

Evolution of photosynthetic reaction centers: insights from the structure of the heliobacterial reaction center

Gregory S. Orf^{1,2,3}, Christopher Gisriel^{1,2,3,4}, and Kevin E. Redding^{1,2,5}

¹School of Molecular Sciences, ²Center for Bioenergy and Photosynthesis, Arizona State University, Tempe, AZ 85287, USA

³G.S.O. and C.G. contributed equally to this work.

⁴Current address: The Biodesign Center for Applied Structural Discover, Arizona State University, Tempe, AZ 85287, USA

⁵To whom correspondence should be addressed. E-mail: kevin.redding@asu.edu. Phone: +1-480-965-0182

ORCID IDs: Gregory S. Orf, 0000-0002-4853-8518; Christopher Gisriel, 0000-0002-8011-0052; Kevin E. Redding 0000-0003-2819-4022

Keywords

Photosynthesis; reaction center; evolution of photosynthesis; sequence alignments; structural alignments; Heliobacteria

Abstract

The proliferation of phototrophy within early-branching prokaryotes represented a significant step forward in metabolic evolution. All available evidence supports the hypothesis that the photosynthetic reaction center (RC) – the pigment-protein complex in which electromagnetic energy (*i.e.* photons of visible or near-infrared light) is converted to chemical energy usable by an organism – arose once in Earth's history. This event took place over 3 billion years ago and the basic architecture of the RC has diversified into the distinct versions that now exist. Using our recent 2.2-Å X-ray crystal structure of the homodimeric photosynthetic RC from heliobacteria, we have performed a robust comparison of all known RC types with available structural data. These comparisons have allowed us to generate hypotheses about structural and functional aspects of the common ancestors of extant RCs and to expand upon existing evolutionary schemes. Since the heliobacterial RC is homodimeric and loosely binds (and reduces) quinones, we support the view that it retains more ancestral features than its homologs from other groups. In the evolutionary scenario we propose, the ancestral RC predating the division between Type I and Type II RCs was homodimeric, loosely bound two quinones, and performed an inefficient, light-driven disproportionation reaction to reduce quinone to quinol. The changes leading to the diversification into Type I and Type II RCs were separate responses to the need to optimize this reaction: the Type I lineage added an [4Fe-4S] cluster to facilitate double reduction of a quinone, while the Type II lineage heterodimerized and specialized the two cofactor branches, fixing the quinone in the Q_A site. After the Type I/II split, an ancestor to Photosystem I fixed its quinone sites and then heterodimerized to bind PsaC as a new subunit, as responses to rising O₂ after the appearance of the oxygen-evolving complex in an ancestor of Photosystem II. These pivotal events thus gave rise to the diversity that we observe today.

Acknowledgements

The authors thank Dr. Gillian Gile (Arizona State University) for consultation on phylogenetic tree construction, Mr. Bill Johnson (Arizona State University) for artistic input on figures, and Dr. Robert Blankenship (Washington University in St. Louis) for sharing the draft genome of *Heliorestis convoluta*. All authors were supported by the

4 **Introduction**

5 The first bacteria to harness photosynthesis 3.0-3.5 Gya began the process of
6 transforming the Earth's environment into one capable of sustaining multicellular life through
7 their manipulation of natural metabolic gradients and atmospheric composition, especially after
8 the evolution of the water-splitting reaction (Hohmann-Marriott and Blankenship 2011; Fischer
9 et al. 2016). The protein complex unique to photosynthesis is the reaction center (RC), which
10 catalyzes uphill electron transfer (ET) across a biological membrane using light energy. Light is
11 first absorbed by antenna pigments and the excitation energy is transferred into the core of the
12 RC, which houses ET cofactors. Upon reaching these cofactors, the excitation energy triggers a
13 primary charge separation (CS) event, and an electron is transferred down a potential gradient
14 formed by the chain of ET cofactors to a terminal acceptor (Nelson and Yocum 2006;
15 Blankenship 2014). The CS event is fundamentally important; it is the moment in which the
16 energy contained within a molecular electronic excited state is converted into the energy of a
17 new, biologically-useful redox state. In general, RCs are classified by their terminal electron
18 acceptor within the protein. The terminal acceptor of Type I RCs is a [4Fe-4S] cluster, which has
19 been observed to reduce soluble ferredoxin proteins (Vassiliev et al. 2001). Type II RCs reduce a
20 mobile quinone to quinol, which then diffuses into the surrounding membrane (Diner et al. 1991;
21 Cardona et al. 2012).

22 Seven extant bacterial phyla contain known photosynthetic representatives:
23 Acidobacteria, Chlorobi, Chloroflexi, Cyanobacteria, Firmicutes, Gemmatimonadetes, and
24 Proteobacteria (Cardona 2015; Fischer et al. 2016). Additionally, an ancient symbiosis between a
25 cyanobacterium and eukaryotic ancestor gave rise to the modern eukaryotic algae and land plant
26 lineages (Keeling 2010). Due to billions of years of evolutionary divergence between
27 photosynthetic organisms and their wide distribution across different branches of the tree of life,
28 the sequences of the core RC polypeptides have also widely diverged (Cardona 2015). Their low
29 sequence homology has made piecing together the evolutionary history of photosynthesis
30 difficult, even though the number of available genomes for photosynthetic organisms has greatly
31 grown in recent years. However, despite the low sequence homology, fully-folded RC
32 polypeptides retain a surprising degree of structural similarity (Sadekar et al. 2006), which is
33 advantageous for inferring functional and evolutionary relationships.

34 Several important questions remain concerning how the RC polypeptides radiated
35 amongst the various prokaryotic lineages, as well as the sequence of events that lead to the
36 distinction in function between Type I and II RCs. Because the core polypeptides of RCs from
37 different phyla share low sequence identity (< 30%), improving the amount of structural data
38 available for RCs, both in resolution and diversity of phototrophic groups represented, is vital in
39 addressing these evolutionary questions. When sequence identity falls below ~25%, colloquially
40 termed the "twilight zone," evolutionary relationships become too difficult to reliably map by
41 simple primary sequence alignments (Doolittle 1986; Rost 1999). Assigning the relationships
42 between groups of phototrophs and their proteins is also complicated by the lack of consensus
43 regarding the evolutionary relationships between prokaryotic phyla in general (Battistuzzi and
44 Hedges 2009). However, structural similarity can be used to infer evolutionary relationships well
45 into the twilight zone (Rost 1999; Sadekar et al. 2006; Cardona 2015; Khadka et al. 2017).

46 In this paper, we have focused on the evolutionary relationships between the modern
47 RCs, leaving open the questions regarding the radiation of bacterial lineages. We make use of

1 previous biophysical conclusions in conjunction with new structural comparisons to generate
2 hypotheses about the defining characteristics of two important RC ancestors: the last common
3 ancestor of Type I RCs, and the last common ancestor of all photosynthetic RCs. Through the
4 lens of our mechanistic analysis, we propose an evolutionary scheme to explain the Type I/Type
5 II split and the various heterodimerization events that have occurred since.

6 7 **The new 2.2-Å HbRC structure from *Heliobacterium modesticaldum* is representative of all** 8 **known HbRCs**

9 Recently, the X-ray crystal structure of the Type I RC from *Heliobacterium* (*Hbt.*)
10 *modesticaldum* was solved at 2.2-Å resolution and deposited in the Protein Data Bank (PDB)
11 under code 5V8K (Gisriel et al. 2017). This organism is a member of the family
12 *Heliobacteriaceae* (commonly called the “heliobacteria”) within the phylum Firmicutes, a family
13 which was first described in 1983 (Gest and Favinger 1983). The heliobacterial RC (HbRC)
14 crystal structure revealed a dimer of the PshA polypeptide with two copies of a novel single-
15 transmembrane helix (TMH) antenna subunit PshX bound at the periphery of PshA, in perfect C₂
16 symmetry. This the first representative structure of a homodimeric RC and confirmed the
17 architecture that was first proposed 25 years ago (Liebl et al. 1993). Because homodimeric RCs
18 are believed to be the evolutionary predecessors of heterodimeric RCs (Nitschke and Rutherford
19 1991; Blankenship 1992), this structure is an important milestone in the study of the evolution of
20 these proteins. A general scheme of the HbRC and the membrane in which it resides is presented
21 in **Fig. 1**. Its Gram-positive host (Vermaas 1994) lacks internal membranes and invaginations
22 (Miller et al. 1986), so the HbRC lies within the single cell membrane, inside of the
23 peptidoglycan cell wall. This membrane architecture is unlike chloroplast-containing organisms,
24 where proteins involved in photosynthesis are contained in an extensive network of stacked
25 thylakoid membranes (for a review, see (Pribil et al. 2014)). Upon a light-absorption event, the
26 HbRC transfers an electron through its ET cofactors, referred to as P₈₀₀ (primary electron donor),
27 ec2 (analogous to the term “accessory” used elsewhere), ec3 (primary electron acceptor,
28 analogous to the term “A₀” used elsewhere), and F_X (a [4Fe-4S] cluster), across the membrane to
29 reduce a soluble ferredoxin on the acceptor side (**Fig. 1**) (Fuller et al. 1985; van de Meent et al.
30 1991; Kleinherenbrink et al. 1994). The now-oxidized HbRC is re-reduced on the donor side by
31 a membrane-anchored cytochrome, cyt *c*₅₅₃ (Prince et al. 1985; Albert et al. 1998).

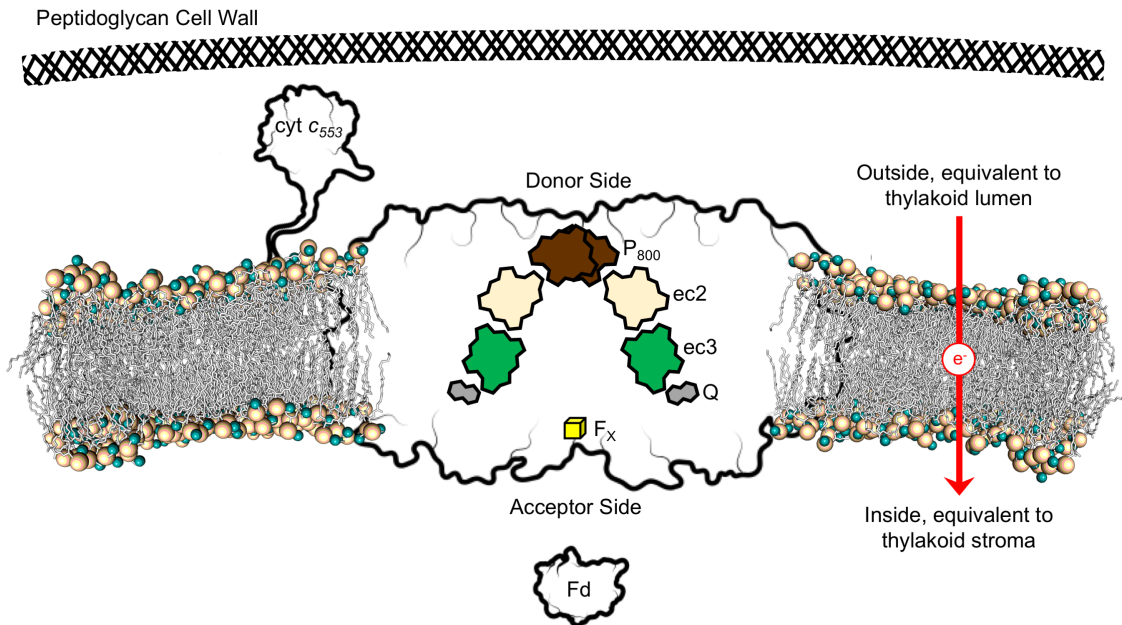


Fig. 1 Membrane architecture of the heliobacteria as it pertains to the HbRC. An outline of the HbRC crystal structure within the membrane, a cyt *c*₅₅₃ homology model, and a PshB1 ferredoxin homology model are shown as they are associated with the heliobacterial membrane. The cofactor names and approximate location within the HbRC are shown in boxes. P₈₀₀ is colored in brown, ec2 is colored in tan, ec3 is colored in green, quinones are colored in grey, and F_x is colored in yellow. The general direction of light-driven ET across the membrane is shown by the red arrow on the right.

The complete amino acid sequence of the HbRC core polypeptide, PshA, is known for four heliobacterial species: *Hbt. modesticaldum* (Sattley et al. 2008), *Hbt. gestii* (Miyamoto et al. 2006), *Heliobacillus (Hba.) mobilis* (Xiong et al. 1998), and *Heliorestis (Hrs.) convoluta* (personal communication from Robert Blankenship, Washington University in St. Louis). To understand the evolutionary implications of the recently solved HbRC structure, it is important to know if the PshA sequence from *Hbt. modesticaldum* is representative of the other known PshA sequences. The four species named above occupy two of the three major clades within the family *Heliobacteriaceae*. The third clade is composed of the genus *Heliophilum (Hph.)*, however no full PshA sequences from this genus are known; only a partial N-terminal sequence of the PshA from *Hph. fasciatum* has been published thus far, so we do not include it in the comparison of the other PshA sequences (Mix et al. 2004; Oh-oka 2007).

The PshA protein from *Hbt. modesticaldum* shares more than 80% sequence similarity with each of the other three known PshA sequences, and 88.7% sequence identity on average. We have performed a sequence alignment of all known PshA sequences using Clustal Omega (**Supplementary Information, Fig. S1**). We have noted the residues comprising the *Hbt. modesticaldum* RC TMHs, the residues providing axial ligand or Hydrogen-bonding (H-bonding) interactions with the ET chain cofactors, and the two cysteines (per PshA) that coordinate the [4Fe-4S] cluster. The sequence alignment shows that all these features are well-conserved in the sequences from other PshA polypeptides. These data support the hypothesis that the HbRC structure from *Hbt. modesticaldum* acts as a good representative for all known heliobacterial species; therefore, we should be able to make informed hypotheses about the evolution of the HbRC with only the single published structure.

Comparison of all known RCs

All known RCs contain the common structural motif of a dimer of five TMHs, referred to here as the “ET domain”, with the ET cofactors sandwiched at the dimer’s interface. The core of all known Type II RCs exists as heterodimers of 5-TMH proteins. In the anoxygenic Type II RCs from Chloroflexi, Proteobacteria, and Gemmatimonadetes, these are called the L and M polypeptides (also known as PufL and PufM, respectively) (Bryant and Frigaard 2006; Zeng et al. 2014), while in the oxygen-evolving Photosystem II (PSII) from cyanobacteria and eukaryotes, these are called D1 and D2 (also known as PsbA and PsbD, respectively) (Barber 2012). The core of the Type I RC from cyanobacteria and eukaryotes, called Photosystem I (PSI), is a heterodimer of the PsaA and PsaB polypeptides (Schubert et al. 1998; Jordan et al. 2001), while the core of the Type I RCs from the anoxygenic Chlorobi (green sulfur bacterial RC; GsbRC), Acidobacteria (chloracidobacterial RC; CabRC), and Firmicutes (heliobacterial RC, HbRC) are homodimers of the PscA or PshA polypeptides (Rémigy et al. 1999; Bryant and Frigaard 2006; Heinnickel and Golbeck 2007; Garcia Costas et al. 2012). In Type I RCs, an additional 6-TMH domain that binds antenna pigments, referred to here as the “antenna domain”, is fused to the N-terminus of the 5-TMH ET domain. A duplication and diversification event of a 6-TMH antenna-binding motif gave rise to the modern CP43 and CP47 (also known as PsbC and PsbB, respectively) proteins that bind antenna pigments and associate with D1 and D2 of PSII, functioning analogously to the antenna domain of the Type I core RC proteins (Schubert et al. 1998; Mix et al. 2005; Cardona 2016a).

All RCs are related to a distant common ancestor that appeared early in the history of life on Earth, well before the radiation of the major bacterial phyla observed today (Olson 1981; Nitschke and Rutherford 1991; Heathcote et al. 2002; Shi et al. 2005; Sadekar et al. 2006; Cardona 2015). As previously mentioned, the low sequence homology between core polypeptides makes reliable sequence alignments difficult to obtain (Rutherford et al. 1996; Cardona 2015), rendering traditional phylogenetic analyses problematic. Despite this, various structural features of RCs are well-conserved (Olson 1981; Olson and Blankenship 2004; Sadekar et al. 2006; Fischer et al. 2016). The recent HbRC structure has added a new class for comparison. In this section, we compare the HbRC to other RCs, starting first with structural and functional features, before moving on to protein sequence comparison. We compare the topologies of phylogenetic trees generated from protein sequence information with that of a phenetic dendrogram constructed using structural data only, including for the first time that from the HbRC. The goal of this comparison is to show that a structure-only analysis is complementary to sequence-only analysis. Then, looking through a functional-mechanistic lens, we identify what we believe to be the key ET cofactor that helps us to understand the observed functional differences between the RC types. This sets the stage for our discussion of the trajectory of RC evolution.

Cofactor identities, orientations, and functions

The ET domain of all RCs coordinate the ET cofactors, which can be chlorophylls (Chls), bacteriochlorophylls (BChls), pheophytins (Pheos), bacteriopheophytins (BPheos), quinones (Q), and [4Fe-4S] clusters. Due to the dimeric nature of the RC core proteins, the ET cofactors line up in two “branches” (Figs. 1 and 2). In all RCs, a special pair of strongly-interacting and closely-associated Chls or BChls ((B)Chls), called “P” (usually denoted with a subscript number, which is the wavelength associated with its photobleaching maximum after CS), is the ET cofactor

1 situated nearest to the donor side (Blankenship 2014). Moving toward the interior of the protein,
2 the next ET cofactor is ec2, followed by ec3 and then Q. In Type II RCs, the quinone is the final
3 electron acceptor within the RC, but in Type I RCs, a [4Fe-4S] cluster called F_X lies on the other
4 side of Q, on the soluble face of the acceptor side (Nelson and Ben-Shem 2004; Nelson and
5 Yocum 2006; Ohashi et al. 2010).

