
Article:
Jackson, P.J., Hitchcock, A., Swainsbury, D.J.K. orcid.org/0000-0002-0754-0363 et al. (6 more authors) (2018) Identification of protein W, the elusive sixth subunit of the Rhodopseudomonas palustris reaction center-light harvesting 1 core complex. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1859 (2). pp. 119-128. ISSN 0006-3002

https://doi.org/10.1016/j.bbabio.2017.11.001

Identification of protein W, the elusive sixth subunit of the Rhodopseudomonas palustris reaction center-light harvesting 1 core complex

Philip J. Jackson, Andrew Hitchcock, David J.K. Swainsbury, Pu Qian, Elizabeth C. Martin, David A. Farmer, Mark J. Dickman, Daniel P. Canniffe, C. Neil Hunter

PII: S0005-2728(17)30160-3
DOI: doi:10.1016/j.bbabio.2017.11.001
Reference: BBABIO 47847

To appear in:

Received date: 25 July 2017
Revised date: 3 November 2017
Accepted date: 6 November 2017

Please cite this article as: Philip J. Jackson, Andrew Hitchcock, David J.K. Swainsbury, Pu Qian, Elizabeth C. Martin, David A. Farmer, Mark J. Dickman, Daniel P. Canniffe, C. Neil Hunter , Identification of protein W, the elusive sixth subunit of the Rhodopseudomonas palustris reaction center-light harvesting 1 core complex. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Bbabio(2017), doi:10.1016/j.bbabio.2017.11.001

Identification of protein W, the elusive sixth subunit of the *Rhodopseudomonas palustris* reaction center-light harvesting 1 core complex

Philip J. Jackson[a,b,¶], Andrew Hitchcock[a,¶], David J.K. Swainsbury[a], Pu Qian[a], Elizabeth C. Martin[a], David A. Farmer[a], Mark J. Dickman[b], Daniel P. Canniffe[a,1] & C. Neil Hunter[a,*]

[a] Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK

[b] ChELSI Institute, Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, UK

[1] Present address: Department of Biochemistry & Molecular Biology, The Pennsylvania State University, USA

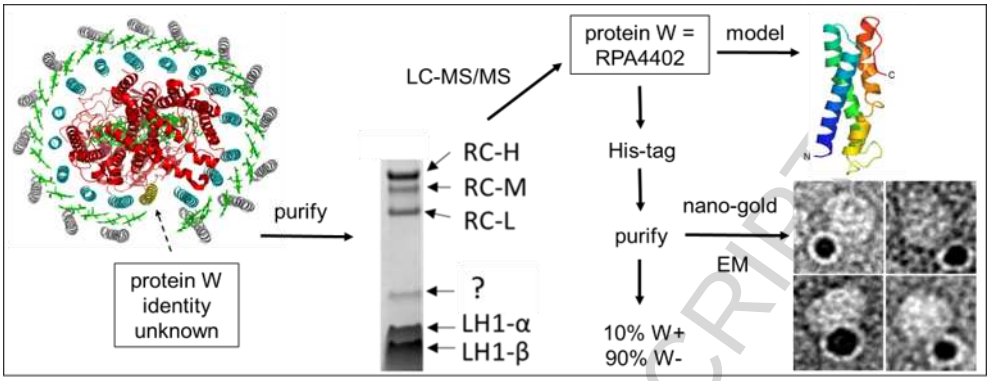
[¶] Both authors contributed equally to this work

* Corresponding author: Email: c.n.hunter@sheffield.ac.uk. Tel: +44 114 222 4191. Fax: +44 114 222 2711

Abstract The X-ray crystal structure of the *Rhodopseudomonas* (*Rps.*) *palustris* reaction center-light harvesting 1 (RC-LH1) core complex revealed the presence of a sixth protein component, variably referred to in the literature as helix W, subunit W or protein W. The position of this protein prevents closure of the LH1 ring, possibly to allow diffusion of ubiquinone/ubiquinol between the RC and the cytochrome bc_1 complex in analogous fashion to the well-studied PufX protein from *Rhodobacter sphaeroides*. The identity and function of helix W have remained unknown for over 13 years; here we use a combination of biochemistry, mass spectrometry, molecular genetics and electron microscopy to identify this protein as RPA4402 in *Rps. palustris* CGA009. Protein W shares key conserved sequence features with PufX homologs, and although a deletion mutant was able to grow under photosynthetic conditions with no discernible phenotype, we show that a tagged version of protein W pulls down the RC-LH1 complex. Protein W is not encoded in the photosynthesis gene cluster and our data indicate that only approximately 10% of wild-type *Rps. palustris* core complexes contain this non-essential subunit; functional and evolutionary consequences of this observation are discussed. The ability to purify uniform RC-LH1 and RC-LH1-protein

W preparations will also be beneficial for future structural studies of these bacterial core complexes.

Graphical abstract



Highlights

- Identification of the protein W subunit of the *Rps. palustris* RC-LH1 core complex.
- The rpa4402 locus encoding protein W is not in the PGC.
- Protein W is present in only a sub-population of core complexes.
- Protein W is dispensable for photosynthetic growth.
- Pure plus/minus protein W core complex preparations will aid structural studies.

Keywords: Photosynthesis, *Rhodopseudomonas palustris*, reaction center-light harvesting 1 (RC-LH1) core complex, helix W, RPA4402, PufX.

Abbreviations: reaction center-light harvesting 1 (RC-LH1); *Rps*. (*Rhodopseudomonas*); *Rba*. (*Rhodobacter*); *Rsp*. (*Rhodospirillum*); matrix-assisted laser desorption/ionisation-time-of-flight (MALDI-TOF) mass spectrometry (MS); overlap-extension polymerase chain reaction (OLE-PCR); bacteriochlorophyll (BChl); *n*-Dodecyl β-D-maltoside (β-DDM); column volumes (CV); immobilised metal affinity chromatography (IMAC); ion-exchange chromatography (IEC); size exclusion chromatography (SEC); atomic force microscopy (AFM); photosynthesis gene cluster

(PGC); transmembrane helix (TMH); ATP-binding cassette (ABC); transmission electron microscopy (TEM).

1. Introduction

Purple phototrophic bacteria perform anoxygenic photosynthesis where, following photon absorption by bacteriochlorophyll (BChl) pigments in the light harvesting (LH) complexes, excitation energy is passed to the special pair BChls in the reaction center (RC). Subsequently, cyclic electron flow between the RC, cytochrome bc_1 complex and cytochrome *c*, generates a transmembrane proton gradient that is used to power ATP synthesis [1]. In the photosynthetic membranes of these bacteria, the peripheral antenna and LH1 complexes surround the RC in an arrangement optimized for efficient energy capture and transfer; the encirclement of the RC by the LH1 antenna is termed the RC-LH1 core complex [2]. The core complexes of some purple bacterial species such as *Rhodobacter* (*Rba*.) *sphaeroides* and *Rba. capsulatus* contain a PufX polypeptide. The 8 Å structure of the dimeric RC-LH1-PufX complex of *Rba. sphaeroides* shows that PufX associates with the extrinsic domain of the RC-H subunit, and with the N-termini of LH1 α1 and β1 on the cytoplasmic side of the membrane, forming a pore in the LH1 ring to permit shuttling of ubiquinone and ubiquinol between the RC and the cytochrome bc_1 complex [3-6].

The 4.8 Å resolution X-ray crystal structure of the *Rps. palustris* core complex showed that the RC is surrounded by an incomplete elliptical LH1 ring consisting of 15 pairs of transmembrane helical α- and β-apoproteins [7]. This arrangement is in contrast to the *Rhodospirillum* (*Rsp*.) *rubrum* and *Thermochromatium* (*T*.) *tepidum* core complexes, which have a complete closed ring of 16 LH1 αβ-heterodimers around the RC [8-10]. In *Rps. palustris* transmembrane helix W prevents closure of the LH1 ring round the RC and LH1 consists of 15 LH1 αβ-heterodimers [7] (Figure 1). The location of helix W and the resulting break in the LH1 outer ring are positioned adjacent to the secondary electron acceptor ubiquinone (UQ_B) binding site in the RC [7]. High-resolution atomic force microscopy (AFM) images of native *Rps. palustris* membranes also reveal a topography gap in the elliptical LH1 assembly around the RC, which the authors attribute to the W subunit [11]. The same study additionally showed that the location of this gap is random with respect to the imposed RC axis,

