francia, F., Venturoli, G., Melandri, B.A. and Oesterhelt, D. (1995). Role of the pufX protein in photosynthetic growth of

rhodobacter sphaeroides 2. PufX is required for efficient ubiquinone ubiquinol exchange between the reaction center Q(B) site and the cytochrome bc(1) complex. *Biochemistry* **34**, 15248-15258.

[73] Parkes-Loach, P.S., Law, C.J., Recchia, P.A., Kehoe, J., Nehrlich, S., Chen, J. and Loach, P.A. (2001). Role of the core region of the PufX protein in inhibition of reconstitution of the core light-harvesting complexes of Rhodobacter sphaeroides and Rhodobacter capsulatus. *Biochemistry* **40**, 5593-5601.

[74] Jamieson, S.J., Wang, P.Y., Qian, P., Kirkland, J.Y., Conroy, M.J., Hunter, C.N. and Bullough, P.A. (2002). Projection structure of the photosynthetic reaction centre-antenna complex of Rhodospirillum rubrum at 8.5 angstrom resolution. *Embo Journal* **21**, 3927-3935.

[75] Tunnicliffe, R.B., Ratcliffe, E.C., Hunter, C.N. and Williamson, M.P. (2006). The solution structure of the PufX polypeptide from Rhodobacter sphaeroides. *Febs Letters* **580**, 6967-6971.

[76] Ratcliffe, E.C. et al. (2011). Experimental evidence that the membrane-spanning helix of PufX adopts a bent conformation that facilitates dimerisation of the Rhodobacter sphaeroides RC-LH1 complex through N-terminal interactions. *Biochimica Et Biophysica Acta-Bioenergetics* **1807**, 95-107.

[77] Crouch, L.I., Holden-Dye, K. and Jones, M.R. (2010). Dimerisation of the Rhodobacter sphaeroides RC-LH1 photosynthetic complex is not facilitated by a GxxxG motif in the PufX polypeptide. *Biochimica Et Biophysica Acta-Bioenergetics* **1797**, 1812-1819.

[78] Semchonok, D.A., Chauvin, J.-P., Frese, R.N., Jungas, C. and Boekema, E.J. (2012). Structure of the dimeric RC-LH1-PufX complex from Rhodobaca bogoriensis investigated by electron microscopy. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 3412-3419.

[79] Aird, A., Wrachtrup, J., Schulten, K. and Tietz, C. (2007). Possible pathway for ubiquinone shuttling in Rhodospirillum rubrum revealed by molecular dynamics simulation. *Biophysical Journal* **92**, 23-33.

[80] Codgell, R.J., Southall, J., Gardiner, A.T., Law, C.J., Gall, A., Roszak, A.W. and Isaacs, N.W. (2006). How purple photosynthetic bacteria harvest solar energy. *Comptes Rendus Chimie* **9**, 201-206.

[81] Scholes, G.D. (2003). Long-range resonance energy transfer in molecular systems. *Annual Review of Physical Chemistry* **54**, 57-87.

[82] Dexter, D.L. (1953). A theory of sensitized luminescence in solids. *Journal of Chemical Physics* **21**, 836-850.

[83] Scholes, G.D., Jordanides, X.J. and Fleming, G.R. (2001). Adapting the Forster theory of energy transfer for modeling dynamics in aggregated molecular assemblies. *Journal of Physical Chemistry B* **105**, 1640-1651.

[84] Scholes, G.D., Fleming, G.R., Olaya-Castro, A. and van Grondelle, R. (2011). Lessons from nature about solar light harvesting. *Nature Chemistry* **3**, 763-774.

[85] Chachisvilis, M., Kuhn, O., Pullerits, T. and Sundstrom, V. (1997). Excitons in photosynthetic purple bacteria: Wavelike motion or incoherent hopping? *Journal of Physical Chemistry B* **101**, 7275-7283.

[86] van Oijen, A.M., Ketelaars, M., Kohler, J., Aartsma, T.J. and Schmidt, J. (1999). Unraveling the electronic structure of individual photosynthetic pigment-protein complexes. *Science* **285**, 400-402.

[87] Polivka, T. and Frank, H.A. (2010). Molecular factors controlling photosynthetic light harvesting by carotenoids. *Accounts of Chemical Research* **43**, 1125-34.