6 After energy transfer to the ET cofactors, an initial CS event occurs; a (B)Chl ET
7 cofactor in an electronic-excited state becomes a good reducing agent, enabling it to reduce its
8 neighbor (B)Chl ET cofactor. The initial radical pair is thought to be P^+ec2^- or $ec2^+ec3^-$.
9 Regardless, within a few picoseconds a second ET event occurs (either from $ec2^-$ to ec3 or from
10 P to $ec2^+$), resulting in P^+ec3^- . From the P^+ec3^- state, more ET reactions can occur as the
11 electron moves down a potential gradient; $ec3^-$ reduces Q or F_X on the acceptor side (which in
12 turn reduces a cellular electron acceptor), and P^+ is re-reduced by a cellular electron donor on the
13 donor side of the membrane (**Fig. 1**), thus opening the RC to another charge separation event.

14 The reduction potentials and activity of the various ET cofactors are tuned by their
15 chemical structures and protein environments, and each RC class has optimized the ET reactions
16 for the specific electron donors and acceptors in their host organisms. The chemical identity of
17 each ET position (**Fig. 2**) follows a distinct pattern:

- 18
- 19 1. P is always a dimer of (B)Chls and ec2 is always a (B)Chl of the same type as P.
- 20 2. The ec3 cofactor is always a Chl *a* (or close derivative) in Type I RCs and is a
21 demetallated version (Pheo or BPheo) of the P/ec2 pigment in Type II RCs.
- 22

23 The orientation and placement of the ET cofactors (**Fig. 2**) also follows a distinct pattern:

- 24
- 25 3. In Type I RCs, ec2 is more perpendicular to the plane of the membrane; whereas in Type
26 II RCs, it is more parallel to the plane of the membrane.
- 27 4. The closest P-to-ec3 center-to-center distance is longer in Type I RCs than it is in Type II
28 RCs.
- 29

30 Finally, the directionality-bias of ET also follows a distinct pattern:

- 31
- 32 5. In Type I RCs, ET can proceed down either branch of the ET chain toward F_X . In
33 homodimeric Type I RCs (*i.e.* the HbRC, GsbRC, and CabRC), the branches are
34 completely symmetric, so ET would not be expected to be biased toward one path. In the
35 heterodimeric PSI, the ET branches are not completely symmetric; however, both
36 branches support ET.
- 37 6. In Type II RCs (all of which are heterodimeric), ET is biased down the ec2 and ec3
38 cofactors of one branch (the A-branch, bound by the L or D1 polypeptide). The quinone
39 on the A-branch (Q_A , bound by the M or D2 polypeptide) is fixed within the RC and
40 performs ET to the mobile quinone on the B-branch (Q_B , bound by the L or D1
41 polypeptide). After Q_B has been reduced to quinol, it diffuses away and is replaced by a
42 fresh quinone. The B-branch ET cofactors ($ec2_B$ and $ec3_B$, bound by the M or D2
43 polypeptide) are not thought to participate in forward ET to the quinones.
- 44

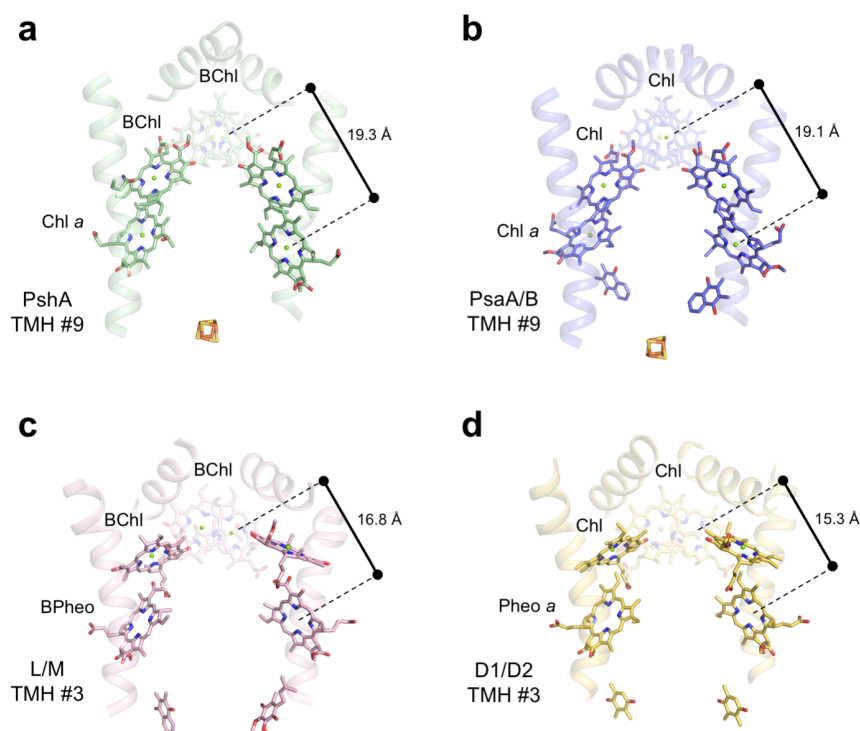


Fig. 2 ET cofactor configuration amongst four representative RC structures. In the HbRC (**a**) and PSI (**b**), P is a dimer of (B)Chls, ec2 is a (B)Chl of the same type as P, and ec3 is a version of Chl *a*. In the PbRC (**c**) and PSII (**d**), P is a dimer of (B)Chls, ec2 is a (B)Chl of the same type as P, and ec3 is the demetallated version (Pheo) of the P/ec2 pigment. Quinones and [4Fe-4S] clusters are shown to illustrate their placement relative to the chlorins. Immediately noticeable is the ec2 orientation difference between Type I and II RCs; in Type I RCs ec2 is perpendicular to the plane of the membrane, whereas in Type II RCs it is parallel to the plane of the membrane. This change pushes ec3 in Type I RCs further away from P (shown as closest center-to-center measurements in the figure) and closer to the [4Fe-4S] cluster. PDB structures 5V8K, 1JB0, 1PRC, and 3WU2 were used to construct this comparison. (B)Chl and quinone tails are omitted for clarity. TMHs and P are made slightly transparent to facilitate viewing the orientations and location of ec2 and ec3.

We have no convincing arguments with regard to the original evolutionary primacy of Chl/Pheo *versus* BChl/BPheo; however, this continues to be debated by other groups (Lockhart et al. 1996; Blankenship et al. 2007; Blankenship 2014; Cardona 2016b; Martin et al. 2017). The fact that Chl derivatives precede BChls in extant biosynthetic pathways would imply that Chl/Pheo were used first (Granick 1957, 1965). Conversely, devolution of the BChl synthesis pathway to end at Chl could have occurred to take advantage of Chl's enhanced oxidation power. We have a slight preference for the original primacy of Chl/Pheo; however, resolving this issue is not essential to the arguments made in this paper, so we will be purposefully ambiguous about it.

The six patterns listed above define the functional differences between the Type I and Type II RC lineages. Some of the factors which determine how the different RCs bind and control their ET cofactors are coded into the primary sequence of each polypeptide and the cofactor biosynthesis pathways present in the host organism. Therefore, the evolutionary trajectory of RC function must be viewed, at least to a certain extent, through the lens of the evolution of the polypeptides. Numerous phylogenetic analyses have been performed on RC

peptides before (Mix et al. 2004; Sadekar et al. 2006; Cardona 2015, 2016a; Fischer et al. 2016; Khadka et al. 2017); however, we feel that it is useful to perform our own analysis to raise some important points about how the results should be interpreted.

ET domain phylogenetic trees and phenetic dendrogram

We first collected the polypeptide sequences for the ET domains of RCs that have structures deposited in the PDB (a list can be found in the **Supplementary Information, Table S1**). We constructed phylogenetic trees, both entirely sequence-based (we use the term “non-structure-based”) and sequence-based but structurally-biased (we use the term “structure-based”; see the **Methods** section). In all 9 non-structure-based phylogenetic trees made from Clustal Omega (Sievers et al. 2011) multiple sequence alignments (MSAs), the PbRC core subunits (L and M) are a monophyletic group, the PSII core subunits (D1 and D2) are a monophyletic group, and the ET domains of PSI (last 5 TMH of PsaA and PsaB) are a monophyletic group (**Fig. 3a**). The HbRC shares a common ancestor with the LCA of the modern PSI polypeptides and the LCA of all modern Type II RC polypeptides, placing its branching point between the PSI and Type II ancestor. This overall topology is maintained by the 9 non-structure-based ET domain trees that were made by varying models and trimming strategies, with bootstrap values for the divergence events always above 70% before the divergence of a single modern polypeptide (*i.e.* L, D1, etc.).

Placement of the root within any of the monophyletic groups would overcomplicate the tree and require drastically different rates of evolution between disparate groups. Because this is unlikely, the root should be placed within the interior branches, most likely between the red and teal diamonds, because this is the current consensus in the field (**Fig. 3a**) (Olson and Blankenship 2004; Sadekar et al. 2006; Cardona 2015; Khadka et al. 2017). If this is the case, then a single early split defined the ET domain function; one descendent would have been the ancestor to all Type I RCs, and the other descendent the ancestor to all Type II RCs. The overall tree alignment is in good agreement with previously-hypothesized evolutionary relationships.

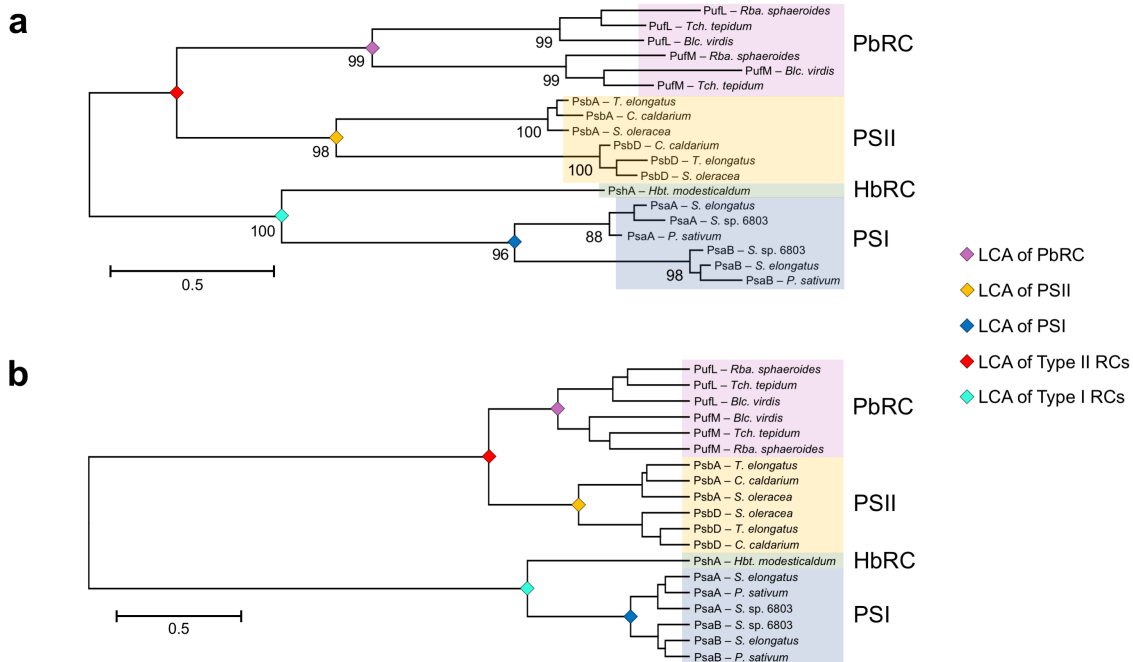


Fig. 3 Phylogenetic tree and phenetic dendrogram of 19 ET domain polypeptides constructed in MEGA7 (Kumar et al. 2016). **(a)** A phylogenetic tree of the 19 ET domain amino acid sequences aligned using Clustal Omega. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (scale bar at bottom left). All positions with less than 50% site coverage were eliminated, leaving 321 positions in the final dataset. The percentage of trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. **(b)** A phenetic dendrogram constructed by creating a matrix of superposition RMSDs of the core polypeptide structures from each of 19 different ET domains for which structures are available. The dendrogram was inferred from the matrix using the UPGMA method (Sneath and Sokal 1973). The optimal dendrogram with the sum of branch lengths = 9.821 is shown and the root is conditionally placed at the deepest node in order to calibrate the branch lengths. The dendrogram is drawn to scale, with branch lengths in the same units as those of the RMSD distances used to infer it (scale bar at bottom left).

To test whether structure-based sequence alignments produce different phylogenetic results, we produced a second set of phylogenetic trees from the 19 ET domains using structure-based MSAs from the PROMALS3D (Pei et al. 2008) and PDBeFOLD (Krissinel and Henrick 2005) servers. These servers first superimpose all the input structures, then produce a sequence alignment by minimizing pairwise distances between residues from the different peptides; since RCs are heavily composed of TMHs, this method forces all TMHs into sequence alignment. This alignment method should be useful for protein families (like RCs) in which the sequences have strongly diverged, yet secondary and tertiary structure have not. The structure-based and non-structure-based MSAs from Clustal Omega and PROMALS3D, respectively, are shown in the **Supplementary Information, Figs. S2 and S3** with highlighted TMH residues (confirmed from the structures) for comparison. In all 18 phylogenetic trees made from structure-based alignments, a low-confidence topology tree is observed. Similar to the non-structure-based phylogenetic trees, the ET domains cluster near one another for PSII, PbRC, and PSI, with the HbRC sharing a more recent common ancestor with PSI than Type II RCs. However, the splitting pattern of L/M and D1/D2 is unresolved, exhibiting low bootstrap values (**Supplementary Information, Fig. S4**). Although this low confidence does not refute the evolutionary scheme inferred by the non-structure-based phylogenetic trees, it probably does signify that either (1) the confidence of non-structure-based ET domain phylogenetic trees are artificially inflated because of poor alignment of sequences, especially within TMHs (for an example, compare the MSAs in **Supplementary Information, Figs. S2 and S3**), (2) seemingly homologous sequences now perform different structural roles, or (3) seemingly homologous structural features are cases of convergent evolution. It is likely that all three issues are relevant to RCs. Furthermore, a limitation of our method is the relatively low number of sequences used in the phylogenetic analysis, intentionally restricted to only sequences corresponding to solved three-dimensional structures.

To compare morphological relationships between the 19 ET domain structures without any reliance on sequence, we constructed a matrix of pairwise superposition RMSDs of the crystal structures, which was then used to create a phenetic dendrogram (**Fig. 3b**). The topology of the dendrogram is essentially identical to that of the non-structure-based phylogenetic trees. In this method, a root is automatically placed at the deepest node in order to calibrate the branching distances to the RMSDs. In **Fig. 3b**, the root is automatically placed between the Type I and Type II lineages, matching the simplest root placement in **Fig. 3a**. Therefore, although the non-structure-based phylogenetic tree confidence may be artificially inflated, the overall topology is still supported by structural comparison. An interesting observation to be gleaned from comparing the two trees in **Fig. 3** is that the structural “rift” between the Type I and II RC groups is actually starker than the sequence comparison would lead one to believe.

1 What our alignment analysis tells us is that sequence-based analysis, even when biased
2 by structural comparison, is simply not enough to provide a detailed picture of the evolutionary
3 path that RCs took from their last common ancestor to the diversity that we observe today. These
4 methods can only tell us *how* the RCs are different. But, to fill in the gaps, we must rely on
5 biochemical and biophysical data to reveal functional aspects of the RCs that cannot be easily
6 predicted from sequence or structure. Quite simply, these data can tell us *why* the RCs are
7 different. Through the recent experimental work on the HbRC from our laboratory and others,
8 outlined in the next section, we have come to the conclusion that the quinone cofactors are the
9 key to understanding why the Type I and Type II RCs differ.

10 *The quinones are the key to understanding the difference between Type I and Type II RCs*

11 Quinones are ubiquitous across the Bacteria and Eukarya domains, and are used by all
12 photosynthetic RCs, as well as by many other proteins in various bioenergetic pathways. In the
13 Type II RCs, the two quinones have distinct roles: one (Q_A) is permanently bound as an
14 intermediate in ET to the other (Q_B), which is the mobile terminal acceptor (Okamura and Feher
15 1992; Okamura et al. 2000). Two CS events are required to sequentially provide the two
16 electrons to fully reduce Q_B ; two protons, which originate from the acceptor-side, must also be
17 provided to Q_B during the process. The quinol then exits the Q_B site and is replaced by a new
18 quinone from the membrane, reinitializing the acceptor side of the RC and increasing the
19 reduction state of the Q-pool (Okamura and Feher 1992; Okamura et al. 2000). In PSI, the
20 phylloquinones are not mobile electron carriers, but are instead tightly-bound cofactors serving
21 as intermediates in ET between $ec3$ and F_X . Therefore, use of quinones as mobile electron
22 carriers has long been thought to be one of the defining characteristics of Type II RCs
23 (Blankenship 1992).

24 Experiments probing the involvement of a quinone in the HbRC ET chain have produced
25 conflicting results (Brok et al. 1986; Trost et al. 1992; Kleinherenbrink et al. 1993; Lin et al.
26 1995; Brettel et al. 1998; van der Est et al. 1998; Oh-oka 2007; Miyamoto et al. 2008; Kondo et
27 al. 2015; Ferlez et al. 2016). It has been clearly shown that the HbRC does not require quinones
28 for forward ET to the terminal electron acceptor, F_X (Kleinherenbrink et al. 1993). The recent
29 HbRC crystal structure revealed that cofactor spacings are slightly different than they are in PSI,
30 which likely explains this (Gisriel et al. 2017). The primary donor (P_{800}) and terminal electron
31 acceptor (F_X) are ~ 2.5 Å closer together (center-to-center) than the analogous cofactors in PSI.
32 Additionally, $ec2$ is moved ~ 2.2 Å closer to F_X and $ec3$ is moved ~ 2.4 Å closer to F_X . Thus, F_X is
33 closer to the set of 6 cofactors that perform primary ET in the HbRC.