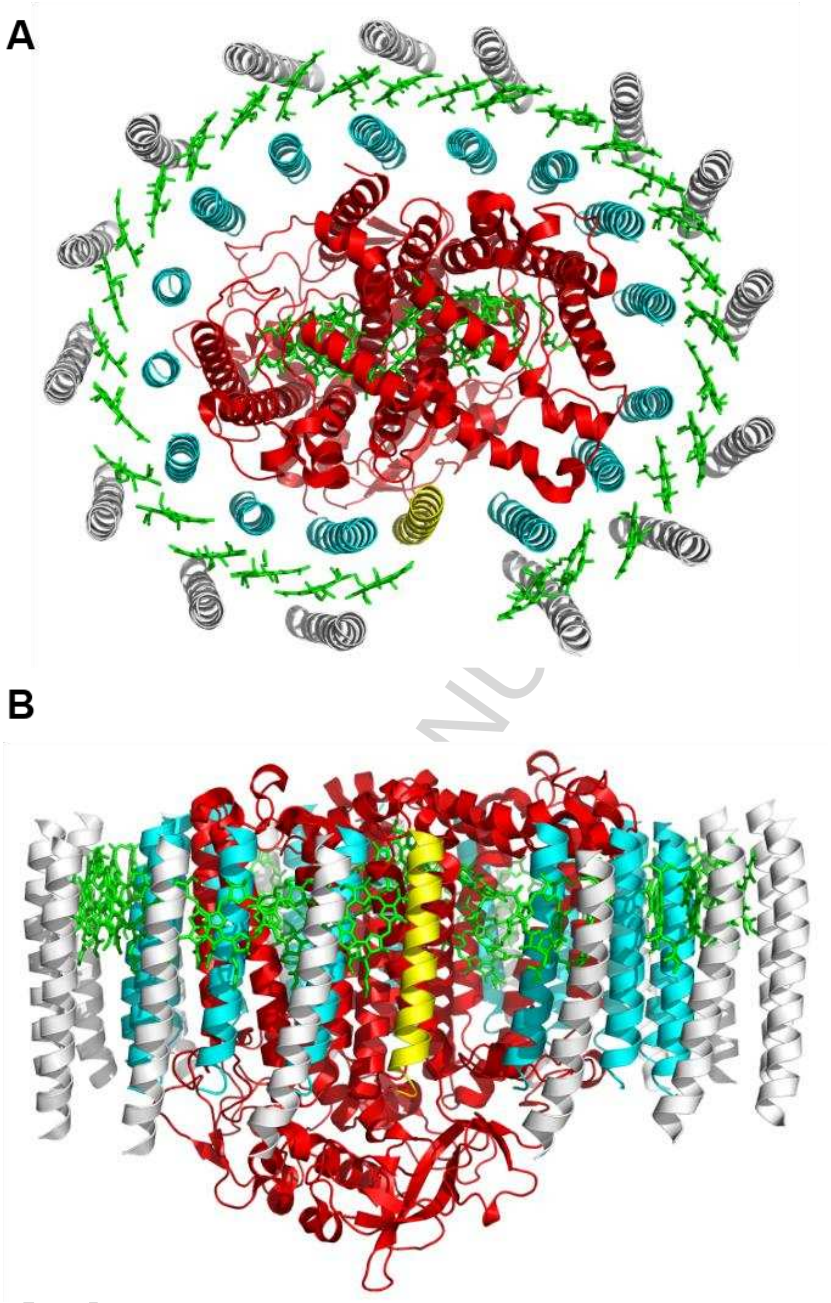


Figure 1. X-ray crystal structure of the *Rps. palustris* core complex showing the position of the W subunit. This figure was made using the data from the 4.8 Å resolution RC-LH1 structure of Roszak *et al*. [7] (PDB ID: 1PYH). **(A)** Schematic model of the RC-LH1 core complex viewed perpendicular to the membrane plane along the pseudo-twofold-axis of the RC and **(B)** as a narrow section viewed parallel to the membrane plane with the cytoplasmic face at the bottom. Transmembrane helices are depicted as ribbons and bacteriochlorophylls/bacteriopheophytins are represented as their respective macrocycles and are coloured green. Core complex subunits are shown in red, LH1 α subunits in blue, LH1 β subunits in white, and the W subunit in yellow.

rather than being restricted to the periapsis of the core complex as seen in the X-ray structure. Whether protein W performs a PufX-like role in allowing ubiquinone/ubiquinol exchange or has another function is open to debate.

Matrix-assisted laser desorption/ionisation-time-of-flight (MALDI-TOF) mass spectrometry (MS) of purified *Rps. palustris* RC-LH1 complexes revealed a sixth protein component with a mass of 10,707 Da [7]. However, this protein was resistant to Edman sequencing and so the identity of the helix W subunit remained unknown. Here, we purified the RC-LH1 core complex from the genome sequenced CGA009 strain of *Rps. palustris* and, in addition to the LH1 polypeptides and the RC L (light), M (medium) and H (heavy) subunits, we observed a sixth protein on stained SDS-PAGE gels. Using a combination of in-solution and in-gel digestion followed by nano-flow liquid chromatography coupled on-line to a mass spectrometer we identified this protein as RPA4402. RPA4402 is a good candidate for protein W, as its predicted molecular mass is similar to that of the protein identified in the *Rps. palustris* 2.1.6 core complex by Roszak *et al.* [7]. Deletion of rpa4402 did not result in significant retardation of either microoxic growth in the dark or photoheterotrophic growth, consistent with the result that the tagged protein pulls down only a small proportion of the total core complexes from purified membranes, suggesting only a sub-population of core complexes contain protein W.

2. Materials and methods

2.1. Growth of Rhodopseudomonas palustris

Rps. palustris CGA009 was grown under microoxic conditions in the dark at 30 ^{o}C with orbital shaking (150 rpm) in either PYE (5 g.l^{-1} each of peptone, yeast extract and succinate) or M22+ with 0.1% casamino acids [12] media supplemented with 20 µg.ml^{-1} chloramphenicol. For phototrophic growth, screw-capped, airtight vessels (ranging in size from 17 ml - 8 L depending on the experiment) completely filled with M22+ medium were illuminated with 30-50 µmol.m^{-2}.s^{-1} white light provided by Osram 116 W halogen bulbs and cultures were grown at room temperature with agitation using magnetic stir bars. Growth was monitored as the optical density at

680 nm (OD_{680}) against a medium blank. For growth on plates media contained 1.5% agar and 80 µg.ml^{-1} kanamycin and/or 10% sucrose as appropriate.

2.2. Strain generation

Strains, primers and plasmids used in this study are listed in Tables 1, S1 and S2 respectively. PCRs were performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs, UK). To delete all but the start and stop codons of the putative protein W encoding rpa4402 open reading frame, approximately 400 bp fragments of upstream and downstream flanking DNA were amplified by PCR with primer pairs W_KO_Up_F/R and W_KO_Down_F/R respectively. These fragments were joined by overlap-extension (OLE)-PCR with primer pair W_KO_Up_F/W_KO_Down_R resulting in an approximately 1.0 kb product, which was digested and cloned into the *Bam*HI and *Hin*dIII sites of the allelic exchange suicide vector pK18mob*sacB* generating plasmid pK18mob*sacB*-WKO. Conjugative transfer of the plasmid from *Escherichia coli* S17-1 to *Rps. palustris* was performed as described previously for *Rba. sphaeroides* [13] with the exception that selection of first transconjugants was performed with 80 µg.ml^{-1} kanamycin on PYE agar. Successful generation of strain Δrpa4402 was confirmed using PCR with primers W_check_F/R and automated DNA sequencing (GATC Biotech, Germany). OLE-PCR was used to generate constructs for adding an N- or C-terminal 10×His tag to RPA4402. To add an N-terminal tag, three PCR products amplified with primer pairs W_KO_Up_F/R, W_N-His_F/R and W_N-His_Down_F/W_KO_Down_R were joined using primers W_KO_Up_F/W_KO_Down_R. For C-terminal tagging the three PCR products were amplified with W_KO_Up_F/R, W_C-His_F/R and W_C-His_Down_F/W_KO_Down_R. OLE-PCR products were cloned into the *Bam*HI and *Hin*dIII sites of pK18mob*sacB* generating plasmids pK18mob*sacB*-HisW and pK18mob*sacB*-WHis, which were transferred to *Rps. palustris* as above. Sequence verification of the tagged strains His-RPA4402 and RPA4402-His was performed as for Δrpa4402.

Table 1. Strains used in this study.

Species and strain	Description	Source/Reference
Escherichia coli		
JM109	Cloning strain.	Promega, UK
S17-1	Conjugative strain for transfer of plasmids to *Rps. palustris.*	[14]
Rhodopseudomonas palustris		
Wild-type	Strain CGA009.	Prof. D. J. Kelly*
Δrpa4402	Unmarked deletion of rpa4402.	This study
His-RPA4402	Sequence encoding 10xHis tag added in frame at the start of rpa4402.	This study
RPA4402-His	Sequence encoding 10xHis tag added in frame at the end of rpa4402.	This study

*Department of Molecular Biology and Biotechnology, University of Sheffield, UK.

2.3. Absorbance spectroscopy and deconvolution of spectra

Absorbance spectra of *Rps. palustris* cells, membranes and core complexes were acquired at room temperature using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies) with appropriate media or buffer baseline correction. Spectra were deconvoluted in Microsoft Excel by fitting spectra of purified RC-LH1 complexes (see below) or peripheral light-harvesting complexes (a by-product of the RC-LH1 purification) to the spectra collected at various stages of preparation to estimate their RC-LH1 content. This was achieved by adaptation of a spreadsheet available online at https://terpconnect.umd.edu/~toh/spectrum/CurveFittingB.html.

2.4. RC-LH1 core complex purification

Photosynthetically grown *Rps. palustris* cells were resuspended in 20 mM HEPES-NaOH pH 7.8 (buffer A) and broken by three passages through a French pressure cell at 18,000 psi. The lysate was applied directly to a 20/40% (w/w) discontinuous sucrose gradient, which was centrifuged at 100,000 × *g* at 4 ^{o}C for 4 h.

Membranes were harvested from the 20/40% interface, diluted 2-fold with buffer A and pelleted by centrifugation at 235,000 x *g* for 60 min. The pellet was re-suspended in buffer A to OD_{875} = 60 and solubilized in 3% *n*-dodecyl β-D-maltoside (β-DDM) at 4 ^{o}C for 30 min in the dark with stirring. Non-solubilized material was pelleted at 211,000 x *g* for 60 min and the supernatant applied to a 20/21.25/22.50/23.75/25% (w/w) discontinuous sucrose gradient in buffer A containing 0.03% β-DDM. After centrifugation at 125,000 x *g* for 16 h the core complex band was harvested and applied to a 2 x 10 cm DEAE-Sepharose ion exchange chromatography (IEC) column (GE Healthcare) pre-equilibrated with buffer A plus 0.03% β-DDM and run with a linear gradient of 0-400 mM NaCl. The core complex, which eluted at ~250 mM NaCl, was concentrated to 1 ml using a Centriprep YM-50 ultrafiltration device (Millipore) and further purified by size-exclusion chromatography (SEC) on a Superdex 200 16/60 column (GE Healthcare) at 0.4 $ml.min^{-1}$.

2.5. In-solution trypsin digestion

Isolated core complex was buffer-exchanged into 100 mM triethylammonium bicarbonate pH 8.5 (TEAB, Sigma-Aldrich) using a centrifugal ultrafiltration device (Microcon Ultracel YM-10, Millipore) and concentrated to 2 $g.l^{-1}$ protein (Bradford assay). 15 µl aliquots were solubilized separately in (1) 0.2% ProteaseMax surfactant (Promega), 5 mM DTT, total volume 20 µl, with incubation at 56 °C for 20 min; (2) 1% SDS, 5 mM DTT, total volume 20 µl, with incubation at 37 °C for 30 min; (3) 60% methanol, total volume 29 µl, with vortexing at room temperature for 30 sec. For digestion, 1.2 µg trypsin (porcine, modified, MS grade, Sigma-Aldrich) was added in (1) 60 µl, (2) 180 µl and (3) 1.2 µl 100 mM TEAB followed by overnight incubation at 37 °C. Digestion (1) was treated with the addition of 0.5% TFA followed by incubation at room temperature for 5 min to hydrolyse the ProteaseMax surfactant. This sample was then desalted using a home-made Poros R2 spin column [15]. Digestion (2) was processed to remove SDS according to Hollingshead *et al*. [15] and digestion (3) was dried by vacuum centrifugation.