[88] Tavan, P. and Schulten, K. (1987). Electronic excitations in finite and infinite polyenes. *Physical Review* B **36**, 4337-4358.

[89] Hashimoto, H. et al. (2004). The very early events following photoexcitation of carotenoids. *Archives of Biochemistry and Biophysics* **430**, 61-69.

[90] Cerullo, G., Polli, D., Lanzani, G., De Silvestri, S., Hashimoto, H. and Cogdell, R.J. (2002). Photosynthetic light harvesting by carotenoids: Detection of an intermediate excited state. *Science* **298**, 2395-2398.

[91] Polli, D., Cerullo, G., Lanzani, G., De Silvestri, S., Yanagi, K., Hashimoto, H. and Cogdell, R.J. (2004). Conjugation length dependence of internal conversion in carotenoids: Role of the intermediate state. *Physical Review Letters* **93** 16

[92] Cerullo, G., Manzoni, C., Luer, L. and Polli, D. (2007). Time-resolved methods in biophysics. 4. Broadband pump-probe spectroscopy system with sub-20 fs temporal resolution for the study of energy transfer processes in photosynthesis. *Photochemical & Photobiological Sciences* **6**, 135-144.

[93] Polli, D., Antognazza, M.R., Brida, D., Lanzani, G., Cerullo, G. and De Silvestri, S. (2008). Broadband pump-probe spectroscopy with sub-10-fs resolution for probing ultrafast internal conversion and coherent phonons in carotenoids. *Chemical Physics* **350**, 45-55.

[94] Herek, J.L., Wohlleben, W., Cogdell, R.J., Zeidler, D. and Motzkus, M. (2002). Quantum control of energy flow in light harvesting. *Nature* **417**, 533-535.

[95] Sashima, T., Koyama, Y., Yamada, T. and Hashimoto, H. (2000). The 1B(u)(+), 1B(u)(-), and 2A(g)(-) energies of crystalline lycopene, beta-carotene, and mini-9-beta-carotene as determined by resonance-Raman excitation profiles: Dependence of the 1B(u)(-) state energy on the conjugation length. *Journal of Physical Chemistry* B **104**, 5011-5019.

[96] Polli, D., Cerullo, G., Lanzani, G., De Silvestri, S., Hashimoto, H. and Cogdell, R.J. (2006). Carotenoid-Bacteriochlorophyll Energy Transfer in LH2 Complexes Studied with 10-fs Time Resolution. Biophysical journal 90, 2486-2497. *Biophysical Journal* **90** 7:2486-2497

[97] Jonas, D.M. (2003). Two-dimensional femtosecond spectroscopy. *Annual Review of Physical Chemistry* **54**, 425-463.

[98] Schlau-Cohen, G.S., Ishizaki, A. and Fleming, G.R. (2011). Two-dimensional electronic spectroscopy and photosynthesis: Fundamentals and applications to photosynthetic light-harvesting. *Chemical Physics* **386**, 1-22.

[99] Carpenter, E.P., Beis, K., Cameron, A.D. and Iwata, S. (2008). Overcoming the challenges of membrane protein crystallography. Current Opinion in *Structural Biology* **18**, 581-586.

[100] Prive, G.G. (2007). Detergents for the stabilization and crystallization of membrane proteins. Methods 41, 388-397.

[101] Bill, R.M., Henderson, P.J.F., Iwata, S., Kunji, E.R.S., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C.G. and Vogel, H. (2011). Overcoming barriers to membrane protein structure determination. *Nature Biotechnology* **29**, 335-340.

[102] Hilgemann, D.W. and Ball, R. (1996). Regulation of cardiac Na+,Ca2+ exchange and K-ATP potassium channels by PIP2. *Science* **273**, 956-959.

[103] Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G.W., Tate, C.G., and Schertler, G.F.X. (2008). Structure of a beta(1)-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-U2.

[104] Kang, H.J., Lee, C. and Drew, D. (2013). Breaking the barriers in membrane protein crystallography. *International Journal of Biochemistry & Cell Biology* **45**, 636-644.

[105] Sonoda, Y., Newstead, S., Hu, N., Alguel, Y., Nji, E., Beis, K., Yashiro, S., Lee, C., Leung, J., Cameron, A.D., Byrne, B., Iwata, S., and Drew, D. (2011). Benchmarking Membrane Protein Detergent Stability for Improving Throughput of High-Resolution X-ray Structures. *Structure* **19**, 17-25.