34 Heliobacteria use menaquinone (MQ) as their sole quinone; approximately 4-5 MQ per
35 RC have been observed in membranes of *Hbt. modesticaldum* and *Hba. mobilis*. MQ has been
36 found associated with purified HbRC in variable amounts, depending on preparation conditions
37 (Trost and Blankenship 1989; Sarrou et al. 2012). The purification and crystallization conditions
38 used to produce the recent HbRC structure removed MQ from the HbRC, so MQ was not
39 resolved in the structure (Gisriel et al. 2017). It was noted, however, that an ill-defined area of
40 non-peptide electron density lies near $ec3$. Although a flat quinone headgroup does not fit into
41 the electron density, an isoprenoid tail approximately the length of that from MQ-9 does. One
42 hypothesis to explain this is that during the purification process, the loosely-bound MQ
43 exchanged with an isoprenoid such as geranylgeranyl phosphate (Gisriel et al. 2017).
44 Interestingly, the location of the electron-dense headgroup of this unassigned density does not lie
45 between $ec3$ and F_X . Instead, it lies ~ 5 Å to the periphery of the conjugated macrocycle of $ec3$. If
46

1 this is the site where MQ normally binds, then it is more likely to be a site for double reduction
2 of a mobile quinone than one for binding a cofactor serving as an ET intermediate between ec3
3 and F_X.

4 Light-driven quinone reduction in *Hbt. modesticaldum* membranes has recently been
5 demonstrated (Kashey et al. 2018), strongly suggesting that the HbRC is responsible for this
6 activity. The data were consistent with a redox cycle between the HbRC and cyt *bc* complex,
7 mediated by MQ (transporting electrons to the cyt *bc*) and the membrane-attached cyt *c₅₅₃*
8 (transporting electrons back to the HbRC). Moreover, the inhibitor tertbutryn, which binds the
9 Q_B mobile quinone site of Type II RCs, inhibited the MQ photoreduction activity. If this
10 peripheral site identified in the HbRC structure is indeed a mobile quinone reduction site, then
11 the HbRC exhibits the ability to use both soluble ferredoxins or lipophilic quinones as terminal
12 electron acceptors. A potential mechanism is the following: (1) CS in the presence of F_X⁻ results
13 in the P₈₀₀⁺ec3⁻F_X⁻ state, (2) ET from ec3⁻ to the nearby MQ results in production of the
14 semiquinone (P₈₀₀⁺MQ⁻F_X⁻), which (3) oxidizes F_X⁻ to yield the fully reduced quinol
15 (P₈₀₀⁺MQH₂). Protonation of the reduced quinone species must also occur twice, but we cannot
16 presently identify a clear protonation channel involving nearby residues or water molecules. The
17 HbRC could thus be the first example a Type I RC performing both Type I and Type II
18 functions, a feature reminiscent of past hypotheses regarding a “Type 1.5 RC” as the ancestor of
19 all RCs (Nitschke et al. 1996; Allen 2005; Sadekar et al. 2006). Since PSI from the *menD1*
20 mutant of *Chlamydomonas* has been shown to be capable of double-reducing plastoquinone
21 inserted into the phyloquinone site (McConnell et al. 2011), and the HbRC appears to also be
22 capable of double-reducing quinones, we hypothesize that the LCA of Type I RCs did so as well.
23 If the LCA of Type I RCs did indeed have the ability to terminally reduce quinone to quinol like
24 modern Type II RCs, then the role of the quinones in the LCA of all RCs is a key to
25 understanding the Type I/Type II split.

26 27 **A New Proposal for the Pathway of RC Evolution**

28 Using informed hypotheses and the idea that the role of quinones is key, we propose a
29 new pathway for the evolution of RCs (**Fig. 4**). It is, of course, impossible to know if our
30 proposed pathway is a faithful recounting of what occurred over 3 Gya. However, we have
31 followed a few guiding principles to be consistent in our hypotheses:

- 32 1. No step should result in a loss of fitness (*i.e.* a potential future gain tomorrow cannot be
33 used as justification for losing something that works well today).
- 34 2. Given the choice between two possible ancestors, the simpler one should be chosen in the
35 absence of a compelling rationale (*i.e.* simpler versions usually precede more complex
36 versions).
- 37 3. Any extinct RCs that predated the LCA of all extant RCs will not be considered, and their
38 function(s) will be ignored (*i.e.* we will not debate the appearance of the first RC, which
39 almost certainly predates the appearance of the LCA of all extant RCs).
- 40 4. We will not conflate host organism evolution with the evolution of the RC polypeptides
41 (*i.e.* we focus only on RC *function* here, so as to avoid too much conjecture on ancient
42 LGT and taxon radiation events).

43 In describing this pathway, we will start at the LCA of all extant RCs, which we term the
44 “ancestral reaction center” (ARC). We will propose the functional role and mechanism of the
45 ARC, which leads to the reasons for why the ARC diversified into the Type I and Type II

- 1
- 2
- 3
- 4

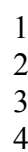


Fig 4. Proposed evolutionary scheme for photosynthetic RCs.

The ARC was a homodimeric RC that reduced quinones

Quinones are found in all known photosynthetic RCs, strongly suggesting that the ARC also used quinones. Reduction of quinones on one side of the membrane and oxidation of quinols on the other side is a major proton pumping mechanism driving ATP synthesis. The discovery that quinones can be fully reduced and perform mobile exchange by the homodimeric HbRC suggests that mobile quinones were an ancestral feature in the Type I lineage, and thus in the ARC as well. If low-potential electron sources were relatively abundant on the early anoxic earth (e.g. Fe^{2+} , S^- , H_2), as is currently thought (Fischer et al. 2016), then ATP production by cyclic electron flow (CEF) was likely to be the primary role of the earliest RCs. Therefore, the idea that the ARC evolved early in the history of life as a homodimeric membrane protein whose sole purpose was to perform the light-driven reduction of quinones to quinols, without the use of an iron-sulfur cluster, is a reasonable hypothesis. Cytochromes and quinones, as well as a protein serving the role of the cyt *bc* complex, would have predated the RC (Furbacher et al. 1996; Dibrova et al. 2013), and their function was to link the electron transport pathways of chemo(litho)trophic metabolism to proton pumping. Thus, the ARC would have fit into this preexistent chain in the first phototrophs to boost ATP production using light energy.

It has been suggested recently that a Type I RC appeared first and Type II RCs evolved from it (Martin et al. 2017). This conclusion was based on three premises, each of which builds upon the previous one: (1) that the organism in which phototrophy first evolved was an obligate autotroph, (2) that the carbon fixation pathway in this organism used low-potential ferredoxins that were reduced with molecular hydrogen (H_2) as electron source, and (3) that use of a Type I RC would have been the best answer to generate reduced ferredoxin and power carbon fixation when using higher potential electron sources (e.g. sulfide). Although this scenario is interesting, the argument that only a Type I RC would have initially been useful is by no means iron-clad. Firstly, if reduced carbon were so scarce that autotrophy was required in the earliest organisms, it is difficult to imagine how the earliest cells could ever have arisen. Thus, it is not inconceivable that the earliest phototrophs made use of abiotically-reduced carbon (Joe et al. 1986; Etiope and Sherwood Lollar 2013). Secondly, while the Wood-Ljungdahl (reductive acetyl-CoA) pathway makes sense as the original autotrophic mechanism, because it is the only carbon fixation pathway that conserves energy, it does not generate nearly enough ATP to drive conversion of all the fixed carbon into other necessary biomolecules (Ragsdale and Pierce 2008). Thus, the appearance of a simpler Type II RC in such an organism to generate additional ATP (via CEF) for the synthesis of those necessary biomolecules might still have been of enormous benefit. In other words, the ARC did not necessarily need to generate a low-potential reductant like ferredoxin in order to provide a selective advantage.

If we take the opinion that mobile quinone reduction in Type II RCs is a derived trait, we must ask the question: *what would have been the driving force for losing F_X in the Type II lineage?* Since it appeared, F_X has functioned very effectively in all Type I RCs to reduce the ferredoxin pool. Moreover, a CEF system in which ferredoxins are reduced by the RC would result in more proton pumping than one in which quinones are reduced by the RC (**Supplementary Information, Fig. S5**). Ferredoxins can be used to reduce NAD(P)^+ , and a NAD(P)H dehydrogenase (i.e. Complex I) can be used to pass the electrons to the quinone pool, resulting in 1-2 additional protons pumped per electron transferred, depending upon the coupling of proton pumping to electron transport in the ancient NAD(P)H dehydrogenase. Thus, if the ARC had an F_X cluster and were part of such a CEF system, to trade that for a system in which less ATP was made would represent a significant decrease in fitness, and would therefore have

been strongly selected against. The simpler hypothesis, therefore, is that the ARC did not contain F_X and the Type I lineage gained it.

If the ARC was indeed homodimeric, lacked an F_X cluster, and reduced mobile quinones, we can infer that it faced an inherent limitation in its chemistry, due to its two symmetric ET branches. Following CS and subsequent ET to one of the quinones and re-reduction of P, the unstable semiquinone radical could have been processed in three different ways (**Fig. 5**):

- 1) **Exchange of the semiquinone ($QH^\bullet/Q^{\bullet-}$) with a new quinone:** Semiquinone disproportionation in the membrane (either spontaneous or catalyzed by an unknown enzyme) could lead to production of a quinol ($2 QH^\bullet \rightarrow Q + QH_2$). Alternatively, the semiquinone could have been re-oxidized in the Q_o (Q_P) site of the cyt *bc* complex or analog thereof.
- 2) **A second CS event in which ET proceeds down the same branch:** After proton-coupled ET to the semiquinone, along with an additional protonation, the mobile quinone will have been converted to a quinol, which can exchange with a new quinone.
- 3) **A second excitation event where ET proceeds down the opposite branch:** This would lead to production of a semiquinone radical on the other side. Subsequently, a *slow proton-coupled radical disproportionation reaction* could occur, accompanied by final protonation of the acceptor. The newly formed quinol could then be replaced by a new quinone.

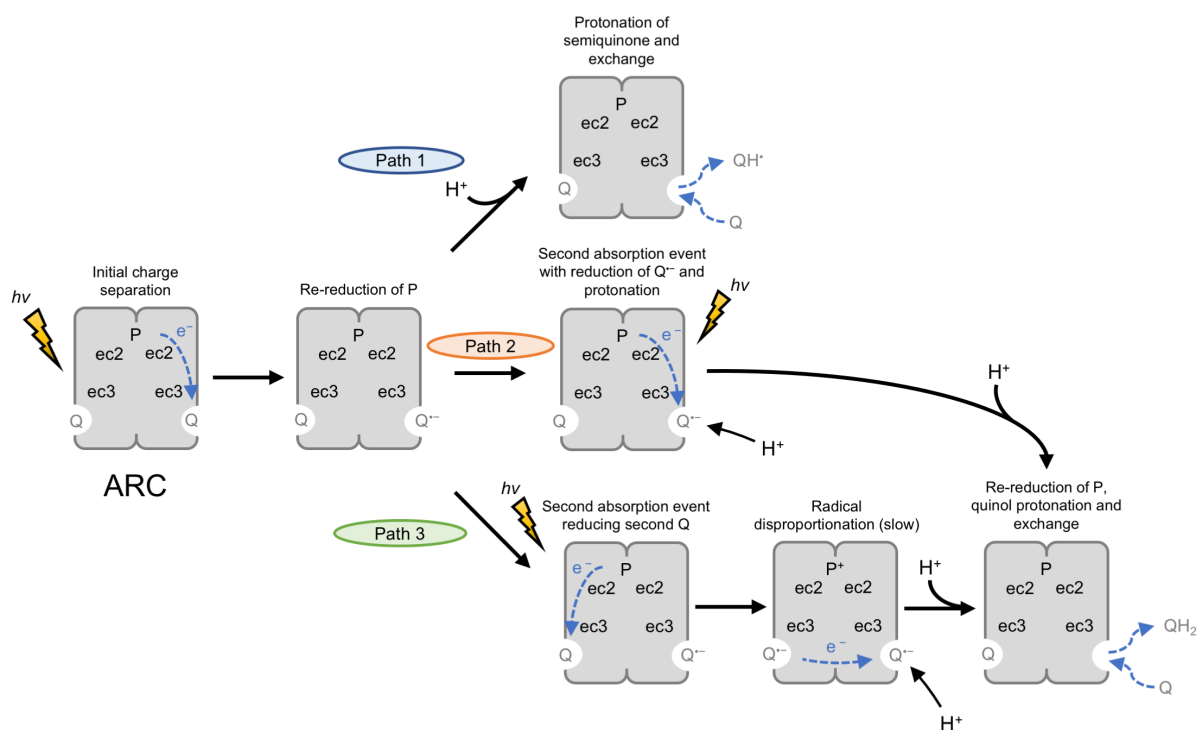


Fig. 5 Proposed mechanism for quinone reduction in the ARC with two mobile quinone sites and no F_X cluster. After the initial light-driven reduction of quinone to semiquinone, three paths exist to fully reduce the semiquinone to quinol. At all points before the re-reduction of P^+ by an external reductant, CR is a likely occurrence.

A major problem for all these pathways is the very low pK_a of the semiquinone ($QH^\bullet \rightleftharpoons Q^{\bullet-} + H^+$), which makes Q unfavorable to protonate. This fact, coupled with the very low reduction potential of the semiquinone anion ($Q^{\bullet-} + e^- \rightarrow Q^{2-}$), led to the suggestion that the second quinone reduction step in an RC must be a proton-coupled ET ($Q^{\bullet-} + e^- + H^+ \rightarrow QH^-$) followed by final protonation ($QH^- + H^+ \rightarrow QH_2$), thus avoiding the unstable intermediates $QH^\bullet$ and Q^{2-} (Graige et al. 1996; Okamura et al. 2000). Thus, Path 1 would not be expected to be very effective, as neither protonation of the semiquinone nor exit of the semiquinone anion from the protein interior to the less polar lipid environment would be favorable. Moreover, Path 2 is unlikely to occur, since the negative charge on the semiquinone anion close to ec3 would disfavor reduction of the nearby ec3 due to electrostatic repulsion. Therefore, excitation of a RC with a semiquinone anion on one branch would likely result in CS on the other branch, as was demonstrated for a mutant PSI that slowed ET from PhQ_A (Santabarbara et al. 2015). Therefore, double reduction of a quinone by Path 2 would not occur until both quinones were in the semiquinone state. Thus, it seems likely that Path 3, in which disproportionation of the two semiquinones (with concomitant protonation of the acceptor quinone) to form QH^- followed by protonation to form QH_2 , would be the major pathway of quinone reduction in the ARC. If this disproportionation reaction was slow, then CS on both branches would be inhibited, leading to CR. Thus, while the ARC would be able to convert light energy into chemical energy by oxidizing a high-potential electron donor, like a cytochrome *c*, and reduce quinone to quinol with uptake of protons from the cytosolic (N) side, it would do so rather inefficiently. **We hypothesize that the Type I/II split represented two different solutions to the same problem: improving the efficiency of this reaction.**

The appearance of F_X in the Type I lineage

The branching point in RC evolution from the ARC to the Type I RC lineage was the acquisition of the F_X cluster. We term this new ancestral form “Proto-RC1” (Fig. 4). In our view, this was the key innovation driving the other changes in the Type I lineage. The [4Fe-4S] binding loop would have been positioned near the cytoplasmic side (as it is today in extant Type I RCs), allowing pre-existing cytosolic [4Fe-4S] cluster insertion machinery to interact with the site after membrane insertion of the core polypeptides and dimerization. The homodimeric nature of Proto-RC1 also means that only two (rather than four) mutations resulting in cysteine would be required in the cytoplasmic loop between the second and third TMH of the ET domain. Thus, this evolutionary step would not have been very difficult.

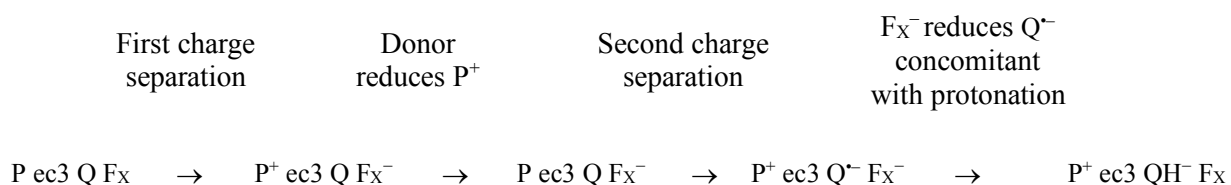
We believe that the main function of F_X in the Proto-RC1 was to facilitate quinone reduction. Nature can tune the reduction potential of Fe-S clusters over a wide range (*i.e.* over 1 V of potential; (Bak and Elliott 2014)) by changing the environment of the cluster. The two reduction potentials associated with quinone reduction (*i.e.* the $Q/Q^{\bullet-}$ and the $Q^{\bullet-}/QH^-$ couples) can also be tuned by their environment, although the first will always be more negative than the second, since the semiquinone is the least stable species. The reduction potential of ec3 is very low (~ -1 V), thus it should always be favorable for this cofactor to reduce either Q or F_X .

We can envision two possible mechanisms of action for the new F_X cluster depending on how the reduction potential of F_X is poised. In the first, the reduction potential of F_X is below that of the $Q/Q^{\bullet-}$ couple, leading to the production of a semiquinone on each side of the RC after two CS events. Here, F_X could serve as a catalyst to accelerate the disproportionation reaction of the two semiquinones. Even if the first ET event ($Q^{\bullet-} F_X Q^{\bullet-} \rightarrow Q F_X^- Q^{\bullet-}$) were uphill, the

second event ($Q F_X^- Q^{\bullet-} + H^+ \rightarrow Q F_X^- QH^-$) would be very favorable. A final protonation would make quinol formation irreversible.