2.6. In-gel trypsin and pepsin digestion

Purified core complex proteins were precipitated using a 2-D clean-up kit (GE Healthcare) following the manufacturer's protocol prior to analysis by SDS-PAGE. Protein bands, stained with Coomassie Blue, were excised and subjected to trypsin digestion according to Pandey *et al*. [16]. For in-gel pepsin digestion, this method was modified by substituting 0.2% formic acid as the post-alkylation wash and digestion solvent.

2.7. Protein identification by nano-flow LC-MS/MS

Digested samples were redissolved in 0.1% TFA, 3% acetonitrile and analysed by nano-flow liquid chromatography (Ultimate 3000 system, Thermo Scientific) coupled on-line to a mass spectrometer (Amazon ion trap, Bruker or Q Exactive HF quadrupole-orbitrap, Thermo Scientific). For protein identification, Amazon MS data-files were first converted to Mascot Generic File (MGF) format using a script provided by Bruker. Amazon MGF and Q Exactive RAW data-files were uploaded into Byonic v. 2.9.38 (Protein Metrics) for searching against the Uniprot reference proteome database for *Rps. palustris* (ID: UP000001426, 4811 entries, 9July 2016).

2.8. His-tagged RPA4402 pulldowns

Phototrophically grown *Rps. palustris* cells (N- or C-terminal His-tagged strains) were resuspended in 20 mM Tris-HCl pH 8.0 (buffer B) and broken by two passages through a French pressure cell at 18,000 psi. The cell lysate was applied directly to a 15/40 % (w/w) discontinuous sucrose gradient, which was centrifuged at 27,000 rpm (~85,000 × *g*) in a Beckman Type 45 Ti rotor at 4 ^{o}C for 10 h. Membranes were harvested from the 15/40% interface and solubilized in 2% β-DDM in 50 ml buffer B containing 200 mM NaCl at 4 ^{o}C for 1 hr in the dark. The solubilized membranes were applied to a 5 ml HisTrap HP column (GE Healthcare) pre-equilibrated with 5 column volumes (CV) of buffer B containing 200 mM NaCl, 20 mM imidazole and 0.03% β-DDM (IMAC buffer) at a flow rate of 5 ml.min^{-1} and the flow-through was retained (see below). The column was washed sequentially with 10 CV buffer B with 20 mM then 40 mM imidazole. Protein that remained bound was eluted with 3 CV buffer B plus 250 mM imidazole and collected as 1 ml fractions. RC-LH1 containing fractions were pooled, concentrated to 1 ml (as above) and injected onto a Superdex 200 16/60 SEC

column pre-equilibrated with buffer B containing 200 mM NaCl and 0.03% β-DDM. Protein was eluted over 150 ml at a flow rate of 0.8 ml.min^{-1} and fractions enriched in RC-LH1 with a ratio of their absorbance at 880 and 803 nm (A880:A803) greater than 3.5 were retained for further analysis.

Core complexes not pulled down by His-tagged RPA4402 were purified from the flow-through by IEC on a 50 ml DEAE-Sepharose column pre-equilibrated with buffer B plus 0.03% β-DDM (DEAE buffer). The flow-through was diluted four-fold with DEAE buffer to reduce [NaCl] to <50 mM and applied to the column at 5 ml.min^{-1}. The column was washed with 2 CV DEAE buffer followed by 5 CV DEAE buffer plus 50 mM NaCl. RC-LH1 complexes were eluted by a linear gradient of 50 to 250 mM NaCl whilst collecting 10 ml fractions. Fractions containing RC-LH1 complexes were diluted to reduce [NaCl] to <50 mM and reapplied to the column for two further rounds of enrichment. RC-LH1 containing fractions with an A880:803 ratio of ≥2 were pooled and further purified by SEC as described above.

2.9. Immunoblotting

To detect the His-tag on RPA4402, immunoblots with an anti-6-His primary antibody (A190-114A, Bethyl Laboratories, USA) were performed and visualized following the method described in Grayson *et al*. [17], with modifications as follows: proteins were transferred to methanol-activated polyvinylidene difluoride membranes in carbonate transfer buffer (10 mM $NaHCO_3$, 3 mM Na_2CO_3, 10%) methanol); membranes were blocked in 50 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.2 % Tween-20 and 5% skimmed milk powder; antibodies were diluted in 50 mM Tris-HCl pH 7.6, 150 mM NaCl and 0.05% Tween-20.

2.10. Ni-NTA nanogold labelling of core complexes

Immobilised metal affinity chromatography (IMAC) purified RC-LH1 complexes (0.083 µM) were incubated in a three-fold molar excess of Ni-NTA nanogold (0.25 µM) (Nanoprobes, USA) in the dark for 1 hr at room temperature in a total volume of 1 ml in 20mM Tris-HCl pH 8, 200 mM NaCl and 0.03 % w/v β-DDM. Excess nanogold was removed by dilution to 15 ml in the same buffer followed by re-concentration to ~250 µl in a 100,000 MWCO spin filter (Millipore) three times.

2.11. Negative stain transmission electron microscopy

Nanogold labelled complexes were diluted to an OD_{875} of 0.3 and 5 µl was loaded onto freshly glow-discharged carbon coated copper grids. The grids were rinsed twice in distilled water and once in 0.75% (w/v) uranyl acetate solution (UA). Grids were stained by incubation for 2 min in UA, dried under vacuum and imaged in a Philips CM100 transmission electron microscope at 52,000x magnification. Images were processed and cropped using ImageJ [18].

2.12. Bioinformatics and structural modelling

Homologs of RPA4402 were identified using BLAST [19] and sequence alignments were performed using CLUSTAL Omega [20]. The TMHMM Server v.2.0 [21] was used to predict transmembrane helices (TMHs). Hydrophobicity profiles were performed using the ProtScale/Transmembrane Tendency application in www.expasy.org [22]. Structural models of RPA4402 were generated using the QUARK [23] and Robetta [24] servers.

3. Results

3.1. Mass spectrometric identification of RPA4402 as a candidate for helix W

Photosynthetic membranes were isolated from WT *Rps. palustris* cells by rate zonal centrifugation. RC-LH1 core complexes were then purified after a further rate zonal centrifugation step followed by anion exchange and size exclusion chromatography. The absorbance spectra in Figure 2A show the expected reduction in the LH2/3 B800 signal and concomitant increase in the B875 peak in the purified core complexes. The difference in absorbance profiles around 500 nm reflect the carotenoid composition ± LH2/3 complexes. This preparation was digested with trypsin by three different in-solution methods. Analysis of the resultant peptides by nano-flow LC-MS/MS and database searching enabled the identification of the abundant, well-characterized components: the three reaction center subunits (L, M and H) and two LH1 polypeptides (α and β), as summarized in Table S3. Of the two surfactant-based methods ProteaseMax, which is specifically designed to enhance

trypsin efficiency, identified all five proteins while LH1-α was not identified after digestion in SDS. Proteolysis in 60% methanol however, enabled the detection of peptides derived from additional proteins, presumably minor contaminants of the preparation, together with an uncharacterized protein RPA4402 as a significant component of the sample. Analysis of the core complex by SDS-PAGE, shown in Figure 2B, revealed a protein band migrating above the LH1 polypeptides, which was identified by in-gel digestion and nano-flow LC-MS/MS as comprising RPA4402 as its highest-scoring constituent (Table S3). Parallel analysis of core complex from an rpa4402 deletion strain (see Figure 2C and section 3.5) at 2x protein loading showed the absence of a band at the RPA4402 position (Figure 2D); in-gel digestion and LC-MS/MS of this region of the gel revealed only minor contaminants, degradation products of RC-H and a presumed LH1-αβ heterodimer (Table S3). RPA4402 has a predicted molecular mass of 10,498 Da, close to the 10,707 Da size of the protein identified in the core complex preparation from *Rps. palustris* 2.1.6 [7]. The other non-RC-LH1 proteins identified by LC-MS/MS (Table S3) either have known functions (e.g. AtpG (RPA0177) and PetB (RPA1193)), are too big (> 175 amino acids/18 kDa) to be candidates for protein W (RPA3961, RPA1961, RPA4760 and RPA2303) or do not have any predicted transmembrane helices (RPA1495) (Table S4).

PufX in *Rhodobacter* spp. is truncated in vivo by the excision of both M1 and the C-terminal 12 (*Rba. sphaeroides*) or 9 (*Rba. capsulatus*) amino acids [25], decreasing the length to 69 or 68 residues, respectively. The rpa4402 gene encoding 102 amino acids would give a potential length difference of around 30 residues compared to processed PufX, so we investigated the possibility that truncation of RPA4402 could generate a functional PufX ortholog. Tryptic digestion of RPA4402 only produces one internal proteotypic peptide with a significant score, shown in Figure S1A, consistent with its observed resistance to this protease under aqueous solution conditions (see above). However, sequence coverage of RPA4402 was increased to four peptides by in-gel pepsin digestion to confirm an intact C-terminus (Figure S1B). Coverage could not be extended to the N-terminus beyond V13 because pepsin (and chymotrypsin) would generate fragments from this part of the sequence that are too small for detection in proteomic analysis. Nevertheless, the results show that

RPA4402 is at least 90 amino acids in length and, in contrast to PufX, it is not truncated at the C-terminus.