[106] Newstead, S., Ferrandon, S. and Iwata, S. (2008). Rationalizing alpha-helical membrane protein crystallization. *Protein Science* **17**, 466-472.

[107] Sonoda, Y., Cameron, A., Newstead, S., Omote, H., Moriyama, Y., Kasahara, M., Iwata, S. and Drew, D. (2010). Tricks of the trade used to accelerate high-resolution structure determination of membrane proteins. *Febs Letters* **584**, 2539-2547.

[108] Chayen, N.E. (1998). Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallographica Section D-Biological Crystallography* **54**, 8-15.

[109] McPherson, A. (1990). Current approaches to macromolecular crystallisation. European Journal of *Biochemistry* **189**, 1-23.

[110] Cherezov, V. (2011). Lipidic cubic phase technologies for membrane protein structural studies. *Current Opinion in Structural Biology* **21**, 559-566.

[111] Caffrey, M. (2009) Crystallizing Membrane Proteins for Structure Determination: Use of Lipidic Mesophases. In *Annual Review of Biophysics* ed.^eds), pp. 29-51

[112] Bose, S.K. (1963) Media for anaerobic growth of photosynthetic bacteria, *Antioch Press, Yellow Springs, Ohio*

[113] Pfennig, N. (1969).Rhodopseudomonas acidophila, SP N, A new species of budding purple nonsulphur bacteria. *Journal of Bacteriology* **99**, 597-&.

[114] Eichler, B. and Pfennig, N. (1988). A new purple sulphur bacterium from stratified fresh water lakes, Amoebacter purpureus SP-NOV. *Archives of Microbiology* **149**, 395-400.

[115] Heckman, K.L. and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols* **2**, 924-932.

[116] Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. (1989). Engineering hybrid genes without the use of restriction enzymes - gene splicing by overlap extension. *Gene* **77**, 61-68.

[117] Schafer, A., Tauch, A., Jager, W., Kalinowski, J., Thierbach, G. and Puhler, A. (1994). Small mobilizable multipurpose cloning vectors derived from the Escherichia coli plasmids pk18 and pk19 - selection of defined deletions in the chromosome of corynebacterium glutamicum. *Gene* **145**, 69-73.

[118] Simon, R., Priefer, U. and Puhler, A. (1983). A broad host range mobilization system for in vivo genetic engineering transposon mutagenesis in gram negative bacteria. *Bio-Technology* **1**, 784-791.

[119] Sambrook, J., Russell, D.W. (2001). Molecular cloning, a laboratory manual. *Cold Spring Harbor Laboratory*

[120] Hernandez, D., Francois, P., Farinelli, L., Osteras, M. and Schrenzel, J. (2008). De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Research* **18**, 802-809.

[121] Mardis, E.R. (2008) Next-generation DNA sequencing methods. *In Annual Review of Genomics and Human Genetics ed.^eds)*, pp. 387-402

[122] Darling, A.E., Mau, B. and Perna, N.T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **5**, e11147.

[123] Ostroumov, E.E., Mulvaney, R.M., Cogdell, R.J. and Scholes, G.D. (2013). Broadband 2D Electronic Spectroscopy Reveals a Carotenoid Dark State in Purple Bacteria. *Science* **340**, 52-56.

[124] van Stokkum, I.H.M., Larsen, D.S. and van Grondelle, R. (2004). Global and target analysis of time-resolved spectra. *Biochimica Et Biophysica Acta-Bioenergetics* **1657**, 82-104.

[125] Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190.

[126] Leslie, A.G.W. and Powell, H.R. (2001) Integration of macromolecular diffraction data. In *Methods in Macromolecular Crystallography* (Turk, D. and Johnson, L., ed.^eds), pp. 61-71

[127] Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D. and Eisenberg, D. (2006). Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences* **103**, 8060-8065.

[128] Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D-Biological Crystallography* **62**, 72-82.

[129] Chae, P.S. et al. (2010). Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nature Methods* **7**, 1003-U90.

[130] Chen, X.J.S., Casini, G., Harrison, S.C. and Garcea, R.L. (2001). Papillomavirus capsid protein expression in Escherichia coli: Purification and assembly of HPV11 and HPV16 L1. *Journal of Molecular Biology* **307**, 173-182.