In the second mechanism, the reduction potential of F_X is above that of the $Q/Q^{\bullet-}$ couple, so the first CS event would lead to reduction of F_X . The second CS would result in the $P^{++}Q^{\bullet-}F_X^-$ state, and ET from F_X to the semiquinone would produce quinol (*i.e.* the unstable $Q^{\bullet-}$ oxidizes F_X^-) (**Reaction Scheme 1**). The latter mechanism is similar to what we have hypothesized for reduction of a mobile quinone in the HbRC (Kashey et al. 2018). As mentioned previously, mutations in the region of the F_X cluster could change the potential of the cluster in a relatively simple fashion. Thus, even if the Proto-RC1 used the first mechanism immediately after introduction of F_X , the transition to use of the second mechanism (**Reaction Scheme 1**) would not be difficult.

Reaction Scheme 1: Proposed mechanism of quinone reduction by Proto-RC1. After the quinone is doubly reduced and protonated to QH^- , a final protonation step would yield QH_2 .



Note that **Reaction Scheme 1** invokes the proton-coupled ET step from F_X^- to $Q^{\bullet-}$ which avoids the unstable $QH^{\bullet}$ and Q^{2-} states (Graige et al. 1996; Okamura et al. 2000). We believe that a similar mechanism exists to prevent these unstable states from accumulating in the extant HbRC; further work will be required to identify proton donors and a pathway from the acceptor side to the site of Q reduction. The fine mechanistic details will depend upon protonation channels and the stability constant of the semiquinone ($K_S = [QH^{\bullet}]^2 / \{[Q][QH_2]\}$). The pK_a of semiquinone is typically low enough to ensure that the dominant species is the anion ($Q^{\bullet-}$) at neutral pH, so this is what we assume is the case with the HbRC and Proto-RC1.

With F_X in place on the cytoplasmic side of the RC with a reduction potential above that of the previous cofactors, the electron generated by the first CS event would always end up there. This situation was one that natural selection could exploit, as any soluble cellular electron carriers with a higher reduction potential could be reduced by F_X^- if they were able to bind to the cytoplasmic face of the RC, even transiently. Indeed, the acceptor side of the HbRC seems to be quite promiscuous, capable of reducing any acceptor with a potential above its own (~ -500 mV) (Ferlez et al. 2016), including those not found in heliobacteria, such as cyanobacterial flavodoxin (Romberger and Golbeck 2012). In all the lineages with Type I RCs, ferredoxins were likely recruited from the genome to interface with the RC, perhaps binding weakly at first (allowing for quinol production most of the time), then interacting progressively better to accept the majority of the electrons from the RC. The lack of homology between the F_A/F_B proteins associated with Type I RCs supports this scenario; there is no universal Type I RC F_A/F_B protein that radiated amongst prokaryotes along with the RC. Simple BLAST searches indicate that PshB, PscB, and PsaC are all more related to ferredoxins from non-photosynthetic organisms than they are to each other (**Supplementary Information, Tables S2, S3, and S4**). For example, in heliobacteria, the PshB ferredoxin found to associate with the HbRC is closely related to clostridial and bacteroidial ferredoxins and shows little homology, other than the typical ferredoxin Fe-S cluster

1 binding motif, with the F_A/F_B proteins of the green sulfur bacteria (PscB) or cyanobacteria
2 (PsaC). In fact, the gene encoding PshB is not even located within the photosynthetic gene
3 cluster (PGC) (Heinnickel et al. 2007; Sattley et al. 2008), indicating it was not laterally
4 transferred to the heliobacterial ancestor with *pshA*.

5 With the increasing use of ferredoxins, rather than quinones, as electron acceptors from
6 the Proto-RC1, the host organism would gain an advantage in ATP production from the longer
7 ET cycle (**Supplementary Information, Fig. S5**). Thus, there would have been a selective
8 advantage in optimizing ET to ferredoxins via F_X instead of Q. One of the ways in which the RC
9 was modified to do this was to use a Chl *a* as ec3, which would have produced a larger driving
10 force for ET to F_X . The other modification was to change the orientation of ec2 to stretch the 6-
11 chlorin system across the membrane, placing ec3 closer to F_X for direct reduction. The change in
12 ec2 orientation would have resulted in changes in the arrangement of the TMHs of the ET
13 domain, with the result that only the prime epimer of the pigment in the P site could be
14 accommodated there. We term the RC that contained all of these changes "LCA-RC1" (**Fig. 4**),
15 the last common ancestor of all extant Type I RCs. We will go into more detail about the
16 changes in ec3, ec2, and the special pair in later sections.

17 As noted above, an expanded electron transport cycle that included a Complex I (NADH
18 dehydrogenase) in addition to Complex III (cyt *bc*) would substantially increase the number of
19 protons pumped per electron transferred (**Supplementary Information, Fig. S5**). This would
20 provide the evolutionary driving force to switch F_X from a cofactor enabling quinone reduction
21 to one whose primary function was to reduce ferredoxins. However, in situations when soluble
22 electron acceptors are in short supply (*i.e.* when the ferredoxin pool is largely reduced, possibly
23 in high-light conditions), then the Type I RC could fall back on quinone reduction as a back-up
24 system to make a smaller amount of ATP via a short cycle involving only Complex III. This
25 seems to be what heliobacteria do now (Kashey et al. 2018).

26 27 *Heterodimerization events in the Type II lineage*

28 In what led to the Type II lineage, a different mechanism was used to speed the
29 inefficient quinone disproportionation reaction: heterodimerization, which allowed for
30 specialization of the two ET branches. In one branch (A-side), the quinone was immobilized and
31 cut off from a proton source, to convert it from a terminal acceptor into an ET intermediate (Q_A).
32 In the other branch (B-side), the reduction potentials of the ec2 and ec3 cofactors were lowered
33 to inhibit CR between the semiquinone and oxidized special pair ($P^+Q_B^-$ state) (Kirmaier et al.
34 1985, 1991; Parson et al. 1990; Warshel et al. 1994). The reduction potentials of Q_A and Q_B were
35 also tuned to favor ET from the former to the latter. The better stability of the mobile
36 semiquinone (Q_B) would allow enough time for reduction of P^+ and a second CS to occur,
37 leading to successful double-reduction of Q_B to quinol. Note that this last step is the same
38 semiquinone disproportionation reaction ($Q_A^- Q_B^- \rightarrow Q_A Q_B H_2$) described for the ARC, but has
39 now been optimized by specialization of each branch and associated quinone. Additionally, CR
40 becomes less probable because the ET cofactors in the B-branch are inactive and because CR is
41 now in the seconds timescale for $P^+Q_A^-$; even if CS occurred in a RC with an empty Q_B site,
42 there would be adequate time for a new quinone to arrive via diffusion before CR occurred.

43 Molecular clock studies have recently shown that the D1/D2 ancestor and the L/M
44 ancestor diverged from each other very early after the origin of photosynthesis, with independent
45 D1/D2 and L/M heterodimerization events occurring soon after the split (Cardona 2016a;
46 Cardona et al. 2017). It is therefore logical that the heterodimerization strategy was used twice,

early and convergently, in the lineages resulting in Ancestral PSII (ancestral D1/D2) and Ancestral PbRC (ancestral L/M). A signature event in the split between D1/D2 and L/M seems to be the loss of the Chl_D/Chl_Z site in the PbRC lineage. As explained later in the section “Recruitment of Antennas”, the driving force for this loss may be the association of the Ancestral PbRC with an LH1-like antenna complex. From the points marked “Ancestral PbRC” on our evolutionary scheme (**Fig. 4**), few changes need to occur to result in the modern PbRC/CfxRC/GmRC lineage. From Ancestral PSII, three changes are required to result in modern PSII, all relating to the antenna domain (CP43/CP47) and the OEC (both discussed in later sections). These changes must have occurred before 2.0-2.5 Gya to precede the rise of oxygen levels in the atmosphere (Fischer et al. 2016).

Accommodation of new function in Type I RCs by changing the ET cofactors

At this point, we will come back to the Type I lineage to explain the cofactor identity and orientation differences which distinguish it from the Type II lineage. In our analysis, the conclusion drawn is that the ec3/ec2/P identities, orientations, and functions of the ARC would most resemble those of modern Type II RCs. We make the argument that with the appearance of F_X in the Type I lineage, the ec3, ec2, and P cofactors were each affected (likely in that order) in a logical way to adjust to its presence, which we outline below:

1. The ec3 cofactor: switching to Chl and moving closer to F_X.

The Type II RCs across the Proteobacteria, Chloroflexi, Gemmatimonadetes, and Cyanobacteria/Eukaryotes use different quinone species (e.g. ubiquinone, menaquinone, plastoquinone, etc.) as the final electron acceptor and have different electron donors (e.g. soluble cyt *c*, bound tetraheme cytochrome, water via the OEC), necessitating adjustment of the ET cofactors to accommodate the change in the energetic gap between donor and acceptor. In Type II RCs, the energetic level of the ec3 site scales with that of the special pair because it is occupied by the demetallated version, Pheo or BPheo ((B)Pheo), of the special pair pigment. For example, in the PbRC, if the special pair contains BChl *a*, then the ec3 site will contain BPheo *a*. Since the reduction potential of a (B)Pheo ($\text{Pheo} + e^- \rightarrow \text{Pheo}^-$) is always ~300 mV more positive than that of the corresponding (B)Chl ($\text{Chl} + e^- \rightarrow \text{Chl}^-$) (Fajer et al. 1975), the (B)Chl⁺(B)Pheo⁻ charge-separated state will be the most thermodynamically favorable one of all the potential (B)Chl-based radical pairs, and initial CS will efficiently produce the P⁺ec3⁻ state. The next ET cofactor after (B)Pheo is the Q_A quinone, whose reduction potential will always be higher than any (B)Pheo. Thus, ET from (B)Pheo to Q_A, whatever its chemical identity, will always be favorable. An exception to the above rule is found in PSII from the Chl *d*-producing cyanobacterium *Acaryochloris marina*, which uses Chl *d* in the special pair but retains Pheo *a* in the ec3 site (Chen et al. 2005; Tomo et al. 2007). Mutants of PbRCs have been produced in which the BPheo is replaced by BChl (the resultant BChl site is termed β) (Kirmaier et al. 1995a, b; Pan et al. 2016). When this change occurs, strong energetic mixing results between the P⁺ec2⁻ and P⁺ β ⁻ states, preventing efficient ET to the quinones and increasing the probability of CR. Therefore, using (B)Pheo as ec3 is probably important for maintaining a redox potential higher than ec2 in Type II RCs.

Invariably, every known Type I RC contains a Chl *a* molecule or Chl *a* derivative in the ec3 position, regardless of the pigment content of the rest of the RC. Even the green sulfur bacteria, which use millions of BChl *c*, *d*, or *e* molecules per cell to construct their chlorosomes and thousands of BChl *a* molecules per cell to construct their antennas and RCs, still make and

1 insert two Chl *a* molecules specifically into the ec3 site of each GsbRC. No reports have thus far
2 been published of Type I RC mutants that successfully replace the Chl *a* at ec3 with another
3 pigment. In Type I RCs, the energetic level of ec3 is thus fixed at the level of Chl *a* and the
4 energy of the P^+ec3^- state will vary as the rest of the pigments (P, ec2) in the RC change.

5 In our view, the shift to use of a Chl *a* in the ec3 site of the Proto-RC1 was to use the F_X
6 cluster as the terminal electron acceptor, which has a much more negative reduction potential
7 than any quinone. Thus, ET from ec3 to F_X will be more favorable if a (B)Chl occupies the ec3
8 site rather than a (B)Pheo. This was accomplished by introducing a residue that could serve as an
9 axial ligand in TMH10; however, it is unknown how Chl *a* is *specifically* selected for this
10 position in the RCs in which it is the minor pigment.

11 The reason for incorporating Chl *a*, as opposed to a BChl, in the ec3 site is likely
12 explained by its particularly low reduction potential. If the LCA-RC1 used Chl *a*, then Pheo *a*
13 would have been originally present in the ec3 site of the ARC, and the addition of an axial ligand
14 in TMH 10 would have led to its replacement by a Chl *a* situated closely to F_X . In the lineages
15 leading to the HbRC and GsbRC, the presence of a Chl *a* in the ec3 site may have been selected
16 for strongly enough, in the face of a shift to use of BChl as the major pigment, that it was
17 retained as the ec3 cofactor. This would explain why the heliobacteria, green sulfur bacteria, and
18 chloracidobacteria bother to make a version of Chl *a*, which they put into no other site than ec3.
19 If the ARC instead used a BChl, then the first change would have been a replacement of BPheo
20 with BChl in the ec3 site of LCA-RC1. Afterwards, the selection for a better reductant for F_X
21 would have led to the use of a Chl *a* derivative as the ec3 cofactor. This might explain why the
22 Chl *a* derivatives used in the HbRC and GsbRC are chemically different and may have distinct
23 biosynthetic pathways (Bryant et al. 2012; Sattley and Swingley 2013). At this point we cannot
24 distinguish between these two scenarios. It is also important to note, however, that it is not
25 obvious which structural features of the ec3 site in the HbRC select for the Chl *a* derivative.

26 The ec3 cofactor in Type I RCs accepts a H-bond from a residue in a surface α -helix on
27 the cytoplasmic side. This is a Ser residue in the HbRC and a Tyr in PSI. The position is
28 conserved, and occurs immediately after TMH10. The H-bond is to the 13¹-keto group, which is
29 part of the conjugated π system. Donation of a H-bond to this functional group should raise ec3's
30 reduction potential ($Chl + e^- \rightarrow Chl^{\cdot-}$). Thus, the H-bond to ec3 would make it a poorer reductant
31 of the quinone, but a better oxidant of $ec2^{\cdot-}$. One might think of this as a way of compensating
32 for the very large shift in ec3's reduction potential caused by its change in chemical identity.
33 Once ec3 changed from (B)Pheo to (B)Chl, the redox gradient from ec2 to ec3 would have been
34 lessened, so donating a H-bond to the 13¹-keto group of ec3 would restore some of this gradient.
35 Note that no ec2 cofactor in any RC is H-bonded in this way. It is also possible that the
36 placement of the H-bond donor on the surface helix served to move ec3 closer to the membrane
37 surface, and thereby closer to F_X .

38 39 2. Changing the binding orientation of ec2

40 In all known RCs, the ec2 (B)Chl is the same type as the major (B)Chl of the RC. When
41 the positions of the RC (B)Chls in all the RC structures are compared, the main difference
42 between Type I and II RCs is the orientation of the ec2 cofactor, which is in turn determined by
43 the position of its axial ligand (**Fig. 6**). In PSII and the PbRC, the surface "P-helix", within the
44 loop between TMH 3 and 4, of D1/D2 and L/M provides the axial ligand for the ec2 cofactors. In
45 L/M, this is a His residue that directly coordinates the central Mg of ec2. In D1/D2, the axial
46 ligand is a water molecule that usually interacts with a Thr side-chain or Ile backbone carbonyl

(for a sequence alignment of this region, see **Supplementary Information, Table S5**). The residue in the analogous position of Type I RCs equivalent is either a Tyr (in the HbRC) or Trp (in PSI). The Trp residue in PSI does not interact with ec2. In the HbRC structure, Tyr510 provides a H-bond to the 13²-ester carbonyl oxygen of ec2. The axial ligand to ec2 in Type I RCs is instead a water molecule H-bonded to a Gln (HbRC) or Asn (PSI) side chain found in TMH 9 (**Figs. 6a and 6b**). This position in the Type II RCs is an Ala or Pro, lacking the ability to coordinate either a (B)Chl central Mg or water (**Figs. 6c and 6d**).

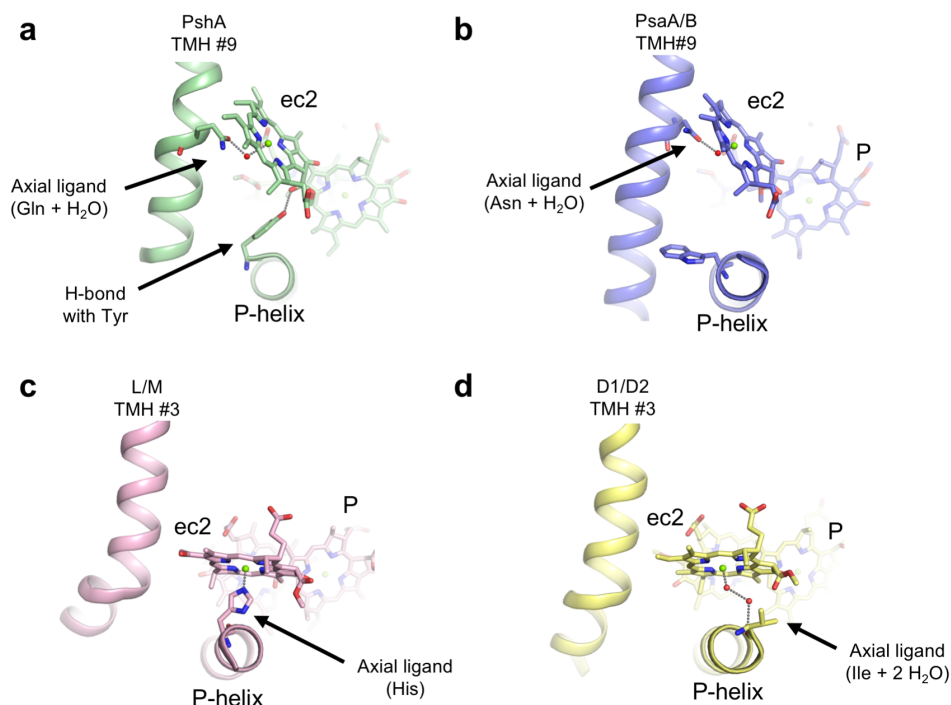


Fig. 6 Differences in ec2 coordination between representative RC structures. (a) In the HbRC, Tyr510 on the P-helix provides a H-bond to the 13¹-keto oxygen of the ec2 BChl g. The homologous residue in PSI (b) is a Trp which does not participate in H-bonding. The homologous residues in Type II RCs provide the axial ligand for ec2 binding: in the PbRC (c) it is a His residue and in PSII (d), it is an Ile residue plus two waters. In the HbRC (a) and PSI (b), the axial ligand for the ec2 central Mg is a water molecule coordinated by a Gln or Asn on TMH 9, but in the PbRC (c) and PSII (d), this position is an Ala or Pro which does not contribute to ec2 coordination. PDB structures 5V8K, 1JB0 (chain A), 1PRC (chain L), and 3WU2 (chain A) were used to construct this comparison. (B)Chl tails are omitted for clarity.