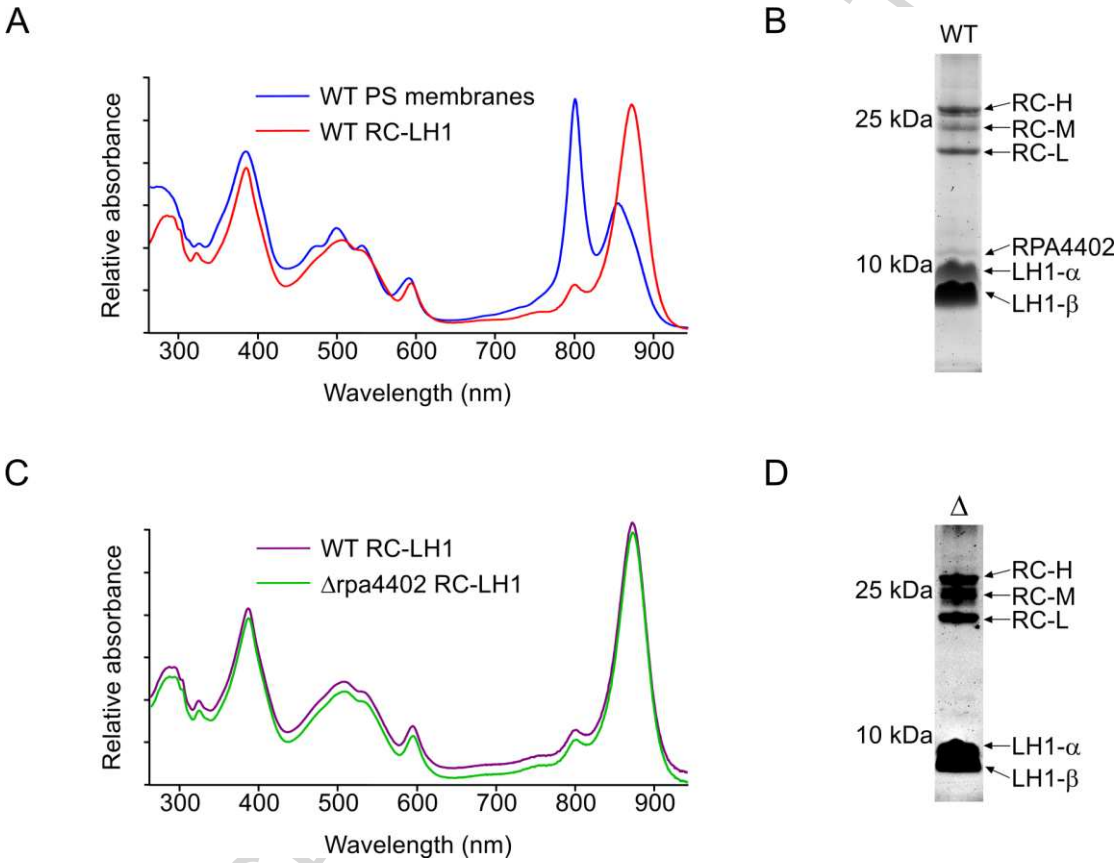


Figure 2. Analyses of photosynthetic membranes and RC-LH1 core complexes by spectroscopy and SDS-PAGE. (A) Room temperature absorbance spectra of wild-type (WT) photosynthetic (PS) membranes (blue) and isolated RC-LH1 complexes (red). **(B)** SDS-PAGE of WT RC-LH1 complexes (10 µg protein) stained with Sypro Ruby (inverted fluorescence image). **(C)** Room temperature absorbance spectra of WT (purple) and Δrpa4402 strain (green) RC-LH1 complexes. **(D)** SDS-PAGE of Δrpa4402 strain RC-LH1 complexes (20 µg protein) stained with Sypro Ruby (inverted fluorescence image).

3.2. RPA4402 is conserved in Rps. palustris strains

Unlike the *pufX* gene of *Rhodobacter* spp., rpa4402 is not located within the *Rps. palustris* photosynthesis gene cluster (PGC), but in a region of the genome occupied by ATP-binding cassette (ABC) transporter subunits (Figure 3). Examination of this same region in six other other *Rps. palustris* strains revealed orthologs of rpa4402, and sequence alignments (Figure 4) show identities of 65-99% with RPA4402, indicating both the genomic location and sequence of protein W are highly conserved in *Rps. palustris* spp.

3.3. RPA4402 is associated with only a sub-population of RC-LH1 core complexes

To further demonstrate that RPA4402 is a component of the core complex, solubilized membranes from strains in which the protein was tagged with a 10xHis tag at either the N- (His-RPA4402) or C-terminus (RPA4402-His) (Figure S2) were used to try to co-purify RC-LH1. Both strains grew comparably to the WT under phototrophic conditions (Figure S3). When membranes from the RPA4402-His strain were applied to an IMAC column, a purple band formed at the top of the column and remained bound throughout the wash steps until the protein was eluted (Figure 5A). Deconvolution of the absorbance spectra of the eluate versus the flow-through using the spectra of the individual purified components allowed us to estimate that approximately 10% of the total RC-LH1 complexes were pulled down by the His-tag on RPA4402 (Figure S4). Immunodetection of the His-tag on RPA4402 showed the protein was present in the load and elution but absent from the flow-through (Figure 5B), and SDS-PAGE analysis of the eluate (Figure 5C) shows a band at approximately the correct predicted molecular weight (11.9 kDa) for RPA4402-His. RPA4402-His remained associated with RC-LH1 complexes purified further by SEC (Figure 5D). Notably, RPA4402-His was clearly visible by Coomassie Blue staining, whereas untagged RPA4402 in WT core complex preparations was only faintly stained by Coomassie Blue, with efficient staining requiring Sypro Ruby (Figure 2B). These data suggest that only a sub-population of core complexes contains a protein W component, and using the tagged strain selectively enriches core complexes containing RPA4402 so that it is in a 1:1 ratio with each of the RC subunits, explaining the easier visualization by Coomassie Blue staining. Consistent with this hypothesis,

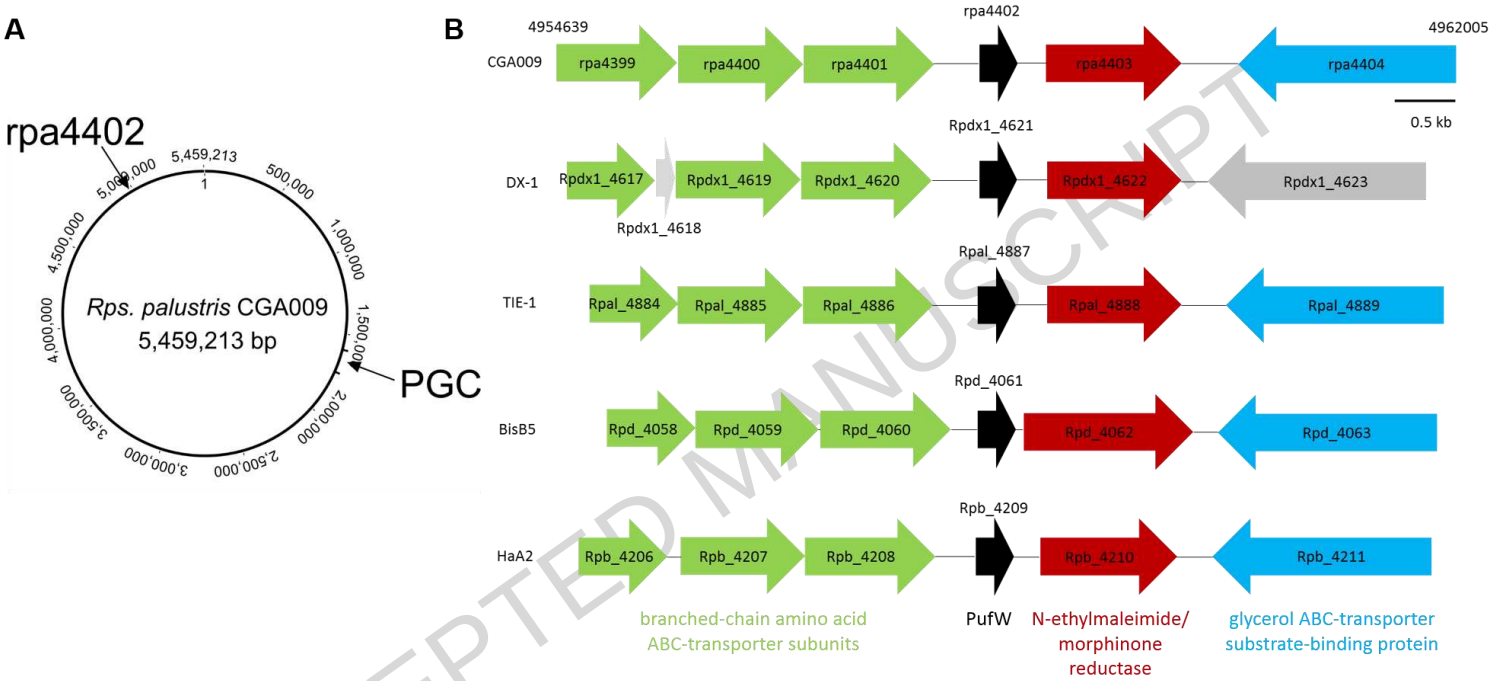


Figure 3. The rpa4402 gene is not part of *Rps. palustris* photosynthesis gene cluster (PGC). (A) The chromosome of *Rps. palustris* strain CGA009 [26]. The rpa4402 gene (position 4958088-4958396) and the ~55-kb PGC (rpa1505-rpa1554, position 1671137-1725322) are indicated with arrows. The scale (in bp) is indicated around the circumference of the circle. **(B)** Zoomed in genomic region (position 4954639-4962005) around the rpa4402 locus. Upstream of rpa4402, rpa4399-rpa4401 encode subunits of a branched chain amino acid ABC-transporter (green). The downstream rpa4403 and rpa4404 genes encode a putative morphinone/N-ethylmaleimide reductase (red) and the substrate-binding protein of an ABC-transporter for glycerol (blue). The corresponding region of the chromosome in other *Rps. palustris* strains (see figure 4 legend for details) reveals conservation of the rpa4402 locus. Putative pseudogenes are shown in grey.

Figure 4. Alignment of RPA4402 with homologs from other *Rhodopseudomonas spp*. Sequence identities (%) relative to *Rps. palustris* CGA009 used in this study are shown on the right, with identical amino acids (*), and conservative (:) and semi-conservative (.) replacements indicated below. The proteins included are from strains: *Rps. palustris* DX-1, *Rps. palustris* TIE-1, *Rps. palustris* BisB5, *Rps. palustris* HaA2, *Rps. palustris* DSM123 and *Rhodopseudomonas* sp. AAP120.