[131] Rohman, M. and Harrison-Lavoie, K.J. (2000). Separation of copurifying GroEL from glutathione-S-transferase fusion proteins. *Protein Expression and Purification* **20**, 45-7.

[132] McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674.

[133] Nogi, T., Fathir, I., Kobayashi, M., Nozawa, T. and Miki, K. (2000). Crystal structures of photosynthetic reaction center and high-potential iron-sulfur protein from Thermochromatium tepidum: Thermostability and electron transfer. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13561-13566.

[134] Nogi, T., Fathir, I., Kobayashi, M., Nozawa, T. and Miki, K. (2000). 1EYS: Crystal structure of the photosynthetic reaction center from a thermophilic bacterium Thermochromatium tepidum. *Protein Data Bank*

[135] Nogi, T., Kobayashi, M., Nozawa, T. and Miki, K. (2000). Crystallization and preliminary crystallographic analysis of the high-potential iron-sulfur protein from Thermochromatium tepidum. *Acta Crystallographica Section D-Biological Crystallography* **56**, 656-658.

[136] Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* **67**, 355-367.

[137] Clayton, R.K. and Haselkorn, R. (1972). Protein components of bacterial photosynthetic membranes. *Journal of Molecular Biology* **68**, 97-105.

[138] Leman, J.K., Mueller, R., Karakas, M., Woetzel, N. and Meiler, J. (2013). Simultaneous prediction of protein secondary structure and transmembrane spans. *Proteins: Structure, Function, and Bioinformatics* **81**, 1127-1140.

[139] von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A., and Bork, P. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Research* **33**, D433-7.

[140] Kelley, L.A. and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363-71.

[141] Schafer, A., Tauch, A., Jager, W., Kalinowski, J., Thierbach, G. and Puhler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**, 69-73.

[142] Boehm, P.S., Southall, J., Cogdell, R.J. and Koehler, J. (2013). Single-Molecule Spectroscopy on RC-LH1 Complexes of Rhodopseudomonas acidophila Strain 10050. *Journal of Physical Chemistry B* **117**, 3120-3126.

[143] Deinum, G., Otte, S.C.M., Gardiner, A.T., Aartsma, T.J., Cogdell, R.J. and Amesz, J. (1991). Antenna organisation of rhodopseudomonas acidophila - a study of the excitation migration. *Biochimica Et Biophysica Acta* **1060**, 125-131.

[144] Metzker, M.L. (2010). Applications of next generation sequencing - Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**, 31-46.

[145] Simmons, A.E., Barrett, S.J., Hunter, C.N. and Cogdell, R.J.R. (2000). Cloning, sequencing and analysis of the pucC genes from Rubrivivax gelatinosus strain 151 and Rhodopseudomonas acidophila strain 10050. *Photosynthesis Research* **65**, 69-82.

[146] Tichy, H.V., Albien, K.U., Gadon, N. and Drews, G. (1991). Analysis of the rhodobacter capsulatus puc operon - the pucC gene plays a central role in the regulation of LH2 (B800-850 complex) expression. *Embo Journal* **10**, 2949-2955.

[147] Fowler, G.J.S., Visschers, R.W., Grief, G.G., Vangrondelle, R. and Hunter, C.N. (1992). Genetically modified photosynthetic antenna complexes with blueshifted absorbency bands. *Nature* **355**, 848-850.

[148] Zuber, H.B., RA. (1993). The chlorophylls. *CRC Press, Boca Raton* (Scheer, H, ed) 627-704.

[149] Roszak, A.W., Moulisova, V., Reksodipuro, A.D.P., Gardiner, A.T., Fujii, R., Hashimoto, H., Isaacs, N.W. and Cogdell, R.J. (2012). New insights into the structure of the reaction centre from Blastochloris viridis: evolution in the laboratory. *Biochemical Journal* **442**, 27-37.

[150] Frank, H.A., Bautista, J.A., Josue, J., Pendon, Z., Hiller, R.G., Sharples, F.P., Gosztola, D. and Wasielewski, M.R. (2000). Effect of the solvent environment on the spectroscopic properties and dynamics of the lowest excited states of carotenoids. *Journal of Physical Chemistry B* **104**, 4569-4577.