The different locations of the axial ligand to ec2 – from the “side” in Type I RCs or from the “top” in Type II RCs – results in a noticeable shift in the positioning of the TMHs to accommodate this change (**Fig. 7**). This is one of the most easily-visualized differences between the two types of RCs, with the ET domain monomers of the Type II RCs having a more severe “twist angle” around each other than those in the Type I RCs. A possible photophysical explanation for the change arises from the need to optimize ET to the final electron acceptor in each RC. In early Type I RC’s that had recently acquired F_X, there would have been selective pressure to optimize ET to F_X, as mentioned previously. A solution to this would be to stretch the ET cofactors out so that ec3 moves closer to F_X, facilitating the final ET step. By changing the axial ligand location of ec2, thus rotating ec2 more perpendicular to the membrane plane, ec2 would be better able to fit in directly between P and ec3, so that the rate (and thereby the

efficiency) of formation of P^+ec3^- would not suffer. This scenario is supported by the measurement of the closest center-to-center distances between P and ec3 (**Fig. 2**). In the HbRC and PSI, these distances are 2-4 Å longer than in the PbRC and PSII. This means that the ec3-to- F_X distance in Type I RCs is 2-4 Å closer than it would be if they had retained the Type II ec2 orientation.

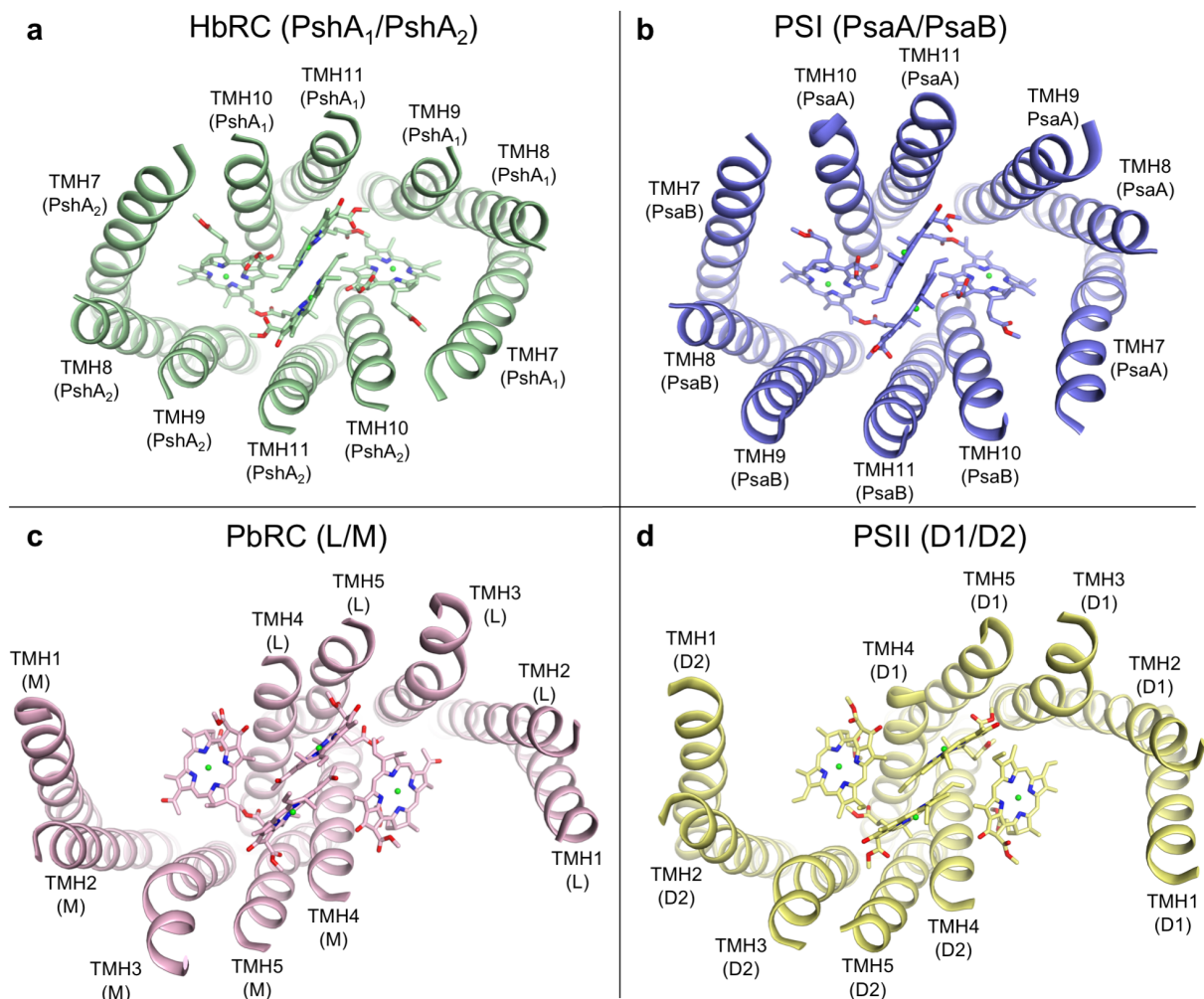


Fig. 7 TMH comparison between RC classes. A donor-side view of the HbRC (**a**, green, PDB ID: 5V8K), PSI (**b**, blue, PDB ID: 1JB0), PbRC (**c**, pink, PDB ID: 1PRC), and PSII (**d**, yellow, PDB ID: 3WU2) TMHs. P and ec2 cofactors are shown for orientation.

Another advantage of the new orientation of ec2 is that the vector of $ec2^+ec3^-$ dipole (assuming that this is the initial charge-separated state, as it is thought to be in PSI; (Müller et al. 2003, 2010)) would be better aligned to take advantage of the static electric field across the RC. As mentioned above, the use of a (B)Chl in place of a B(Pheo) in the ec3 site would lower the reduction potential of ec3, and thereby raise the energy of the $ec2^+ec3^-$ state. The perpendicular orientation of ec2 would make the $ec2^+ec3^-$ dipole more perpendicular to the membrane plane, thus perpendicular to the direction of the static electric field across the membrane, caused by the overall negative charge on the donor side and the positive charge on the acceptor side of the RC

(Gunner et al. 1996; Gisriel et al. 2017). This would lower the energy of the $ec2^+ec3^-$ ion pair back down.

The use of a Tyr in the P-helix as a H-bond donor to $ec2$ in the HbRC raises an intriguing possibility for the transition from the Type II to the Type I orientation of $ec2$. We suggest that in the ARC, a residue in this position (likely a His) was the axial ligand to $ec2$. After the introduction of a residue in TMH9 that could serve as axial ligand, the change in axial ligand was facilitated by conservation of the interaction with the P-helix residue, although its character was changed: from axial ligand of the central Mg(II) to H-bond donor to a peripheral functional group (carbonyl O of the methyl ester substituent on ring E). Note that this carbonyl group is not part of the conjugated π system of the BChl g in the HbRC, and would therefore have little effect on its redox or optical properties. Even a deprotonated imidazole can donate a H-bond, and a phenol can serve as an axial ligand, so it is not clear if a His or a Tyr were the ancestral residue before the change in $ec2$ orientation. In any case, the use of Tyr allows a strong H-bond without introducing positive charge in the vicinity of P, which may explain the presence of a Tyr in all HbRCs known today (**Supplementary Information, Fig. S1**).

1. Prime chlorophylls (Chl') in the special pair

The functional role of the prime isomer of (B)Chl (*e.g.* BChl g' and Chl a'), which possesses reversed stereochemistry about the 13^2 carbon in ring E of the tetrapyrrole, in the special pair of Type I RCs has long been a question. The HbRC uses a BChl g' homodimer as P_{800} and the GsbRC is predicted to use a BChl a' homodimer as P_{840} (Prince et al. 1985; Kobayashi et al. 1992; Ohashi et al. 2010; Sarrou et al. 2012). Chl a - and Chl d -containing PSI use a Chl a /Chl a' or Chl d /Chl d' heterodimer as P_{700} or P_{740} , respectively (Fromme et al. 2001; Akiyama et al. 2002; Ohashi et al. 2010). Therefore, in Type I RCs, the homodimeric variety employs homodimeric BChl'/BChl' as the special pair, while the heterodimeric variety employs heterodimeric Chl/Chl' as the special pair. Initial high-resolution X-ray crystal structures of PSI revealed that the P_{700} Chl a' participates in a H-bonding network. Specifically, these H-bonds are from PsaA-Tyr735 (to the phytyl chain ester group), PsaA-Thr743 (to the 13^1 -keto O) and a water molecule (to the C- 13^2 ester group). This water molecule may additionally interact with PsaA residues Ser607, Thr743, Tyr603, and Gly739. Because H-bonding influences the spin density distribution of excited states, it was initially hypothesized that the majority of spin density for P_{700}^+ would be localized to the non-hydrogen-bonded Chl a , which would bias ET toward one branch of the ET chain (Jordan et al. 2001; Fromme et al. 2001). However, this idea was cast into doubt after it was shown that both branches of PSI were active in electron transfer (Guergova-Kuras et al. 2001; Muhiuddin et al. 2001; Poluektov et al. 2005; Santabarbara et al. 2005).

The situation is different in the HbRC: the two BChl g' molecules of P_{800} do not participate in any H-bonding whatsoever. This indicates that the H-bonding environment around the special pair neither dictates, nor is required for, the presence of the prime epimer. The best explanation for the presence of prime epimer is minimization of steric clash; the (B)Chls' simply fit better into the cavity afforded for the special pair at the Type I RC dimer interface. In the HbRC, if the P_{800} BChl g' molecules were replaced with BChl g molecules, the 13^2 -methoxycarbonyl group would likely sterically clash with residues of TMH 11, particularly Thr598 and Cys601. This conclusion is reinforced by the observation of a second BChl g' per PshA in the HbRC structure, which is found in the antenna domain and is coordinated by His36 of TMH1. There is also no H-bonding between this BChl g' and the protein. However, if a

normal BChl *g* were in this site, its 13²-methoxycarbonyl group would likely clash with a neighboring antenna BChl *g* that is coordinated by a water molecule via Lys596 of TMH11. Furthermore, if the Chl *a'* of P₇₀₀ in PSI were a Chl *a*, it would likely clash with Phe598 of TMH9. As shown in **Fig. 7**, with the change in ec2 orientation in Type I RCs comes a subtle change in the orientation of the TMHs. The changes in the orientation of TMH9 and TMH11 specifically will affect the binding of P. We believe that the result of these changes is that only the prime epimer of the pigment in the P site could be accommodated due to steric effects.

No enzyme responsible for making the C13² epimer of any (B)Chl has been identified. An intriguing hypothesis put forward by Webber and Lubitz (Webber and Lubitz 2001) was that the PSI polypeptide itself catalyzed the conversion of Chl *a* to a Chl *a'*, which would then be stabilized by optimal H-bonding to the nearby water molecule coordinated by several PsaA residues (Tyr603, Ser607, Thr743, Gly739). Residue PsaA-Thr739 would play a crucial role in the postulated mechanism, as the H-bond it donates to the 13¹-keto oxygen could stabilize an enolate intermediate after deprotonation of C13². However, mutation of this Thr to Ala did not result in loss of Chl *a'* (Li et al. 2004), largely disproving this hypothesis. The simplest explanation is that (B)Chl synthesis can result in a small amount of the C13² epimer. The enzyme responsible for closure of ring E might even be capable of interconverting the two isomers at a low rate. However, when the C13² epimer is made, selection for it at a specific site would be according to whichever isomer fits best. Although both epimers could compete for any given (B)Chl site, the low amount of the prime C13² epimer would result in it losing this competition except for those sites where the much more abundant (non-prime) epimer could not bind.

With all of these lines of evidence taken together, we believe that there is no energetic role for (B)Chl'. There is only a structural role – and a relatively trivial one. We also believe it likely that the LCA of Type I RCs (Proto-RC1) contained two (B)Chl' in its special pair, driven by the change in ec2 orientation. Thus, if the ARC contained the same ec2 orientation as modern Type II RCs, then it likely did not contain prime (B)Chls as P. The shift to use of the C13² epimer in the P site would have been an automatic consequence of the change in ec2 orientation and concomitant rearrangement of the ET domain TMHs.

The evolution of Photosystem I was profoundly affected by the rise of atmospheric oxygen

Now that we have explained the changes in the ET cofactors that are common to all Type I RCs, we will move on to discussing the changes that resulted in the split between the aerobic PSI and anaerobic HbRC/GsbRC/CabRC lineages. The key to understanding these changes lies in the need to deal with dioxygen, which was the product of an ancestor of PSII that had acquired the oxygen-evolving complex (OEC).

The rise of dioxygen forced most anaerobic organisms containing a homodimeric Type I RC to avoid aerobic conditions, due to the intense stress they would have experienced. This stress would have come in the form of singlet oxygen (¹O₂), produced from a reaction between O₂, which has a triplet ground-state (³O₂), and the P triplet state (³P), which is generated by CR of P⁺ec3⁻. ET from ec3 to F_X ($\tau \approx 0.8$ ns) is only ~20-fold faster than CR of P⁺ec3⁻ ($\tau \approx 20$ ns). This ratio of forward:reverse ET rate is lower than any other ET step in Type I RCs and means that CR would occur in a small but significant fraction of CS events, even if F_X were oxidized and available to accept an electron. Some organisms with Type I RCs that could not avoid aerobic conditions overcame this oxidative stress by modifying their Type I RC. We assume that most of these changes necessary to fashion PSI, especially the last ones, took place in the

1 presence of a water-oxidizing Type II RC (either PSII or a close ancestor). We term the Type I
2 RCs that followed this scenario “Ancestral PSI”. Every ET-domain modification that occurred
3 on the way from the LCA-RC1 to Ancestral PSI can be explained by the avoidance of singlet
4 oxygen production, which is achieved by minimizing the P^+ec3^- CR reaction.

5 To successfully avoid CR of the P^+ec3^- state, the quinones in the LCA-RC1 were moved
6 further into the interior of the RC to serve as tightly-bound ET intermediates between $ec3$ and
7 F_X , losing mobile quinone reduction activity in the process. It must be emphasized that this
8 change does *not* enhance the rate of ET from $ec3$ to F_X ; the LCA-RC1 could already efficiently
9 do this on the sub-ns timescale, as it does in the HbRC now (Chauvet et al. 2013; Ferlez et al.
10 2016). In fact, the overall ET rate from P to F_X is much slower in PSI using the tightly-bound
11 phylloquinone (tens-to-hundreds of ns) than it is in the HbRC (< 1 ns). Slowing the rate of
12 forward ET may seem counter-intuitive from a gain-of-fitness perspective. However, the point of
13 this change was not to increase the overall rate of ET to F_X , but to minimize P^+ec3^- CR. The key
14 is that CR of the P^+Q^- state in PSI does not produce a $^3P_{700}$ triplet (Warren et al. 1993). Thus,
15 once the electron is on the quinone, CR is “safe”. In order to complete the transition to
16 permanently-bound quinone, the protonation channels were lost so that the quinone could not be
17 double reduced and/or leave the site. Note that if the quinone-binding site moved, the new site
18 would be cut off from these channels automatically; afterwards the residues making up the
19 channel could be lost slowly over time. The final effect is that the electron is always transferred
20 rapidly from $ec3$ to the quinone, due to the short distance and very strong driving force. The
21 system loses so much energy that the next step in PSI (ET from Q to F_X) is almost isoenergetic
22 or slightly uphill (Santabarbara et al. 2005). The fact that this last step slows the overall rate of
23 ET from P to F_X does not matter in the end, since the net electron throughput of the RC is not
24 determined by this rate; it is instead determined by the diffusion rate of exogenous electron
25 donors and acceptors to and from the RC.

26 A concerted movement of the $ec3$ and quinone sites during the transition from LCA-RC1
27 to PSI may not have been so difficult. Key residues responsible for binding these two cofactors
28 are next to each other in the polypeptide sequence (**Fig. 8**). The quinone-binding site of PSI is
29 made up of 3 key residues: a Trp that stacks aromatically with the naphthoquinone headgroup on
30 one side, and a Phe and Leu sidechain on the other side. On TMH10, the axial ligand to $ec3$ (Ser
31 in HbRC, Met in PSI) is right before the Phe (Met in the HbRC). On the surface helix
32 immediately after TMH10, the H-bond donor to $ec3$ (Ser in the HbRC, Tyr in PSI) is right before
33 the Trp (Arg in the HbRC). The Leu, which is on the loop immediately before TMH11, is present
34 in the HbRC. Thus, one could imagine that only two mutations, each of which affected two
35 neighboring residues, could have accomplished most of the work of pushing $ec3$ up slightly and
36 creating a new quinone-binding site. The change of axial ligand from Ser to Met would have
37 tuned the reduction potential of $ec3$ and allowed its subsequent move; the change of Met to Phe
38 may not have affected the structure much at all. Using a longer residue as H-bond donor (Ser to
39 Tyr) would have had the effect of pushing $ec3$ slightly from the membrane surface, allowing the
40 quinone to bind by aromatically stacking with the new Trp residue, with the Phe and Leu on the
41 other side to provide a hydrophobic environment for the menaquinone headgroup.