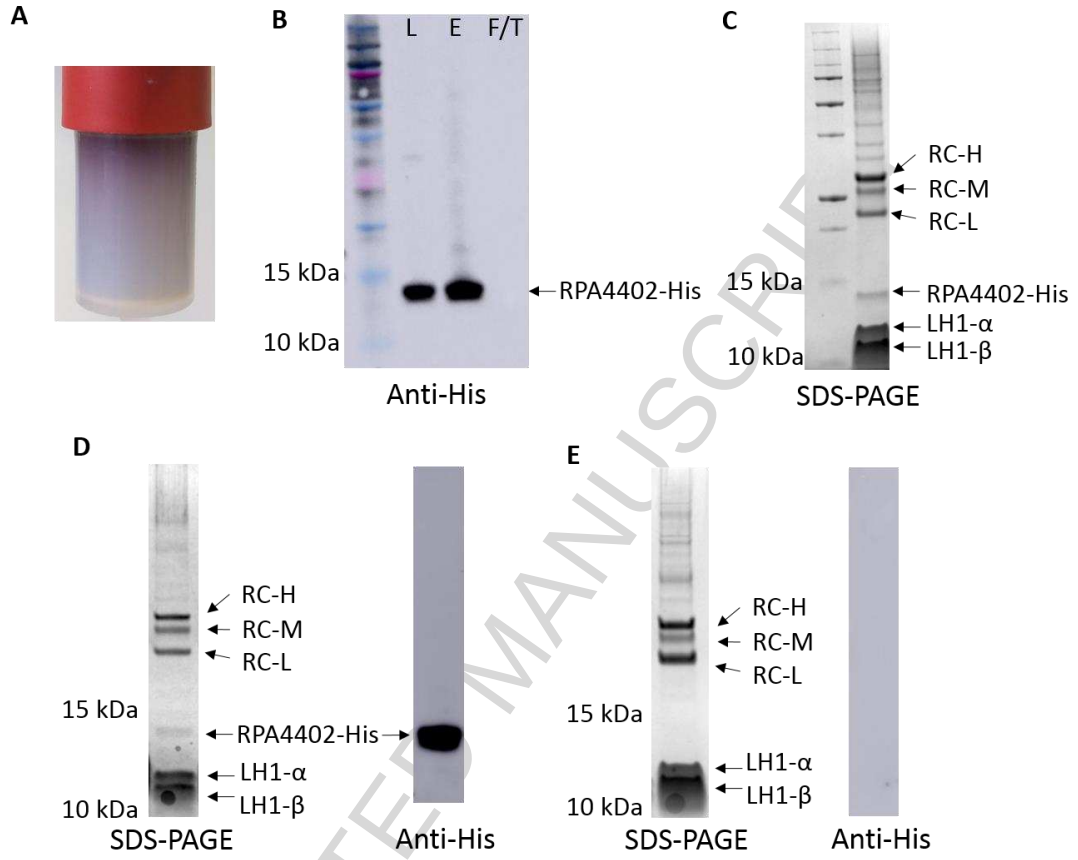


Figure 5. RPA4402-His pulls down RC-LH1. (A) The HisTrap HP IMAC column (GE Healthcare) remained coloured after loading of solubilized RPA4402-His membranes and washing with 40 mM imidazole. **(B)** Immunodetection of the His-tag shows RPA4402 is present in the load (L) and elution (E) fractions but not the flow-through (F/T). **(C)** Coomassie Blue stained SDS-PAGE of the concentrated IMAC eluate. The RC, LH1 and RPA4402 core complex subunits are labelled. **(D)** Further purification of RC-LH1-RPA4402-His by SEC. **(E)** Purification of RC-LH1 core complexes lacking RPA4402-His from the IMAC flow-through by IEC and SEC.

the remaining approximately 90% of RC-LH1 core complexes purified from the flow-through by IEC and SEC did not contain any RPA4402 when analysed by either SDS-PAGE or immunoblotting (Figure 5E).

Conversely to the situation with C-terminally His-tagged RPA4402, no coloured protein remained bound when solubilized membranes from N-terminally His-tagged RPA4402 (His-RPA4402) were applied to an IMAC column (Figure S5A), and all the RC-LH1 was present in the flow-through (determined by deconvolution of absorbance spectra as described above, data not shown). This observation indicates that the N-terminal tag either prevents production of RPA4402 or its incorporation into core complexes, or that the tagged protein is present but the tag is not available to bind to the column, or that the N-terminus is cleaved removing the tag. To investigate which of these scenarios may be true, solubilized membranes, the IMAC flow-through and eluate, and core complexes purified from the flow-through by IEC and SEC were analysed by SDS-PAGE and immunoblots probed with an anti-His antibody (Figure S5B, C). No anti-His signals corresponding to His-RPA4402 were detected in any of the IMAC fractions, and pertinently neither in the solubilized membranes. In addition, there was no evidence of an RPA4402 band on a stained SDS-gel of IEC/SEC-purified core complexes from the IMAC flow-through. Taken together, these findings imply that His-RPA4402 is not produced, otherwise it would be detectable in the stained SDS-gel either with, or after truncation without, the N-terminal His-tag.

Absorption spectra of the IMAC-binding core complexes containing RPA4402-His (Figure. 5D) and the flow-through fraction (Figure 5E) were compared after normalisation to the RC pheophytin signal at 760 nm, as shown in Figure S6. LH1 BChl values of 1.024 and 1.100 respectively imply that core complexes containing RPA4402-His have a lower LH1 BChl content than those in the IMAC flow-through. This finding therefore suggests the possibility that core complexes lacking RPA4402 have an additional LH1αβ pair.

3.4. Electron microscopy

IMAC and SEC purified RC-LH1-RPA4402-His core complexes were labelled with gold nano-beads and visualized by negative stain transmission electron microscopy (TEM). Figure 6 shows that the majority of complexes are elliptical and associated

with a single nano-bead at the periphery of the complex, which is consistent with the expected position of the His-tag on RPA4402. Measurement of the complexes gives an average diameter of 13.4 ± 1.5 nm (n = 29), the expected size of the RC-LH1 complex [7]. In the light of this evidence that RPA4402 is associated with the *Rps. palustris* core complex we henceforth refer to it as protein W.

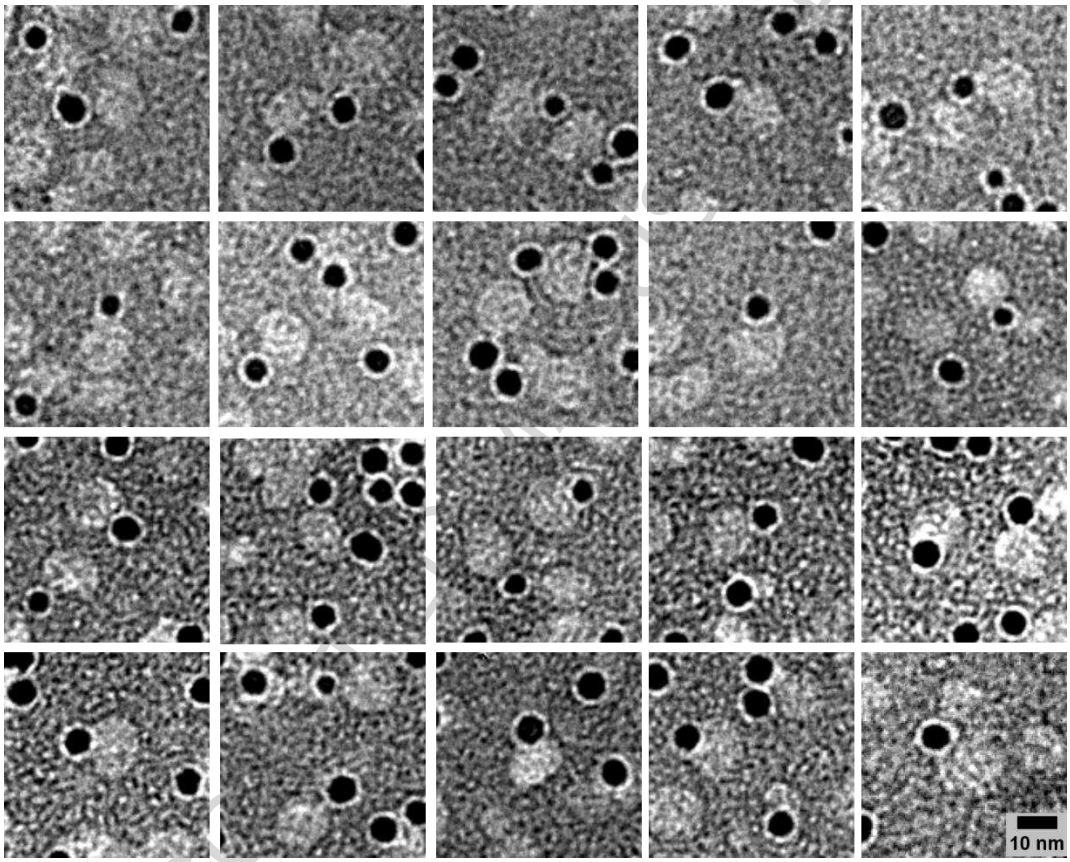


Figure 6. TEM of nanogold labelled core complexes. Core complexes labelled with 5 nm Ni-NTA Nanogold, negatively stained with uranyl acetate and imaged at 52 000 x magnification. Each panel is cropped from the wide-field images at 50 x 50 nm to show multiple objects of similar composition and an average diameter of 13.4 ± 1.5 nm (n=29).

3.5. The absence of protein W does not impair phototrophic growth of Rps. palustris

An unmarked Δrpa4402 knockout strain, in which all but the start and stop codons of the gene were deleted, was generated using the pK18mob*sacB* allelic exchange vector as described in the Methods section and Figure S2. To test whether removal of protein W from the *Rps. palustris* core complex had a deleterious effect on photosynthesis, phototrophic growth curves were performed. Growth of the mutant mimicked the isogenic WT parent (Figure 7A) and the whole cell absorbance spectra were indistinguishable between the two strains (Figure 7B). Therefore, protein W is neither required for photosynthetic competency nor formation of core complexes, at least under standard laboratory growth conditions.

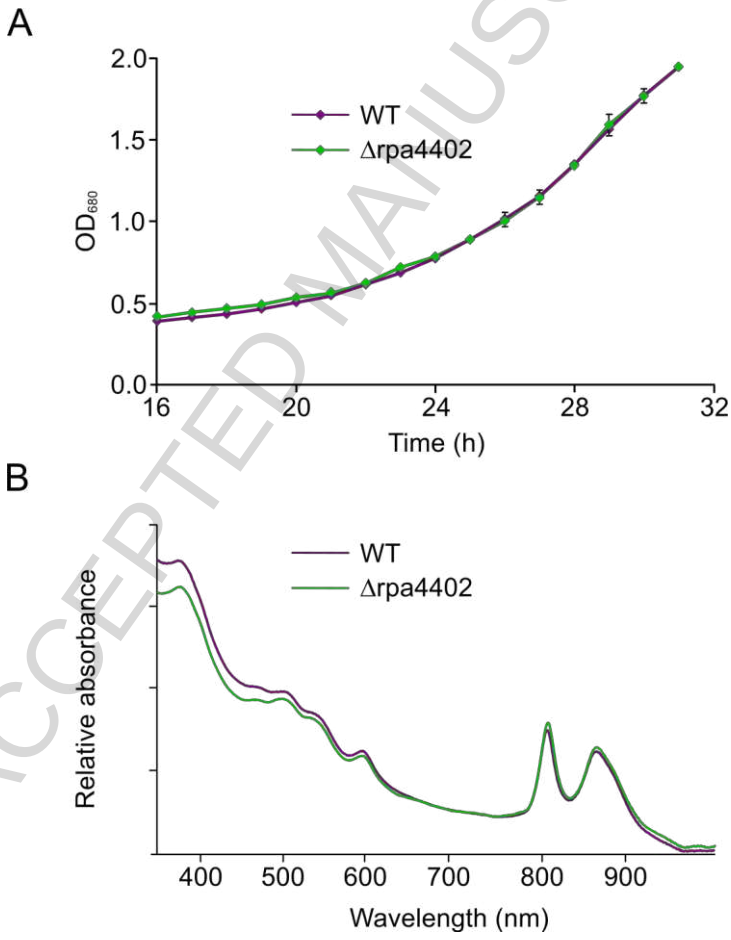


Figure 7. RPA4402 is not required for photosynthetic competence in *Rps. palustris*. **(A)** Photosynthetic growth curves of the WT (purple) and Δrpa4402 (green) strains. Error bars represent the standard deviation from the mean of three biological replicate cultures. **(B)** Absorbance spectra of the same strains measured at room temperature and normalised at 680 nm.

3.6. Sequence alignment of protein W with orthologs of PufX in relation to transmembrane helix predictions

To explore the hypothesis that protein W is related to PufX, the sequence alignment shown in Figure 8A was constructed in the format designed by Tsukatani *et al*. [27] and Holden-Dye *et al*. [28]. Although protein W has, according to its translated gene sequence, 102 amino acids while authentic purple bacterial PufX proteins have 75-83 and has no other obvious homology, the alignment reveals all six of the conserved features indicated by these authors as characteristic of PufX proteins (highlighted in red). Despite this pattern match in the sequences, a comparison of hydropathy profiles and transmembrane helix (TMH) predictions between protein W and PufX from *Rba. sphaeroides* (Figure 8B) indicates that protein W has three TMHs while PufX has one. This prediction of a single TMH for *Rba. sphaeroides* PufX, also the case for the other PufX orthologs (Figure 8A), is supported by a structural model based on both NMR [29, 30] and X-ray data [6]. The spatial relationship between the conserved amino acids and the predicted TMH arrangement shows that the MXXG motif and a G closely flank the TMH in PufX, while in protein W MXXG is integral to TMH2 and G is predicted to be five residues distant from the membrane. Furthermore, the C-terminal AA/AP/GP and P motifs occur inside TMH3 of protein W rather than residing in the periplasmic space (deduced from the X-ray model described by Qian *et al*. [6]) in PufX.

4. Discussion

Purple bacterial photosynthetic core complexes comprise a central RC complex surrounded by a ring of LH1 αβ heterodimers. In *Rhodospirillum* and *Thermochromatium* spp. for example, the LH1 ring is continuous around the RC with 16 αβ pairs [8-10]. In *Rhodobacter* spp. there is an additional component of the ring, PufX, which replaces the αβ pair adjacent to the UQ_B site in the RC and creates a gap to facilitate ubiquinone/ubiquinol exchange with the cytochrome bc_1 complex [31, 32]. A PufX-like component, referred to as protein W, was detected by X-ray crystallography of *Rps. palustris* core complex [7] (Figure 1). Refinement of the X-ray structure by single-molecule spectroscopy [33, 34] provided support for the physical gap modelled in the X-ray crystal structure and observed by AFM [11].

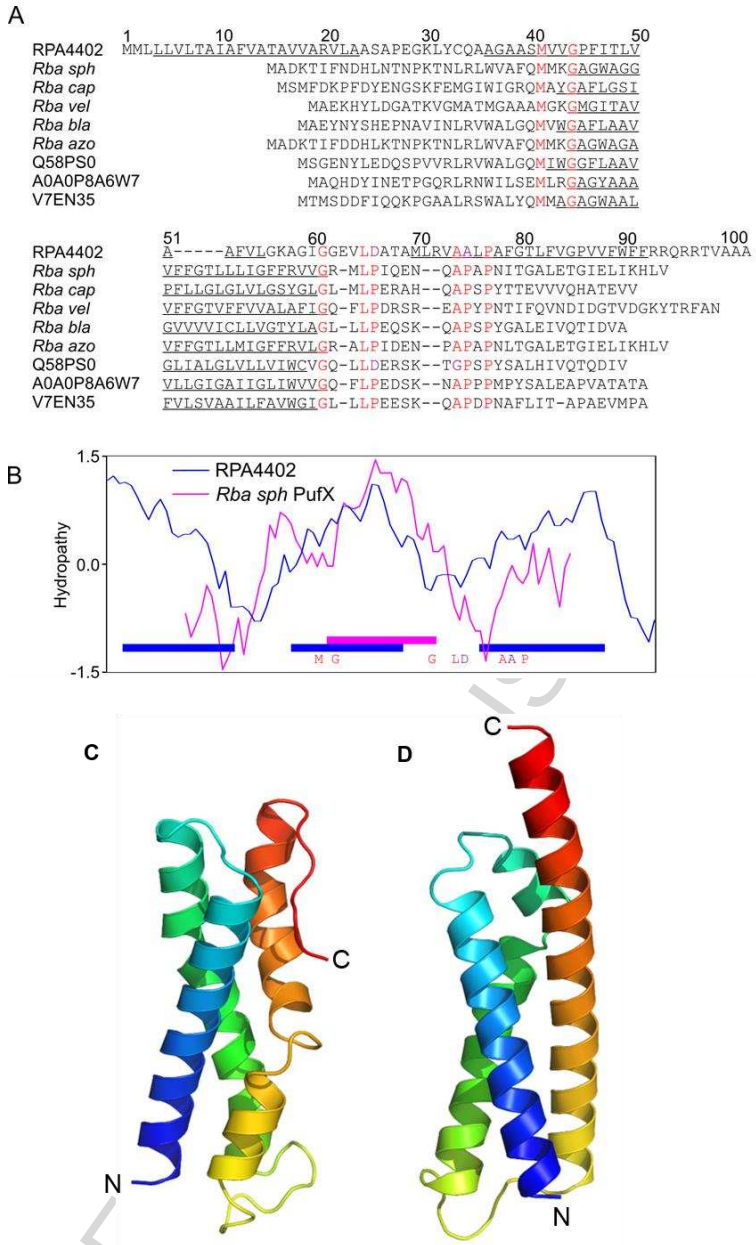


Figure 8. Sequence alignment of protein W with PufX orthologs from species of *Rhodobacteraceae*. (A) Alignment of protein W with PufX orthologs from *Rba. sphaeroides* (*Rba sph*), *Rba. capsulatus* (*Rba cap*), *Rba. veldkampii* (*Rba vel*), *Rba. blasticus* (*Rba bla*), *Rba. azotoformans* (*Rba azo*) and hypothetical membrane proteins from proteobacteria DelRiverFos13D03 (Q58PS0), HL-91 (A0A0P8A6W7) and CACIA14H1 (V7EN35). Q58PS0 is referred to as AY912081 in Holden-Dye *et al*. [28] and is included because it displays two exceptions in the identical regions (P>D and A>G, indicated in magenta), one of which is shared by protein W. A0A0P8A6W7 and V7EN35 are additional to the ones shown in Holden-Dye *et al*. [28] and were identified by BLAST searching using *Rba. sphaeroides* PufX as the input sequence. Identical residues are highlighted in red with exceptions in magenta and sequences predicted to be in transmembrane helices (TMHs) underlined. **(B)** Hydrophobicity profiles for protein W (blue) and *Rba. sphaeroides* PufX (magenta) with the bars indicating the positions of predicted TMHs. **(C-D)** Models of protein W generated by the Robetta **(C)** and QUARK **(D)** servers.