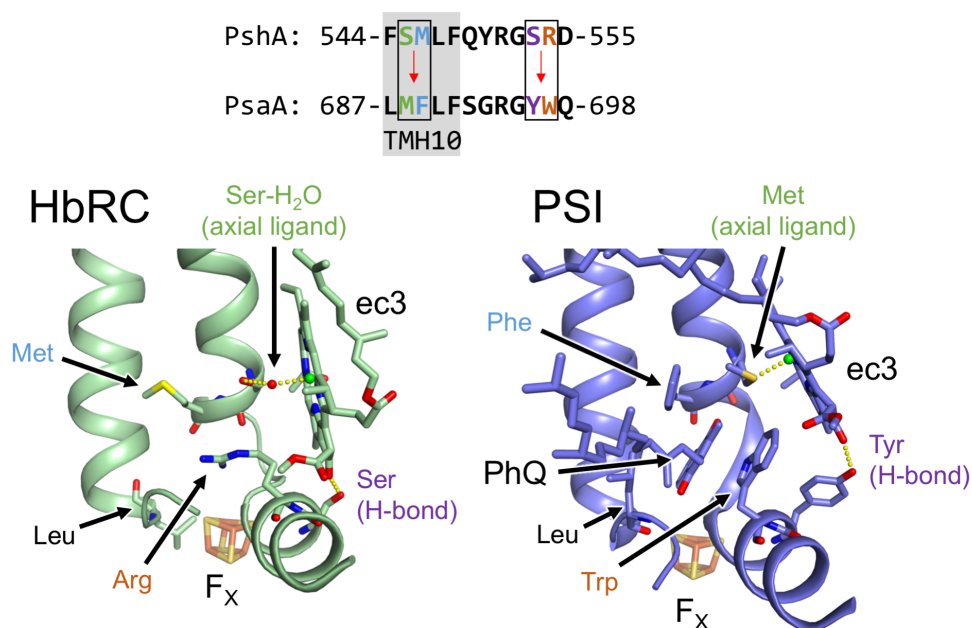


Fig. 8 Comparison of the ec3 and quinone binding sites in the HbRC and PSI. Two sets of mutations define the major differences in the region. The first set, adjacent residues at the end of TMH10, control the ec3 axial ligand and the top of the quinone pocket. The second set, adjacent residues in a surface helix after TMH10, control an ec3 H-bond donor and the interior of the quinone pocket. In particular, the side chain of an Arg residue present in the HbRC blocks the putative Q pocket from penetrating the space between ec3 and F_X. In PSI, this residue is a Trp, whose side chain “flips upward” (relative to the Arg in the HbRC) to tightly bind a PhQ molecule into the space between ec3 and F_X.

As the residues in the two phyloquinone-binding sites in PSI are highly conserved between the PsaA and PsaB subunits, the burying and isolation of the quinone sites must have occurred in the homodimeric state, before diversification of the core polypeptide into PsaA and PsaB. It has been argued that the asymmetry of the quinone sites in PSI is due to removing the risk of ¹O₂ production: as the semiquinone in the Q_A site is lower in energy than in the Q_B site, CR from F_X would precede through Q_A, subsequently tunneling to the ground state (Rutherford et al. 2012). From our argument above, it seems likely that in the homodimeric state the quinone would have been more like Q_A (*i.e.* lower in energy in the semiquinone state). After formation of the heterodimer, one of the quinones was free to raise its energy, although the reason for this is presently unclear.

Ancestral PSI would have dealt with a second issue related to the presence of atmospheric oxygen: long-lived F_X⁻ states can generate superoxide (O₂⁻), another reactive oxygen species (ROS). However, unlike singlet oxygen, a biological remediation pathway does exist to eliminate superoxide in the form of the enzyme superoxide dismutase. The product of this enzyme is hydrogen peroxide, which can be reduced to water by peroxidases or dismutated by catalase. In effect, Ancestral PSI sacrificed its ability to reduce mobile quinones to prevent formation of an ROS that the cell could not remediate (singlet oxygen), in the process allowing occasional formation of an ROS the cell could remediate (superoxide). The sacrifice was not a large one; the loss of this ability in PSI would have further strengthened the linear electron flow pathway from PSII (which could still reduce quinone) to NADPH.

1 A way to spatially separate the charge-separated state even further was to recruit a
2 modified version of the ferredoxin acceptor as a permanently bound subunit. In Modern PSI, this
3 is the PsaC subunit, which contains the F_A/F_B clusters and is bound to the PsaA/PsaB core. As
4 one would expect, the crystal structures of PSI reveal that PsaC binds asymmetrically to the
5 PsaA/PsaB heterodimer. Therefore, heterodimerization likely occurred in Ancestral PSI to
6 encourage stronger binding of PsaC, resulting in Modern PSI. The result of this would be to
7 further stabilize the CS state, increasing its lifetime to ~ 100 ms. Without the F_A/F_B clusters, if the
8 acceptor pool were heavily reduced, the $Q^+F_X^-$ state could eventually accumulate under high
9 light, leading to detrimental CR of P^+ec3^- . With the additional two clusters, PSI would have to
10 accumulate 4 electrons – on all 3 Fe-S clusters as well as one of the quinones – before P^+ec3^- CR
11 would occur, making this situation very unlikely. It should be noted that this can occur under
12 high light in anaerobic conditions, but the lack of atmospheric oxygen removes the danger of
13 singlet oxygen or superoxide, thus explaining why an ancestor to the HbRC/GsbRC/CabRC did
14 not fix their quinones or heterodimerize; they were not forced to acquire a mechanism to avoid
15 ROS.

16 *Recruitment of the antenna domain*

17 Finally, we will consider a last piece of the evolutionary puzzle unrelated to the tuning of
18 the redox function of the ET cofactors: the recruitment of the antennas. If we imagine that the
19 ARC had an antenna domain fused to the N-terminus of the ET domain, similar to what is found
20 in modern Type I RCs, this domain must have been lost in the PbRC lineage and split off into a
21 separate subunit in the Ancestral PSII lineage. The latter change is a relatively minor one, but the
22 former is a problem. What possible advantage would the ancestor to the PbRC have gained by
23 *losing* so many antenna pigments? One cannot use the argument that this allowed the LH1
24 antenna complex to bind to it, as LH1 would not have been “waiting in the wings” for the
25 antenna domain to be lost so it could associate with the ET domain. As mentioned before, we
26 will avoid any step that results in a loss of fitness unless there is a compelling argument for it.
27 There is no such argument here. It seems much more likely that the ARC lacked an antenna
28 domain and the lineages that evolved from it developed their own antennas, which over time may
29 have been transferred around via LGT. In this final section, we will consider the LGT events in
30 which the 6-TMH antenna domain found in HbRC/GsbRC/CabRC/PSI/PSII radiated, how the
31 presence of antenna (B)Chls in the ET domain relate to the antenna domain radiation, and
32 finally, the evolution and radiation of single-TMH antenna subunits.

33 1. LGT of the 6-TMH antenna domain

34 During analysis of the new HbRC structure, a curious finding was the HbRC antenna
35 domain and CP43/CP47 antenna structurally align with lower RMSD than the HbRC antenna
36 domain and PSI antenna domains do (Gisriel et al. 2017). There is also a distinct β -hairpin
37 structure conserved between the HbRC, CP43, and CP47, on the donor side of the antenna
38 domains, which does not appear in PSI (**Supplementary Information, Fig. S6**). This offers a
39 surprising contrast with the superpositions of the ET domains and points to a more complicated
40 evolutionary history than previously thought.

41 The simplest explanation we can propose is that Ancestral PSII gained its antenna
42 domain via LGT from an ancient Type I RC that already contained the antenna domain (**Fig. 4**).
43 The LGT event would have transferred a single gene for the antenna domain without fusing it to
44 the Ancestral PSII core (which had already heterodimerized). Such an event would have
45
46

1 occurred before the Ancestral PSII gained an OEC, because the antenna domain is a requisite
2 partner for binding the OEC in Modern PSII. From there, the unfused antenna polypeptide of
3 Ancestral PSII would have later diverged into modern CP43 and CP47.

4 If the Type I RC ancestor serving as the donor of the antenna domain transferred the
5 antenna domain to Ancestral PSII *before* the split of the homodimeric Type I RCs and PSI, it
6 would imply that both the HbRC antenna domain and CP43/CP47 have structurally diverged
7 little while the PSI antenna domain has structurally diverged more, a scenario we have no
8 justification for. If, however, the ancient Type I RC that transferred the antenna domain did so
9 *after* the split of the homodimeric Type I RCs and PSI, *and* was an ancestor in the
10 HbRC/GsbRC/CabRC lineage, it would provide a good explanation as to why the HbRC antenna
11 domain and CP43/CP47 remain relatively similar.

12 An alternative evolutionary scheme in which the ancestral antenna originated in
13 Ancestral PSII and was transferred via LGT to a lineage containing Proto-RC1 is less
14 compelling. Due to the distribution of the antenna domains, the ancient LGT event would have
15 had to occur from Ancestral PSII *before* the Type I lineages split. As explained above, the fact
16 that the HbRC antenna and CP43/CP47 have structurally diverged little, while PSI has
17 structurally diverged more, cannot be easily explained by this scenario. Secondly, the pattern of
18 (B)Chls that are bound by the ET domain but function as antenna pigments (discussed below)
19 does not support this scenario.

20 21 2. Antenna (B)Chl sites in the ET domains

22 There are several antenna (B)Chls in the Type I RCs that are coordinated by residues in
23 the ET domain (**Fig. 9**). In the HbRC, 7 BChl *g* molecules are bound in this fashion by PshA.
24 Almost twice as many Chl *a* are bound by the ET domains of PsaA (11) and PsaB (12) in
25 cyanobacterial PSI; these include the intimately-associated “red-absorbing” chlorophylls (Jordan
26 et al. 2001) which are not found in the HbRC. In sharp contrast, the PSII core polypeptides (D1
27 and D2) each bind a single antenna Chl, called Chl_Z/Chl_D (or sometimes D1-Chl_Z/D2-Chl_Z)
28 (Zouni et al. 2001). No antenna sites are found in the structures of the L/M heterodimer of the
29 PbRC (Deisenhofer et al. 1985); the site of the His residue used to bind Chl_Z/Chl_D in PSII is
30 generally a Phe or Cys residue in the PbRC. Based upon our sequence alignments, the
31 Chloroflexi RC (CfxRC) and Gemmatimonadetes RC (GmRC) are not expected to contain the
32 Chl_D/Chl_Z site either (**Supplementary Information, Fig. S7**). All antenna (B)Chls bound by the
33 ET domains are shown in **Figs. 9a**.

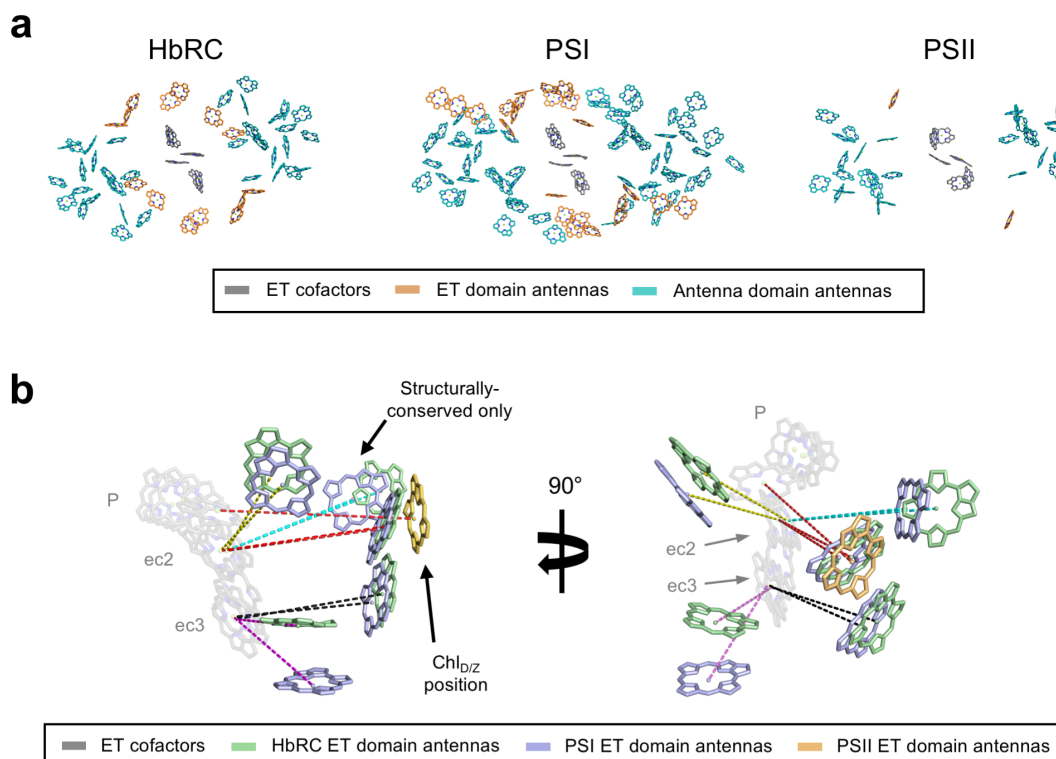


Fig. 9 Comparison of antenna sites provided by the ET domains of the HbRC, PSI, and PSII. **(a)** All pigment sites are shown from donor-side view and are color-coded according to their role and position. **(b)** Structurally-conserved ET domain antenna sites within 20 Å (center-to-center) of any ET cofactors are shown from an in-membrane perspective and are color-coded by RC. Pigment macrocycles are shown alone to facilitate viewing. In **(b)**, dotted lines connect the antenna site center-to-center with the closest ET cofactor; lines are color-coded by conserved site.

There are four ET domain-bound antenna (B)Chl sites that are conserved in the HbRC and PSI in terms of sequence (axial ligand) and spatial position. One of these sites is analogous to Chl_Z/Chl_D in PSII. There is also a fifth site in which (B)Chls are in a similar position in the structure, but the coordinating residues are not homologous (**Figs. 9b**). Three of the five aforementioned sites are situated more closely to ec2, while the others lie more closely to ec3. We believe energy transfer from the antenna to ec3 to be unlikely in the HbRC because ec3 is a Chl *a* derivative, thus requiring an uphill step from the lower-energy BChl *g* antenna. The same situation is likely to hold true in the GsbRC and CabRC, since their antenna pigments are BChl *a* and ec3 is also a Chl *a*. Of course, this is not the case in PSI, where all the pigments are Chl *a*.

Since the Chl_Z/Chl_D site is conserved across the HbRC, GsbRC, PSI, and PSII, and lies most closely to the ec2 site, a logical conclusion is that it is the major bridging site that energetically connects the antenna domain (or CP43/CP47) to the ET chain, with energy transfer usually proceeding to the ET chain to initiate CS through the ec2 site, which does not require uphill energy transfer. Curiously, the sequence of the CabRC predicts a Trp residue at the Chl_Z/D site, which may not ligate a BChl (see **Supplementary Information, Fig. S7**). Since none of the anoxygenic Type II RCs (which lack the 6-TMH antenna domain) contain a Chl_D/Chl_Z-type bridging site, but PSII and most known Type I RCs do, a simple evolutionary scenario to explain this is that the Chl_D/Chl_Z site was present in the ARC and pre-dates the Type

I/II split. This site may have been important for either aiding in energy transfer from a now-extinct primordial antenna complex or increasing photon absorption by the ET domain. Following this scenario, this antenna site was lost in the last common ancestor of the PbRC/CfxRC/GmRC. The drive to lose this site may have been the recruitment of an ancient LH1-like antenna. In the three-dimensional structures from extant PbRCs, the Chl_D/Chl_Z site occupies the same space as an LH1 polypeptide; inclusion of a BChl at this position would likely clash with LH1 and inhibit its association with the RC (**Supplementary Information, Fig. S7**). Considering that the closest distance between any BChl in LH1 and the ET core in the PbRC is about 11 Å longer than the distance between the Chl_D/Chl_Z site and the ET core, the ancestor of the PbRC may have sacrificed a close distance between antenna pigments and the ET core to gain a large increase in antenna size. The Chl_D/Chl_Z site was probably advantageous for Ancestral PSII to maintain, if only for using it as an antenna site. This advantage was greatly magnified after the recruitment of ancestral CP43/CP47 and the OEC, to enhance energy transfer efficiency to the ET cofactors (Schelvis et al. 1994; Lince and Vermaas 1998) and to act in a photoprotective role in oxygenic conditions (Schweitzer and Brudvig 1997; Schweitzer et al. 1998).

The larger number of antenna sites in the Type I ET domains (*i.e.* 6-12) implies that the antenna domain has been present in this lineage longer, enough time for the Type I RCs to add more bridging sites between ET and antenna domains. If PSII was the progenitor of the antenna domain, one would expect that it would have evolved more ET domain-bound antenna sites to aid in energy transfer to the ET core. Therefore, according to our evolutionary scheme, the antenna domain was gained by an ancestor of all modern Type I RCs, after the acquisition of F_X but before the diversification of the HbRC/GsbRC/CabRC and PSI. So, the first complex labelled in our evolutionary scheme to contain this domain is LCA-RC1. After the split between the lineages that led to modern homodimeric RCs and PSI, an ancestor to modern homodimeric RCs laterally transferred this antenna domain to an ancestor of PSII.

3. Evolution of single-TMH extrinsic antenna subunits

A surprising discovery in the recent HbRC structure was the presence of two single-TMH subunits, one on each side of the core. Although this polypeptide had not previously been documented or annotated as a coding sequence in the genome, clear positive density in the F_O-F_C map indicated its presence. Its primary sequence was identified by crystallography and mass spectrometry (Gisriel et al. 2017). Attempts to identify a homolog to the *pshX* gene in the *Heliorestis convoluta* draft genome have so far been unsuccessful. The single-TMH subunits PsaI and PsaJ are present in PSI in the same general position, but no sequence homology can be seen between these polypeptides and PshX. Interestingly, the gene encoding PshX lies outside of the photosynthetic gene cluster (PGC), which contains *pshA*, the genes for cyt *c*₅₅₃, the cyt *bc* complex, and pigment synthesis. This suggests that the gene encoding PshX may not have been transferred with the other genes in the PGC and was an independent evolutionary invention of the Heliobacteria. This implies that small transmembrane subunits may be cases of convergent evolution. Its small size (31 residues) makes its *de novo* evolution conceivable. Genomic DNA sequences capable of encoding hydrophobic peptides of sufficient length to cross the membrane are not very rare (Hemm et al. 2008). If one of these had some affinity for the HbRC and managed to render the complex more stable in some way, then it would be selected for. The same sort of selection may well explain the numerous small hydrophobic subunits at the periphery of PSI, PSII, cyt *b*_{6f}, NADH dehydrogenase, etc. Often, deletion of such subunits

1 results in a lower steady-state level of the complex due to more rapid degradation (*e.g.* PsaJ in
2 eukaryotic PSI (Hansson et al. 2007)). In any case, the PshX subunit found today has two
3 antenna pigment (BChl *g*) binding sites, and thus contributes to 4 of the 54 antenna BChls (*i.e.*
4 7.4%). The conclusion that transmembrane proteins may associate with the RC promiscuously,
5 exhibiting convergent evolution, allows us to remain uncommitted as to whether the LCA of
6 Type I RCs may have exhibited transmembrane subunits other than the dimeric core.

7 8 **Conclusion and Outlook**

9 The recent 2.2-Å crystal structure of the HbRC, as well as its unexpected quinone
10 reduction activity, has caused us to re-evaluate our thinking on the evolution of the
11 photochemical RCs. Our proposed evolutionary trajectory rests upon the hypothesis that the last
12 common ancestor of all RCs was a homodimeric complex that functioned to reduce quinone to
13 quinol using an inefficient radical disproportionation reaction, allowing for a new innovation in
14 the biosphere: light-driven cyclic electron flow. All of the diversification presently observed
15 across the disparate phototrophic taxa are attributable to the need to, first, optimize this reaction,
16 and much later, adjust to rising oxygen concentrations. We acknowledge the limitations inherent
17 in trying to understand protein-level changes that occurred over three billion years ago.
18 However, we believe that our approach, which considers every *functional* aspect of the RC, can
19 help fill in informational gaps that genetic and structural data alone cannot fill. Recent advances
20 in structural biology methods allow for optimism that representative RC structures from more
21 phototrophic groups will eventually be solved, and recent advances in metagenomic analyses
22 allow for optimism that new and interesting phototrophs will be discovered. Marrying these
23 future findings with our underlying functional analysis will help to further refine this
24 evolutionary picture.

25 26 **Methods**

27 *Sequence alignments and structural visualization*

28 For non-structure-based sequence alignments, each sequence was retrieved from the
29 Uniprot database (The UniProt Consortium 2017) and Clustal Omega (Sievers et al. 2011) was
30 used for initial sequence alignments. For structure-based sequence alignments, the RC structures
31 were gathered from the protein data bank (PDB) (Berman et al. 2000), trimmed according to the
32 domain of interest in PyMol (2014), and all amino acids were changed into alanines using Coot
33 (Emsley et al. 2010). **Table S1 (Supplementary Information)** provides a list of all structures
34 used for the structure-based phylogenetic analysis for the ET domains. Two structure-based
35 MSAs were constructed for each dataset: one from the PROMALS3D server (Pei et al. 2008) and
36 one from the Protein Data Bank in Europe PDBeFold server (Krissinel and Henrick 2005).
37 Together with the non-structure-based method, this resulted in a total of 3 MSAs (two structure-
38 based and one sequence-based). When showing the three-dimensional overlay of RC structures,
39 the *super* function of PyMol was used. When comparing RC polypeptides which do not have
40 available structures (such as the CfxRC), these sequences were fetched from the Uniprot
41 database.

42 43 *MSA trimming and model selection*

44 We applied 3 different data trimming strategies to each MSA: no gap deletion, 50% site
45 coverage cutoff, and full deletion for positions containing gaps. To determine which model
46 should be used to assess the evolutionary relationships between the non-structure-based MSAs

(Clustal Omega) and structure-based MSAs (PROMALS3D and PDBeFOLD), the “Find Best Protein Model” function of the MEGA7 software (Kumar et al. 2016), which ranks 56 protein substitution models according to Bayesian and Akaike information criteria and the Maximum Likelihood fit, was used. The top three amino acid substitution models for each MSA were variations of LG (Le and Gascuel 2008) and WAG (Whelan and Goldman 2001) (**Supplementary Information, Table S6**). The three models for each data set were identified with the top-scoring specification ((Whelan and Goldman 2001) with or without 4 gamma distributed rate categories, and with or without proportions of invariant sites or amino acid frequencies from the data). Phylogenetic trees were built under each of the top three models, for a total of 27 phylogenetic trees (3 MSAs *times* 3 trimming methods *times* 3 best substitution models; **Supplementary Information, Table S7**). Support for nodes in these 27 trees were assessed by 100 bootstrap replicates.

The resulting 3 groups of 9 trees were examined for topological variation and a tree that exhibited the most common and well-supported topology was chosen to represent the overall data and re-created with 1000 bootstrap replicates: one for the non-structure-based tree (**Fig. 3a**) and one for the structure-based tree (**Supplementary Information, Fig. S4**) For all trees, a pairwise distance matrix was estimated with a JTT model and the Neighbor-Join and BioNJ algorithms were applied to identify the initial tree for the heuristic search. The final non-structure-based tree was made with the WAG model (Whelan and Goldman 2001) with frequencies (+F) and gamma-distributed rates (+G, 4 categories, parameter = 8.1159), which exhibited a highest log likelihood value of -7448.88, and the structure-based tree was made from the PROMALS3D MSA with the LG model (Le and Gascuel 2008) with frequencies (+F), gamma-distributed rates (+G, 4 categories, parameter = 5.6588), and invariant sites (+I), which exhibited a maximum log likelihood value of -8785.34.

Phenetic dendrogram construction

The 19 individual trimmed structures (see “Sequence alignments” section above) were superimposed onto one another in pairwise fashion, biased by TMHs, using the *super* function in PyMol (2014) and a triangular RMSD matrix was created (for the PyMol script formula, see **Supplementary Information, Table S8**). The matrix was converted into MEGA format, read into MEGA7 (Kumar et al. 2016), and a phenetic dendrogram was inferred using the UPGMA method (Sneath and Sokal 1973). The optimal dendrogram with the sum of branch length = 9.821 is shown. The dendrogram is drawn to scale, with branch lengths in the same units as those of the RMSD distances used to infer it. The UPGMA method adds a root at the deepest node automatically in order to calibrate the RMSD distance scale.

References

- Akiyama M, Miyashita H, Kise H, et al (2002) Quest for minor but key chlorophyll molecules in photosynthetic reaction centers - Unusual pigment composition in the reaction centers of the chlorophyll d-dominated cyanobacterium *Acaryochloris marina*. *Photosynth Res* 74:97–107. doi: 10.1023/A:1020915506409
- Albert I, Rutherford AW, Grav H, et al (1998) The 18 kDa cytochrome c553 from *Heliobacterium gestii*: Gene sequence and characterization of the mature protein. *Biochemistry* 37:9001–9008. doi: 10.1021/bi9731347
- Allen JF (2005) A redox switch hypothesis for the origin of two light reactions in photosynthesis. *FEBS Lett* 579:963–968. doi: 10.1016/j.febslet.2005.01.015
- Bak DW, Elliott SJ (2014) Alternative FeS cluster ligands: Tuning redox potentials and chemistry. *Curr Opin Chem Biol* 19:50–58. doi: 10.1016/j.cbpa.2013.12.015
- Barber J (2012) Photosystem II: The water-splitting enzyme of photosynthesis. *Cold Spring Harb Symp Quant Biol* 77:295–307. doi: 10.1101/sqb.2012.77.014472
- Battistuzzi FU, Hedges SB (2009) A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol* 26:335–343. doi: 10.1093/molbev/msn247
- Berman HM, Westbrook J, Feng Z, et al (2000) The protein data bank. *Nucleic Acids Res* 28:235–242. doi: 10.1093/nar/28.1.235
- Blankenship RE (2014) *Molecular Mechanisms of Photosynthesis*, 2nd edn. John Wiley & Sons, Ltd., Southern Gate, Chichester, West Sussex, UK
- Blankenship RE (1992) Origin and Early Evolution of Photosynthesis. *Photosynth Res* 33:91–111. doi: 10.1104/pp.110.161687
- Blankenship RE, Sadekar S, Raymond J (2007) The Evolutionary Transition from Anoxygenic to Oxygenic Photosynthesis. In: Falkowski PG, Knoll AH (eds) *Evolution of Primary Producers in the Sea*. Elsevier, pp 21–35
- Brettel K, Liebl W, Liebl U (1998) Electron transfer in the heliobacterial reaction center: Evidence against a quinone-type electron acceptor functioning analogous to A1 in Photosystem I. *Biochim. Biophys. Acta - Bioenerg.* 1363:175–181.
- Brok M, Vasmel H, Horikx JTG, Hoff AJ (1986) Electron transport components of *Heliobacterium chlorum* investigated by EPR spectroscopy at 9 and 35 GHz. *FEBS Lett* 194:322–326. doi: 10.1016/0014-5793(86)80110-7
- Bryant DA, Frigaard N-U (2006) Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* 14:488–496. doi: 10.1016/j.tim.2006.09.001
- Bryant DA, Liu Z, Li T, et al (2012) Comparative and Functional Genomics of Anoxygenic Green Bacteria from the Taxa Chlorobi, Chloroflexi, and Acidobacteria. In: Burnap R, Vermaas W (eds) *Functional Genomics and Evolution of Photosynthetic Systems*. Springer Netherlands, Dordrecht, pp 47–102
- Cardona T (2015) A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynth Res* 126:111–134. doi: 10.1007/s11120-014-0065-x
- Cardona T (2016a) Reconstructing the Origin of Oxygenic Photosynthesis: Do Assembly and Photoactivation Recapitulate Evolution? *Front Plant Sci* 7:1–16. doi: 10.3389/fpls.2016.00257

- 1 Cardona T (2016b) Origin of Bacteriochlorophyll a and the Early Diversification of Photosynthesis. PLoS One
2 11:e0151250. doi: 10.1371/journal.pone.0151250
- 3 Cardona T, Sanchez-Baracaldo P, Rutherford AW, Larkum AWD (2017) Molecular evidence for the early evolution
4 of photosynthetic water oxidation. doi: 10.1101/109447
- 5 Cardona T, Sedoud A, Cox N, Rutherford AW (2012) Charge separation in Photosystem II: A comparative and
6 evolutionary overview. Biochim Biophys Acta - Bioenerg 1817:26–43. doi: 10.1016/j.bbabi.2011.07.012
- 7 Chauvet A, Sarrou J, Lin S, et al (2013) Temporal and spectral characterization of the photosynthetic reaction center
8 from *Heliobacterium modesticaldum*. Photosynth Res 116:1–9. doi: 10.1007/s11120-013-9871-9
- 9 Chen M, Telfer A, Lin S, et al (2005) The nature of the photosystem II reaction centre in the chlorophyll d-
10 containing prokaryote, *Acaryochloris marina*. Photochem Photobiol Sci 4:1060. doi: 10.1039/b507057k
- 11 Deisenhofer J, Epp O, Miki K, et al (1985) Structure of the protein subunits in the photosynthetic reaction centre of
12 *Rhodospseudomonas viridis* at 3 Å resolution. Nature 318:618–624. doi: 10.1038/318618a0
- 13 Dibrova D V, Cherepanov DA, Galperin MY, et al (2013) Evolution of cytochrome bc complexes: From membrane-
14 anchored dehydrogenases of ancient bacteria to triggers of apoptosis in vertebrates. Biochim Biophys Acta -
15 Bioenerg 1827:1407–1427. doi: 10.1016/j.bbabi.2013.07.006
- 16 Diner BA, Petrouleas V, Wendoloski JJ (1991) The iron-quinone electron-acceptor complex of photosystem II.
17 Physiol Plant 81:423–436. doi: 10.1111/j.1399-3054.1991.tb08753.x
- 18 Doolittle RF (1986) Of urfs and orfs: a primer on how to analyze derived amino acid sequences. University Science
19 Books, Mill Valley, CA
- 20 Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr Sect D
21 Biol Crystallogr 66:486–501. doi: 10.1107/S0907444910007493
- 22 Etiope G, Sherwood Lollar B (2013) Abiotic Methane on Earth. Rev Geophys 51:276–299. doi: 10.1002/rog.20011
- 23 Fajer J, Brune DC, Davis MS, et al (1975) Primary charge separation in bacterial photosynthesis: oxidized
24 chlorophylls and reduced pheophytin. Proc Natl Acad Sci U S A 72:4956–60.
- 25 Ferlez BH, Cowgill JB, Dong W, et al (2016) Thermodynamics of the electron acceptors in *Heliobacterium*
26 *modesticaldum*: An exemplar of an early homodimeric type I photosynthetic reaction center. Biochemistry
27 55:2358–2370. doi: 10.1021/acs.biochem.5b01320
- 28 Fischer WW, Hemp J, Johnson JE (2016) Evolution of Oxygenic Photosynthesis. Annu Rev Earth Planet Sci
29 44:647–683. doi: 10.1146/annurev-earth-060313-054810
- 30 Fromme P, Jordan P, Krauß N (2001) Structure of photosystem I. Biochim Biophys Acta - Bioenerg 1507:5–31. doi:
31 10.1016/S0005-2728(01)00195-5
- 32 Fuller RC, Sprague SG, Gest H, Blankenship RE (1985) A unique photosynthetic reaction center from
33 *Heliobacterium chlorum*. FEBS Lett 182:345–349. doi: 10.1016/0014-5793(85)80330-6
- 34 Furbacher PN, Tae GS, Cramer WA (1996) Evolution and origins of the cytochrome bc1 and b6f complexes. In:
35 Baltscheffsky H (ed) Origin and Evolution of Biological Energy Conservation. Wiley-VCH, New York, pp
36 221–253
- 37 Garcia Costas AM, Liu Z, Tomsho LP, et al (2012) Complete genome of *Candidatus Chloracidobacterium*
38 *thermophilum*, a chlorophyll-based photoheterotroph belonging to the phylum Acidobacteria. Environ

1 Microbiol 14:177–190. doi: 10.1111/j.1462-2920.2011.02592.x

2 Gest H, Favinger JL (1983) Heliobacterium chlorum, an anoxygenic brownish-green photosynthetic bacterium
3 containing a “new” form of bacteriochlorophyll. Arch Microbiol 136:11–16. doi: 10.1007/BF00415602

4 Gisriel C, Sarrou I, Ferlez B, et al (2017) Structure of a symmetric photosynthetic reaction center–photosystem.
5 Science (80-) 357:1021–1025. doi: 10.1126/science.aan5611

6 Graige MS, Paddock ML, Bruce JM, et al (1996) Mechanism of proton-coupled electron transfer for quinone (QB)
7 reduction in reaction centers of *Rb. sphaeroides*. J Am Chem Soc 118:9005–9016. doi: 10.1021/ja960056m

8 Granick S (1957) Speculations on the Origins and Evolution of Photosynthesis. Ann N Y Acad Sci 69:292–308. doi:
9 10.1111/j.1749-6632.1957.tb49665.x

10 Granick S (1965) Evolution of Heme and Chlorophyll. ACADEMIC PRESS INC.

11 Guergova-Kuras M, Boudreaux B, Joliot A, et al (2001) Evidence for two active branches for electron transfer in
12 photosystem I. Proc Natl Acad Sci U S A 98:4437–4442. doi: 10.1073/pnas.081078898

13 Gunner MR, Nicholls A, Honig B (1996) Electrostatic Potentials in Rhodopseudomonas viridis Reaction Centers:
14 Implications for the Driving Force and Directionality of Electron Transferase. J Phys Chem 100:4277–4291.
15 doi: 10.1021/jp9519070

16 Hansson A, Amann K, Zygadlo A, et al (2007) Knock-out of the chloroplast-encoded PSI-J subunit of photosystem I
17 in *Nicotiana tabacum*. FEBS J 274:1734–1746. doi: 10.1111/j.1742-4658.2007.05722.x

18 Heathcote P, Fyfe PK, Jones MR (2002) Reaction centres: the structure and evolution of biological solar power.
19 Trends Biochem Sci 27:79–87. doi: 10.1016/S0968-0004(01)02034-5

20 Heinrickel M, Golbeck JH (2007) Heliobacterial photosynthesis. Photosynth Res 92:35–53. doi: 10.1007/s11120-
21 007-9162-4

22 Heinrickel M, Shen G, Golbeck JH (2007) Identification and characterization of PshB, the dicluster ferredoxin that
23 harbors the terminal electron acceptors FA and FB in Heliobacterium modesticaldum. Biochemistry 46:2530–
24 2536. doi: 10.1021/bi0622165

25 Hemm MR, Paul BJ, Schneider TD, et al (2008) Small membrane proteins found by comparative genomics and
26 ribosome binding site models. Mol Microbiol 70:1487–1501. doi: 10.1111/j.1365-2958.2008.06495.x

27 Hohmann-Marriott MF, Blankenship RE (2011) Evolution of photosynthesis. Annu Rev Plant Biol 62:515–548. doi:
28 10.1146/annurev-arplant-042110-103811

29 Joe H, Kuma K, Paplawsky W, et al (1986) Abiotic photosynthesis from ferrous carbonate (siderite) and water. Orig
30 Life Evol Biosph 16:369–370. doi: 10.1007/BF02422078

31 Jordan P, Fromme P, Witt HT, et al (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å
32 resolution. Nature 411:909–917. doi: 10.1038/35082000

33 Kashey TS, Luu DD, Cowgill JC, et al (2018) Light-driven quinone reduction in heliobacterial membranes.
34 Photosynth Res 1–9. doi: 10.1007/s11120-018-0496-x

35 Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol
36 Sci 365:729–748. doi: 10.1098/rstb.2009.0103