Rps. palustris 2.1.6 was used in the study of Roszak *et al*. [7] and *Rps. palustris* DSM123 by Scheuring *et al*. [11], but here we used strain CGA009 as availability of its full genome sequence [26] enabled the identification of rpa4402 as the locus encoding protein W. This may explain the 209 Da difference in MW between 2.1.6 protein W (10,707 Da) and CGA009 RPA4402/protein W (10,498 Da), considering that the MWs of the protein W orthologs in Figure 4 show variation over a range of 363 Da; as far as we are aware the genome of strain 2.1.6 has not been sequenced. Our data show that *Rps. palustris* cells appear to contain two distinct populations of core complexes, a large (~90%) subset of RC-LH1 and a smaller (~10%) sub-population RC-LH1-protein W complexes. Why only some core complexes have protein W and others do not is unknown, as is whether the protein W occupancy level is constant or variable. On the assumption that the X-ray structure determined by Roszak *et al*. [7] and the AFM topographs of *Rps. palustris* membranes [11] are comparable, our finding is unexpected. Possible explanations are (a) the *Rps. palustris* strains used in these earlier studies, unlike our CGA009 strain, express protein W at levels which enable 100% protein W occupancy, or (b) crystallization induced the enrichment of a subpopulation of RC-LH1-protein W complexes, or (c) the membrane patches imaged by AFM represented domains that were populated exclusively by protein W containing core complexes. Further studies are needed to address this variability, which may be relevant to the underlying mechanism that governs core complex assembly in *Rps. palustris*, and to determine whether core complexes that do not contain protein W have an extra LH1αβ pair, as suggested in Section 3.3 and Figure S6or simply a gap in the LH1 ring. We note that in *Rps. palustris* the gene encoding protein W is distal from the RC and LH1 genes, in contrast to the fixed stoichiometry of PufX in the core complexes of *Rba. sphaeroides* or *Rba. capsulatus*, which is a consequence of *pufX* in a single transcriptional unit along with the other genes encoding LH1 and RC polypeptides [35, 36]. The methods described for purification of both of these complexes allow preparation of homogeneous samples with or without protein W, which will also be useful for future structural biology work with the *Rps. palustris* core complex.

Deletion of *pufX* from *Rba.* species prevents photosynthetic growth [37, 38] unless LH1 is also deleted [39], or LH1 is modified by suppressor mutants or

mutagenesis to prevent ring closure [3, 40-42] or to have lowered carotenoid content [43]. Here, deletion of rpa4402 did not have any effect on the photosynthetic growth of *Rps. palustris* under our standard laboratory growth conditions; this is in agreement with rpa4402 not being identified in a Tn-seq screen of genes essential for phototrophic growth [44]. Further work is needed in order to identify conditions under which protein W is required, which will allow additional insight into its role in the core complex of this metabolically versatile photosynthetic model organism.

Sequence comparisons (Figure 8) show that protein W is significantly longer than the recognised purple bacterial PufX proteins. Also, our identification of an intact C-terminus in protein W (Figure S1B), in contrast to the C-terminal truncation seen in PufX [25], confirms that protein W is not processed to become more PufX-like in length. It is evident therefore that protein W is not a PufX ortholog in the accepted sense. Nevertheless, the conserved sequence motifs identified by Tsukatani *et al*. [27] and Holden-Dye *et al*. [28] as being a common characteristic of PufX proteins are all present in protein W (Figure 8). PufX does not have a catalytic or pigment binding function, and it is only required to bind to RCs and/or LH1 polypeptides in a way that prevents closure of the LH1 ring. Thus, there is scope for substantial divergence of sequence and even structure for PufX orthologs, especially when the PufX does not have to mediate dimerization of core complexes. For example, the identity between the PufX proteins of *Rba. sphaeroides* and *Rba. capsulatus* is only 23.5%, but a strain of *Rba. sphaeroides* containing the *Rba. capsulatus* PufX protein could grow photosynthetically [45]. Consequently, protein W in the more distantly related *Rps. palustris* might not have to bear much similarity to PufX in order to perform an analogous function.

A BLAST sequence similarity search revealed six orthologs of protein W, shown in Figure 4, all of which belong to *Rhodopseudomonas* species with sequence identities in the range 65 – 99%. Two of these homologs were assigned as sodium/hydrogen antiporters. Unexpectedly, the search results contained no other homologs even from other purple bacteria, implying that protein W is unique to this particular genus. In terms of sequence similarity, the nearest significant matches included the C-terminal 100-residue sections of several ABC transporter permease subunits. The highest scoring match, from *Microbacterium GCS4* (accession number A0A0L6HZA3) displays

36% sequence identity in the aligned C-terminal region, which also corresponds almost exactly with three predicted TMHs (Figure S7). This homology between the three-TMH protein W and the C-terminal three TMHs of a bacterial ABC transporter permease is consistent with the position of rpa4402 within a region of the *Rpa. palustris* genome containing ABC transporter subunits (Figure 3). The orthologs of rpa4402 in the other *Rhodopseudomonas* spp. are located in the genome in equivalent positions (Figure 3B), suggesting that the rpa4402 open reading frame corresponds to the C-terminal three TMH domains of an ancestral eight TMH ABC transporter permease; we translated the DNA sequence upstream of rpa4402 to the stop codon of the previous open reading frame but found no evidence of the N-terminal section of such a protein in this intergenic region. Structural models of protein W support the three TMH prediction (Figure 8C-D) but in the structure of the core complex determined by Roszak *et al*. [7] protein W is a single TMH, and we could not model a three TMH protein into the gap in the 15-membered α/β LH1 ring. However, the size of the protein identified by mass spectrometry (10,707 Da) in the study by Roszak *et al*. [7] was substantially larger than a single TMH, similar to that of RPA4402. It is therefore possible that the assembly of the RC-LH1-protein W core complex is similar to that of the *Rba. sphaeroides* RC-LH1-PufX complex, in which a pre-assembled PufX-LH1α1β1 sub-complex attaches to the RC near the Q_B site and initiates the encirclement of the RC by a further 13 LH1α1β1 heterodimers [46-49]. In the ~10% of core complexes with protein W in *Rps. palustris*, the initiator PufX-LH1α1β1 sub-complex, with a total of 3 TMHs, could be replaced by the 3 TMH protein W subunit, followed by subsequent addition of 14 LH1αβ units. This possibility is supported by the comparison of the LH1 BChl absorbance of core complexes with (1.024) and without (1.100) protein W shown in Figure S6, consistent with the elimination of one LH1αβ pair and its associated two BChls. Substitution is also evident in a Δ*pufX* strain of *Rba. sphaeroides*; the position normally occupied by PufX is taken by an additional, in this case 2 LH1αβ pairs [49]. The discrepancy between this suggested model, with 14 LH1αβ heterodimers and a 3 TMH protein W, and the published structure (15 LH1αβ heterodimers and a 1 TMH protein W) could be accommodated if a feature originally attributed to an LH1αβ pair was re-assigned as two of the TMHs of protein

W.It will be interesting to determine the structures of RC-LH1 and RC-LH1-protein W core complexes in the future.

5. Conclusion

In this study, we have developed a method for isolating *Rps. palustris* RC-LH1 core complexes in a highly purified state. Minimal interference from contaminants enabled the unambiguous identification of a sixth component, previously referred to as protein W in *Rps. palustris* strain 2.1.6, which is encoded by the rpa4402 locus in *Rps. palustris* CGA009. Structural models, based on earlier crystallographic and AFM data suggest that this protein is integrated into the LH1 ring of the core complex where it creates a channel to allow efficient transfer of ubiquinol/ubiquinone between the RC and cytochrome bc_1 complexes. Furthermore, C-terminally His-tagged protein W pulls down RC-LH1 and enables binding of nano-gold beads to core complexes. Therefore, protein W in *Rps. palustris* appears to have an equivalent function to PufX in *Rhodobacter* spp., as originally proposed by Roszak *et al*. [7].

This similarity between protein W and PufX, however, is not straightforward. The rpa4402 deletion mutant does not impair phototrophic growth or core complex assembly under the conditions used in this study, in contrast to PufX, which is obligatory for phototrophy in *Rba. sphaeroides*. This apparent expendability may be reflected in the observation that only ~10% of core complexes appear to harbour protein W. Also, as evident from the sequences and structural models of protein W and PufX, these proteins are not homologous. PufX could have evolved from an ancestral RC cytochrome subunit; the *pufC* gene resides in the photosynthesis gene cluster of some purple bacterial species [50]. Protein W, on the other hand, displays sequence identity with the C-terminal section of an ABC transporter permease, consistent with the position of rpa4402 in a region of the genome occupied by ABC transporter subunits. Despite their lack of relationship, the one structural similarity between protein W and PufX is the presence of six sequence motifs that are shared by all PufX orthologs. We propose that these motifs form recognition sites for the integration of protein W/PufX into nascent core complexes and have evolved in different purple bacterial genera by mutations in unrelated ancestral genes. While the role of *Rhodobacter* PufX in enabling phototrophic growth is now established,

further studies might elucidate the environmental conditions under which protein W becomes necessary for photosynthetic growth in *Rps. palustris*.

Acknowledgments

This work was supported by the European Research Council (Advanced Award 338895); the Biotechnology and Biological Sciences Research Council (BBSRC) UK (Award numbers BB/M000265/1, BB/M011151/1 and BB/M012166/1); the Photosynthetic Antenna Research Center (PARC), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences (Award number DE-SC0001035); and the European Commission (Marie Skłodowska-Curie Global Fellowship 660652).

Author contributions

P.J.J., D.P.C. and C.N.H. conceived the study. P.J.J., A.H., D.J.K.S., P.Q., M.J.D., D.P.C. and C.N.H. designed the experiments. P.J.J., A.H., D.J.K.S., P.Q., E.C.M., D.A.F. M.J.D. and D.P.C. performed the experiments and/or analysed the results. P.J.J., A.H., D.J.K.S. and C.N.H. wrote the manuscript.

References

[1] M. Sener, J. Strümpfer, S. Abhishek, C.N. Hunter, K. Schulten, Overall energy conversion efficiency of a photosynthetic vesicle, eLife 5 (2016) e09541.

[2] P.A. Bullough, P. Qian, C.N. Hunter, Reaction center-light-harvesting core complexes of purple bacteria, in: C.N. Hunter, F. Daldal, M.C. Thurnauer, J.T Beatty (Eds.), The Purple Phototrophic Bacteria, Springer, Dordrecht, 2008, pp 155-179.

[3] W.P. Barz, A. Verméglio, F. Francia, G. Venturoli, B.A. Melendri, D. Oesterhelt, Role of 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, (1995) 15248-15258.

[4] R. Comayras, C. Jungas, J. Lavergne, Functional consequences of the organization of the photosynthetic apparatus in *Rhodobacter sphaeroides* - I.

Quinone domains and excitation transfer in chromatophores and reaction center antenna complexes, J. Biol. Chem. 280 (2005) 11203–11213.