37 Khadka B, Adeolu M, Blankenship RE, Gupta RS (2017) Novel insights into the origin and diversification of

- photosynthesis based on analyses of conserved indels in the core reaction center proteins. *Photosynth Res* 131:159–171. doi: 10.1007/s11120-016-0307-1
- Kirmaier C, Gaul D, DeBey R, et al (1991) Charge separation in a reaction center incorporating bacteriochlorophyll for photoactive bacteriopheophytin. *Science* (80-) 251:922–927. doi: 10.1126/science.2000491
- Kirmaier C, Holten D, Parson WW (1985) Picosecond-photodichroism studies of the transient states in *Rhodospseudomonas sphaeroides* reaction centers at 5 K. Effects of electron transfer on the six bacteriochlorin pigments. *BBA - Bioenerg* 810:49–61. doi: 10.1016/0005-2728(85)90205-1
- Kirmaier C, Laporte L, Schenck CC, Holten D (1995a) The Nature and Dynamics of the Charge-Separated Intermediate in Reaction Centers in which Bacteriochlorophyll Replaces the Photoactive Bacteriopheophytin. 1. Spectral Characterization of the Transient State. *J Phys Chem* 99:8903–8909. doi: 10.1021/j100021a067
- Kirmaier C, Laporte L, Schenck CC, Holten D (1995b) The Nature and Dynamics of the Charge-Separated Intermediate in Reaction Centers in which Bacteriochlorophyll Replaces the Photoactive Bacteriopheophytin. 2. The Rates and Yields of Charge Separation and Recombination. *J Phys Chem* 99:8910–8917. doi: 10.1021/j100021a068
- Kleinherenbrink FAM, Chiou H-C, LoBrutto R, Blankenship RE (1994) Spectroscopic evidence for the presence of an iron-sulfur center similar to Fx of Photosystem I in *Heliobacillus mobilis*. *Photosynth Res* 41:115–123. doi: 10.1007/BF02184151
- Kleinherenbrink FAM, Ikegami I, Hiraishi A, et al (1993) Electron transfer in menaquinone-depleted membranes of *Heliobacterium chlorum*. *Biochim Biophys Acta - Bioenerg* 1142:69–73. doi: 10.1016/0005-2728(93)90085-T
- Kobayashi M, van de Meent EJ, Oh-oka H, et al (1992) Pigment composition of heliobacteria and green sulfur bacteria. In: Murata N (ed) *Research in Photosynthesis*, Vol. 1. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 393–396
- Kondo T, Itoh S, Matsuoka M, et al (2015) Menaquinone as the secondary electron acceptor in the Type I homodimeric photosynthetic reaction center of *Heliobacterium modesticaldum*. *J Phys Chem B* 119:8480–8489. doi: 10.1021/acs.jpcc.5b03723
- Krissinel E, Henrick K (2005) Multiple Alignment of Protein Structures in Three Dimensions. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp 67–78
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870–1874. doi: 10.1093/molbev/msw054
- Le SQ, Gascuel O (2008) An Improved General Amino Acid Replacement Matrix. *Mol Biol Evol* 25:1307–1320. doi: 10.1093/molbev/msn067
- Li Y, Lucas MG, Konovalova T, et al (2004) Mutation of the putative hydrogen-bond donor to P700 of photosystem I. *Biochemistry* 43:12634–12647. doi: 10.1021/bi036329p
- Liebl U, Mockensturm-Wilson M, Trost JT, et al (1993) Single core polypeptide in the reaction center of the photosynthetic bacterium *Heliobacillus mobilis*: structural implications and relations to other Photosystems. *Proc Natl Acad Sci U S A* 90:7124–7128.
- Lin S, Chiou HC, Blankenship RE (1995) Secondary electron transfer processes in membranes of *Heliobacillus mobilis*. *Biochemistry* 34:12761–12767.
- Lince MT, Vermaas W (1998) Association of His117 in the D2 protein of photosystem II with a chlorophyll that

- affects excitation-energy transfer efficiency to the reaction center. Eur J Biochem 256:595–602.
- Lockhart PJ, Larkum AWD, Steel MA, et al (1996) Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc Natl Acad Sci 93:1930–1934. doi: 10.1073/pnas.93.5.1930
- Martin WF, Bryant DA, Beatty JT (2017) A physiological perspective on the origin and evolution of photosynthesis. FEMS Microbiol Rev 205–231. doi: 10.1093/femsre/fux056
- McConnell MD, Cowgill JB, Baker PL, et al (2011) Double reduction of plastoquinone to plastoquinol in Photosystem I. Biochemistry 50:11034–11046. doi: 10.1021/bi201131r
- Miller KR, Jacob JS, Smith U, et al (1986) *Heliobacterium chlorum*: cell organization and structure. Arch Microbiol 146:111–114. doi: 10.1007/BF00402335
- Mix LJ, Haig D, Cavanaugh CM (2005) Phylogenetic analyses of the core antenna domain: Investigating the origin of photosystem I. J Mol Evol 60:153–163. doi: 10.1007/s00239-003-0181-2
- Mix LJ, Harmer TL, Cavanaugh CM (2004) Sequence of the core antenna domain from the anoxygenic phototroph *Heliophilum fasciatum*: implications for diversity of reaction center type I. Curr Microbiol 48:438–440. doi: 10.1007/s00284-003-4221-3
- Miyamoto R, Iwaki M, Mino H, et al (2006) ESR Signal of the Iron–Sulfur Center F X and Its Function in the Homodimeric Reaction Center of *Heliobacterium modesticaldum*. Biochemistry 45:6306–6316. doi: 10.1021/bi0519710
- Miyamoto R, Mino H, Kondo T, et al (2008) An electron spin-polarized signal of the P800+ A1(Q)- state in the homodimeric reaction center core complex of *Heliobacterium modesticaldum*. Biochemistry 47:4386–4393. doi: 10.1021/bi701612v
- Muhiuddin IP, Heathcote P, Carter S, et al (2001) Evidence from time resolved studies of the P700+ /A1- radical pair for photosynthetic electron transfer on both the PsaA and PsaB branches of the photosystem I reaction centre. FEBS Lett 503:56–60. doi: 10.1016/S0014-5793(01)02696-5
- Müller MG, Niklas J, Lubitz W, Holzwarth AR (2003) Ultrafast Transient Absorption Studies on Photosystem I Reaction Centers from *Chlamydomonas reinhardtii*. 1. A New Interpretation of the Energy Trapping and Early Electron Transfer Steps in Photosystem I. Biophys J 85:3899–3922. doi: 10.1016/S0006-3495(03)74804-8
- Müller MG, Slavov C, Luthra R, et al (2010) Independent initiation of primary electron transfer in the two branches of the photosystem I reaction center. Proc Natl Acad Sci 107:4123–4128. doi: 10.1073/pnas.0905407107
- Nelson N, Ben-Shem A (2004) The complex architecture of oxygenic photosynthesis. Nat Rev Mol Cell Biol 5:971–982. doi: 10.1038/nrm1525
- Nelson N, Yocum CF (2006) Structure and Function of Photosystems I and II. Annu Rev Plant Biol 57:521–565. doi: 10.1146/annurev.arplant.57.032905.105350
- Nitschke W, Mattioli T, Rutherford AW (1996) The Fe-S type photosystems and the evolution of photosynthetic reaction centers. In: Baltscheffsky H (ed) Origin and Evolution of Biological Energy Conservation. VCH, New York, pp 177–203
- Nitschke W, Rutherford AW (1991) Photosynthetic Reaction Centers - Variations On A Common Structural Theme. Trends Biochem Sci 16:241–245.
- Oh-oka H (2007) Type I Reaction Center of Photosynthetic Heliobacteria. Photochem Photobiol 83:177–186. doi: 10.1562/2006-03-29-IR-860

- 1 Ohashi S, Iemura T, Okada N, et al (2010) An overview on chlorophylls and quinones in the photosystem I-type
2 reaction centers. *Photosynth Res* 104:305–319. doi: 10.1007/s11120-010-9530-3
- 3 Okamura MY, Feher G (1992) Proton transfer in reaction centers from photosynthetic bacteria. *Annu Rev Biochem*
4 61:861–96. doi: 10.1146/annurev.bi.61.070192.004241
- 5 Okamura MY, Paddock ML, Graige MS, Feher G (2000) Proton and electron transfer in bacterial reaction centers.
6 *Biochim Biophys Acta* 1458:148–163. doi: 10.1016/S0005-2728(00)00065-7
- 7 Olson JM (1981) Evolution of photosynthetic reaction centers. *Biosystems* 14:89–94.
- 8 Olson JM, Blankenship RE (2004) Thinking about the evolution of photosynthesis. In: Govindjee, Beatty JT, Gest
9 H, Allen JF (eds) *Discoveries in Photosynthesis. Advances in Photosynthesis and Respiration*, vol 20.
10 Springer, Dordrecht, pp 1073–1086
- 11 Pan J, Saer RG, Lin S, et al (2016) Electron Transfer in Bacterial Reaction Centers with the Photoactive
12 Bacteriopheophytin Replaced by a Bacteriochlorophyll through Coordinating Ligand Substitution.
13 *Biochemistry* 55:4909–4918. doi: 10.1021/acs.biochem.6b00317
- 14 Parson WW, Chu ZT, Warshel A (1990) Electrostatic control of charge separation in bacterial photosynthesis. *BBA*
15 - *Bioenerg* 1017:251–272. doi: 10.1016/0005-2728(90)90192-7
- 16 Pei J, Kim B-H, Grishin N V (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments.
17 *Nucleic Acids Res* 36:2295–2300. doi: 10.1093/nar/gkn072
- 18 Poluektov OG, Paschenko S V, Utschig LM, et al (2005) Bidirectional electron transfer in photosystem I: Direct
19 evidence from high-frequency time-resolved EPR spectroscopy. *J Am Chem Soc* 127:11910–11911. doi:
20 10.1021/ja053315t
- 21 Pribil M, Labs M, Leister D (2014) Structure and dynamics of thylakoids in land plants. *J. Exp. Bot.* 65:1955–1972.
- 22 Prince RC, Gest H, Blankenship RE (1985) Thermodynamic properties of the photochemical reaction center of
23 *Heliobacterium chlorum*. *BBA - Bioenerg* 810:377–384. doi: 10.1016/0005-2728(85)90224-5
- 24 Ragsdale SW, Pierce E (2008) Acetogenesis and the Wood-Ljungdahl pathway of CO₂ fixation. *Biochim Biophys*
25 *Acta - Proteins Proteomics* 1784:1873–1898. doi: 10.1016/j.bbapap.2008.08.012
- 26 Rémy H-W, Stahlberg H, Fotiadis D, et al (1999) The reaction center complex from the green sulfur bacterium
27 *Chlorobium tepidum*: a structural analysis by scanning transmission electron microscopy. *J Mol Biol* 290:851–
28 858. doi: 10.1006/jmbi.1999.2925
- 29 Romberger SP, Golbeck JH (2012) The FX iron-sulfur cluster serves as the terminal bound electron acceptor in
30 heliobacterial reaction centers. *Photosynth Res* 111:285–290. doi: 10.1007/s11120-012-9723-z
- 31 Rost B (1999) Twilight zone of protein sequence alignments. *Protein Eng Des Sel* 12:85–94. doi:
32 10.1093/protein/12.2.85
- 33 Rutherford AW, Mattiolo TA, Nitschke W (1996) The FeS-type photosystems and the evolution of photosynthetic
34 reaction centers. In: Baltscheffsky H (ed) *Origin and evolution of biological energy conversion*. New York, pp
35 177–204
- 36 Rutherford AW, Osyczka A, Rappaport F (2012) Back-reactions, short-circuits, leaks and other energy wasteful
37 reactions in biological electron transfer: Redox tuning to survive life in O₂. *FEBS Lett.* 586:603–616.
- 38 Sadekar S, Raymond J, Blankenship RE (2006) Conservation of Distantly Related Membrane Proteins:

- 1 Photosynthetic Reaction Centers Share a Common Structural Core. *Mol Biol Evol* 23:2001–2007. doi:
2 10.1093/molbev/msl079
- 3 Santabarbara S, Bullock B, Rappaport F, Redding KE (2015) Controlling Electron Transfer between the Two
4 Cofactor Chains of Photosystem I by the Redox State of One of Their Components. *Biophys J* 108:1537–
5 1547. doi: 10.1016/j.bpj.2015.01.009
- 6 Santabarbara S, Heathcote P, Evans MCW (2005) Modelling of the electron transfer reactions in Photosystem I by
7 electron tunnelling theory: The phylloquinones bound to the PsaA and the PsaB reaction centre subunits of PS
8 I are almost isoenergetic to the iron-sulfur cluster FX. *Biochim. Biophys. Acta - Bioenerg.* 1708:283–310.
- 9 Sarrou I, Khan Z, Cowgill JB, et al (2012) Purification of the photosynthetic reaction center from *Heliobacterium*
10 *modesticaldum*. *Photosynth Res* 111:291–302. doi: 10.1007/s11120-012-9726-9
- 11 Sattley WM, Madigan MT, Swingley WD, et al (2008) The genome of *Heliobacterium modesticaldum*, a
12 phototrophic representative of the Firmicutes containing the simplest photosynthetic apparatus. *J Bacteriol*
13 190:4687–4696. doi: 10.1128/JB.00299-08
- 14 Sattley WM, Swingley WD (2013) Properties and Evolutionary Implications of the Heliobacterial Genome. In:
15 Beatty JT (ed) *Advances in Botanical Research Vol. 66: Genome Evolution of Photosynthetic Bacteria*.
16 Academic Press Inc., pp 67–97
- 17 Schelvis JPM, van Noort PI, Aartsma TJ, van Gorkom HJ (1994) Energy transfer, charge separation and pigment
18 arrangement in the reaction center of Photosystem II. *Biochim Biophys Acta - Bioenerg* 1184:242–250. doi:
19 10.1016/0005-2728(94)90229-1
- 20 Schubert W-D, Klukas O, Saenger W, et al (1998) A common ancestor for oxygenic and anoxygenic photosynthetic
21 systems. *J Mol Biol* 280:297–314. doi: 10.1006/jmbi.1998.1824
- 22 Schweitzer RH, Brudvig GW (1997) Fluorescence quenching by chlorophyll cations in photosystem II.
23 *Biochemistry* 36:11351–11359. doi: 10.1021/bi9709203
- 24 Schweitzer RH, Melkozernov AN, Blankenship RE, Brudvig GW (1998) Time-resolved fluorescence measurements
25 of photosystem II: the effect of quenching by oxidized chlorophyll Z. *J Phys Chem B* 102:8320–8326. doi:
26 10.1021/jp982098y
- 27 Shi T, Bibby TS, Jiang L, et al (2005) Protein interactions limit the rate of evolution of photosynthetic genes in
28 cyanobacteria. *Mol Biol Evol* 22:2179–89. doi: 10.1093/molbev/msi216
- 29 Sievers F, Wilm A, Dineen D, et al (2011) Fast, scalable generation of high quality protein multiple sequence
30 alignments using Clustal Omega.
- 31 Sneath PHA, Sokal RR (1973) *Numerical Taxonomy*. Freeman, San Francisco, CA, CA
- 32 The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45:D158–D169.
33 doi: 10.1093/nar/gkw1099
- 34 Tomo T, Okubo T, Akimoto S, et al (2007) Identification of the special pair of photosystem II in a chlorophyll d-
35 dominated cyanobacterium. *Proc Natl Acad Sci U S A* 104:7283–7288. doi: 10.1073/pnas.0701847104
- 36 Trost JT, Blankenship RE (1989) Isolation of a photoactive photosynthetic reaction center-core antenna complex
37 from *Heliobacillus mobilis*. *Biochemistry* 28:9898–9904.
- 38 Trost JT, Brune DC, Blankenship RE (1992) Protein sequences and redox titrations indicate that the electron
39 acceptors in reaction centers from heliobacteria are similar to Photosystem I. *Photosynth Res* 32:11–22. doi:

10.1007/BF00028794

- van de Meent EJ, Kobayashi M, Erkelens C, et al (1991) Identification of 81-hydroxychlorophyll a as a functional reaction center pigment in heliobacteria. *BBA - Bioenerg* 1058:356–362. doi: 10.1016/S0005-2728(05)80131-8
- van der Est A, Hager-Braun C, Leibl W, et al (1998) Transient electron paramagnetic resonance spectroscopy on green-sulfur bacteria and heliobacteria at two microwave frequencies. *Biochim Biophys Acta - Bioenerg* 1409:87–98. doi: 10.1016/S0005-2728(98)00152-2
- Vassiliev IR, Antonkine ML, Golbeck JH (2001) Iron-sulfur clusters in type I reaction centers. *Biochim Biophys Acta - Bioenerg* 1507:139–160. doi: 10.1016/S0005-2728(01)00197-9
- Vermaas WFJ (1994) Evolution of heliobacteria: Implications for photosynthetic reaction center complexes. *Photosynth Res* 41:285–294. doi: 10.1007/BF02184169
- Warren P V, Golbeck JH, Warden JT (1993) Charge recombination between P700⁺ and A1⁻ occurs directly to the ground state of P700 in a Photosystem I core devoid of FX, FB, and FA. *Biochemistry* 32:849–857. doi: 10.1021/bi00054a016
- Warshel A, Chu ZT, Parson WW (1994) On the energetics of the primary electron-transfer process in bacterial reaction centers. *J Photochem Photobiol A Chem* 82:123–128. doi: 10.1016/1010-6030(94)02010-8
- Webber AN, Lubitz W (2001) P700: the primary electron donor of photosystem I. *Biochim Biophys Acta - Bioenerg* 1507:61–79. doi: 10.1016/S0005-2728(01)00198-0
- Whelan S, Goldman N (2001) A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol Biol Evol* 18:691–699. doi: 10.1093/oxfordjournals.molbev.a003851
- Xiong J, Inoue K, Bauer CE (1998) Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*. *Proc Natl Acad Sci* 95:14851–14856. doi: 10.1073/pnas.95.25.14851
- Zeng Y, Feng F, Medová H, et al (2014) Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc Natl Acad Sci U S A* 111:7795–7800. doi: 10.1073/pnas.1400295111
- Zouni A, Witt H-T, Kern J, et al (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* 409:739–743. doi: 10.1038/35055589
- (2014) The PyMOL Molecular Graphics System, Version 1.8. Schrödinger LLC <http://www.pymol.org>. doi: 10.1038/hr.2014.17