[5] P. Qian, P.A. Bullough, C.N. Hunter, 3-D reconstruction of a membrane-bending complex: the RC-LH1-PufX core dimer of *Rhodobacter sphaeroides*, J. Biol. Chem. 283 (2008) 14002-14011.

[6] P. Qian, M.Z. Papiz, P.J. Jackson, A.A. Brindley, I.W. Ng, J.D. Olsen, M.J. Dickman, P.A. Bullough, C.N. Hunter, Three-dimensional structure of the *Rhodobacter sphaeroides* RC-LH1-PufX complex: dimerization and quinone channels promoted by PufX, Biochemistry, 52 (2013)7575-7585.

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

[8] S. Karrasch, P.A. Bullough, R. Ghosh, The 8.5 Å projection map of the light-harvesting complex I from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits, EMBO J. 14 (1995) 631-638.

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

[10] S. Niwa, L.J. Yu, K. Takeda, Y. Hirano, T. Kawakami, Z.Y. Wang-Otomo, K. Miki, Structure of the LH1-RC complex from *Thermochromatium tepidum* at 3.0 Å, Nature 508 (2014) 228–232.

[11] S. Scheuring, R.P. Gonçalves, V. Prima, J.N. Sturgis, The photosynthetic apparatus of *Rhodopseudomonas palustris*: structures and organization, J. Mol. Biol. 358 (2006) 83-96.

[12] C.N. Hunter, G. Turner, Transfer of genes coding for apoproteins of reaction center and light-harvesting LH1 complexes to *Rhodobacter sphaeroides*, Microbiology 134 (1988) 1471-1480.

[13] A. Hitchcock, P.J. Jackson, J.W. Chidgey, M.J. Dickman, C.N. Hunter, D.P. Canniffe, Biosynthesis of chlorophyll *a* in a purple bacterial phototroph and assembly into a plant chlorophyll-protein complex, ACS Synth. Biol. 5 (2016) 948-954.

[14] R. Simon, U. Priefer, A. Pühler, A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram-negative bacteria, Nature Biotechnol. 1 (1983) 784–791.

[15] S. Hollingshead, J. Kopecná, P.J. Jackson, D.P. Canniffe, P.A. Davison, M.J. Dickman, R. Sobotka, C.N. Hunter, Conserved chloroplast open-reading frame ycf54 is required for activity of the magnesium protoporphyrin monomethylester oxidative cyclase in *Synechocystis* PCC 6803, J. Biol. Chem. 287 (2012) 27823-27833.

[16] A. Pandey, J.S. Anderson, M. Mann, Use of mass spectrometry to study signaling pathways, Sci. STKE. 2000 (2000) pl1.

[17] K.J. Grayson, K.M. Faries, X. Huang, P. Qian, P. Dilbeck, E.C. Martin, A. Hitchcock, C. Vasilev, J.M. Yuen, D.M. Niedzwiedzki, G. J. Leggett, D. Holten, C. Kirmaier, C.N. Hunter, Augmenting light coverage for photosynthesis through YFP-enhanced charge separation at the *Rhodobacter sphaeroides* reaction centre, Nature. Commun. 8 (2017) 13972.

[18] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to Image J: 25 years of image analysis, Nature Methods 9 (2012).671-675.

[19] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389-3402.

[20] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539.

[21] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567-580.

[22] G. Zhou, E. London, An amino acid “transmembrane tendency” scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: relationship to biological hydrophobicity, Prot. Sci. 15 (2006) 1987-2001.

[23] D. Xu, Y. Zhang, *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field, Proteins 80 (2012) 1715-1735.

[24] D.E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server, Nucleic Acids Res. 32 (2004) W526-531.

[25] P.S. Parkes-Loach, C.J. Law, P.A. Recchia, J. Kehoe, S. Nehrlich, J. Chen, P.A. Loach, 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 (2001) 5593-5601.

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

[27] Y. Tsukatani, K. Matsuura, S. Masuda, K. Shimada, A. Hiraishi, K.V. Nagashima, Phylogenetic distribution of unusual triheme to tetraheme cytochrome subunit in the reaction center complex of purple photosynthetic bacteria, Photosynth. Res. 79 (2004) 83-91.

[28] K. Holden-Dye, L.I. Crouch, M.R. Jones, Structure, function and interactions of the PufX protein, Biochim. Biophys. Acta. 1777 (2008) 613-630.

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

[30] E.C. Ratcliffe, R.B. Tunnicliffe, I.W. Ng, P.G. Adams, P.Qian, K. Holden-Dye, M.R. Jones, M.P. Williamson, C.N. Hunter, 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, Biochim. Biophys. Acta 1807 (2011) 95-107.

[31] J. Lavergne, A. Verméglio, P. Joliot, Functional coupling between reaction centers and cytochrome bc_1 complexes, in: C.N. Hunter, F. Daldal, M.C. Thurnauer, J.T Beatty (Eds.), The Purple Phototrophic Bacteria, Springer, Dordrecht, 2008, pp 509-536.

[32] M.L. Cartron, J.D. Olsen, M. Sener, P.J. Jackson, A.A. Brindley, P. Qian, M.J. Dickman, G.J. Leggett, K. Schulten, C.N. Hunter, Integration of energy and electron transfer processes in the photosynthetic membrane of *Rhodobacter sphaeroides*. Biochim. Biophys. Acta 1837 (2014) 1769-1780.

[33] M.F. Richter, J. Baier, T. Prem, S. Oellerich, F. Francia, G. Venturoli, D. Oesterhelt, J. Southall, R.J. Cogdell, J. Köhler, Symmetry matters for the electronic structure of core complexes from *Rhodopseudomonas palustris* and *Rhodobacter sphaeroides* PufX^{-}, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 6661-6665.

[34] M.F. Richter, J. Baier, J. Southall, R.J. Cogdell, S. Oellerich, J. Köhler, Refinement of the X-ray structure of the RC LH1 core complex from *Rhodopseudomonas palustris* by single-molecule spectroscopy, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 20280-20284.

[35] G Klug, S.N. Cohen, Pleiotropic effects of localized *Rhodobacter capsulatus puf* operon deletions on production of light-absorbing pigment–protein complexes, J. Bact. 170 (1988) 5814–5821.

[36] J.K. Lee, B.S. DeHoff, T.J. Donohue, R.I. Gumport, S. Kaplan, Transcriptional analysis of *puf* operon expression in *Rhodobacter sphaeroides* 2.4.1. and an intercistronic transcription terminator mutant, J. Biol. Chem. 264 (1989) 19354–19365.

[37] J.W. Farchaus, H. Gruenberg, D. Oesterhelt, Complementation of a reaction center-deficient *Rhodobacter sphaeroides pufLMX* deletion strain *in trans* with *pufBALM* does not restore the photosynthesis-positive phenotype, J. Bacteriol. 172 (1990) 977-985.

[38] J.W. Farchaus, W.P. Barz, H. Grunberg, D. Oesterhelt, Studies on the expression of the PufX polypeptide and its requirement for photoheterotrophic growth in *Rhodobacter sphaeroides*, EMBO J. 11 (1992) 2779-2788.

[39] P. McGlynn, C.N. Hunter, M.R. Jones, The *Rhodobacter sphaeroides* PufX protein is not required for photosynthetic competence in the absence of a light harvesting system, FEBS Lett. 349 (1994) 349-353.

[40] T.G. Lilburn, C.E. Haith, R.C. Prince, J.T. Beatty, Pleiotropic effects of *pufX* gene deletion on the structure and function of the photosynthetic apparatus of *Rhodobacter capsulatus*, Biochim. Biophys. Acta 1100 (1992) 160-170.

[41] W.P. Barz, F. Francia, G. Venturoli, B.A. Melendri, A. Verméglio, D. Oesterhelt, 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 (1995) 15235-15247.

[42] P. McGlynn, W.H.J. Westerhuis, M.R. Jones, C.N. Hunter, Consequences for the organisation of reaction centre/LH1 complexes of *Rhodobacter sphaeroides* arising from deletion of amino acid residues from the C terminus of the LH1α polypeptide, J. Biol. Chem. 271 (1996) 3285-3292.

[43] J.D. Olsen, E.C. Martin, C.N. Hunter, The PufX quinone channel enables the light-harvesting 1 antenna to bind more carotenoids for light collection and photoprotection, FEBS Lett. 591 (2017) 573-580.

[44] J. Yang, L. Yin, F. H. Lessner, E. S. Nakayasu, S. H. Payne, K. R. Fixen, L. Gallagher, C. S. Harwood, Genes essential for phototrophic growth by a purple alphaproteobacterium, Environ. Microbiol. 19 (2017) 3567–3578.

[45] T.K. Fulcher, J.T Beatty, M.R. Jones, Demonstration of the key role played by the PufX protein in the functional and structural organization of native and hybrid bacterial photosynthetic core complexes, J. Bact. 180 (1998) 642–646.

[46] R.J. Pugh, P. McGlynn, M.R. Jones, C.N.Hunter, The PufX polypeptide in the core complex of *Rhodobacter sphaeroides*: transcription and translation of the gene, order of assembly of the core and topology of PufX in the membrane, Biochim. Biophys. Acta. 1366 (1998) 301-316.

[47] J.D. Olsen, P.G. Adams, C.N. Hunter, Aberrant assembly intermediates of the -LH1-PufX core complex of *Rhodobacter sphaeroides* imaged by atomic force microscopy, J. Biol. Chem 289 (2014) 29927-29936.

[48] D.J. Mothersole, P.J. Jackson, C. Vasilev, J.D. Tucker, A.A. Brindley, M.J. Dickman, C.N. Hunter, PucC and LhaA direct efficient assembly of the light-harvesting complexes in *Rhodobacter sphaeroides*, Mol. Microbiol. 99 (2016) 307-327.

[49] P. Qian, E.C. Martin, I.W. Ng, C.N. Hunter, The C-terminus of PufX plays a key role in dimerisation and assembly of the reaction center light-harvesting 1 complex from *Rhodobacter sphaeroides*, Biochim. Biophys. Acta 1858 (2017) 795-803.

[50] O. Hucke, E. Schiltz, G. Drews, A. Labahn, Sequence analysis reveals new membrane anchor of reaction center-bound cytochromes possibly related to PufX, FEBS Lett. 535 (2003) 166-170.