doi:10.1016/SS0022-2386(02)00822-7 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 322, 1025–1037

JMB

A Cytochrome *b* Origin of Photosynthetic Reaction Centers: an Evolutionary Link between Respiration and Photosynthesis

Jin Xiong[1] and Carl E. Bauer[2]*

[1]*Department of Biology, Texas A&M University, College Station, TX 77843, USA*

[2]*Department of Biology Indiana University, 1001 E Third Street, Jordan Hall 142 Bloomington, IN 47405, USA*

**Corresponding author*

The evolutionary origin of photosynthetic reaction centers has long remained elusive. Here, we use sequence and structural analysis to demonstrate an evolutionary link between the cytochrome *b* subunit of the cytochrome bc_1 complex and the core polypeptides of the photosynthetic bacterial reaction center. In particular, we have identified an area of significant sequence similarity between a three contiguous membrane-spanning domain of cytochrome *b*, which contains binding sites for two hemes, and a three contiguous membrane-spanning domain in the photosynthetic reaction center core subunits, which contains binding sites for cofactors such as (bacterio)chlorophylls, (bacterio)pheophytin and a non-heme iron. Three of the four heme ligands in cytochrome *b* are found to be conserved with the cofactor ligands in the reaction center polypeptides. Since cytochrome *b* and reaction center polypeptides both bind tetrapyrroles and quinones for electron transfer, the observed sequence, functional and structural similarities can best be explained with the assumption of a common evolutionary origin. Statistical analysis further supports a distant but significant homologous relationship. On the basis of previous evolutionary analyses that established a scenario that respiration evolved prior to photosynthesis, we consider it likely that cytochrome *b* is the evolutionary precursor for type II reaction center apoproteins. With a structural analysis confirming a common evolutionary origin of both type I and type II reaction centers, we further propose a novel "reaction center apoprotein early" hypothesis to account for the development of photosynthetic reaction center holoproteins.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: distant homology; ancestral sequence reconstruction; Bayesian phylogeny; structural comparison; threading

Introduction

The origin and evolution of photosynthesis are extremely complex biological issues. There are two current views concerning the origin of photosynthesis on Earth. The first suggests that photosynthesis began during the prebiotic phase on Earth and was coupled with the origin of life.[1,2] This view has been called into question, especially after the recent rebuttal of the 3.5 billion-year-old "cyanobacterial fossils".[3] The second, which is based on molecular phylogenetic analyses, suggests that photosynthesis appeared sometime after the origin of life and that it arose in Bacteria after divergence of Bacteria from Archaea.[4,5] This second view has become more broadly accepted.

Since the photochemical reaction is performed by photosynthetic reaction centers, it is reasonable to believe that the evolution of photosynthesis follows the evolution of reaction centers.[6–8] Reaction centers are classified as either type I or type II on the basis of their unique photochemical structures and functions. Cyanobacteria and chloroplasts contain both types of reaction centers that perform a linear electron transfer, with the type II reaction center oxidizing water to produce oxygen. Anoxygenic photosynthetic bacteria, namely, purple bacteria, green sulfur bacteria, and heliobacteria, use

Abbreviations used: LHII-β, light harvesting II-β; RMSD, root-mean-square; JTT, Jones, Taylor and Thornton amino acid substitution matrix; BI, Bayesian inference; MCMCMC, Metropolis-coupled Markov chain Monte Carlo.

E-mail address of the corresponding author: cbauer@bio.indiana.edu

0022-2836/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved

one reaction center type (either type I or type II) and perform a cyclic electron transfer through the cytochrome bc/bc_1 complex.

All photosynthetic reaction centers have two major components, membrane-spanning apopolypeptides and tetrapyrrole photopigments embedded in the apopolypeptides. Crystal structural analysis has demonstrated that both type I and type II reaction centers exhibit significant conservation at the tertiary structural level with respect to the topology of transmembrane helices as well as positions of the Mg–tetrapyrrole photopigments.[9–11] This has been interpreted as evidence that these two reaction centers most likely evolved from a common origin. However, despite significant structural conservation, there remains almost no sequence conservation between the type I and type II reaction center apopolypeptides. Consequently, it has not been possible to construct a unified reaction center apopolypeptide phylogeny that includes all photosynthetic lineages. This has hampered efforts to reconstruct the evolutionary history of photosynthesis.

Recently, we circumvented the problem by determining evolutionary relationships of all photosynthetic lineages using genes and enzymes that are involved in the Mg–tetrapyrrole biosynthetic pathway. Using a rooted phylogenetic approach, we observed that the purple bacterial lineage is most likely to be the first lineage to evolve photosynthetic pigments. This is followed by the green sulfur bacterial lineage that subsequently bifurcated into green non-sulfur and cyanobacterial lineages, with the heliobacterial lineage being the latest to receive pigment biosynthesis genes.[8,12] What is unclear is whether the evolutionary path of the reaction center apoproteins is congruent with that of the tetrapyrrole biosynthetic enzymes.

Although the evolutionary path giving rise to type I and type II reaction center apoproteins remains an enigma, there has been some effort to reveal the ancestral origin of reaction center apoproteins. For example, Mulkidjanian & Junge[13] suggested that a putative 11-transmembrane helix protein providing a UV protection role may be an ancestor for the reaction centers. Because of the unclear nature of the proposed UV-protector, the hypothesis has remained unsupported. Meyer[14] suggested that photosynthetic reaction centers may have originated from the cytochrome *b* subunit of the cytochrome bc_1 complex. This proposal was based on the observation that both types of proteins are membrane-spanning with bound tetrapyrroles and quinones and function in electron transfers. However, since this hypothesis lacked supporting evidence on the basis of sequence analysis, it has not been broadly accepted. Vermaas,[15] Olson[16] and Baymann *et al.*[17] speculated that the last common ancestor of the reaction center should resemble a current type I reaction center. However, the nature of a protein ancestor preceding the type I photosynthetic reaction center remained unanswered. Here, we present the first evidence for the origin of photosynthetic reaction centers, which is supported by sequence and structural analysis. We conclude that a pre-existing cytochrome *b*-like transmembrane protein may be the evolutionary precursor of the photosynthetic reaction centers.

Results

Identification of similarity between cytochrome *b* and photosynthetic reaction centers

Sequence homology between cytochrome *b* of the bc_1 complex and photosynthetic reaction centers was discovered fortuitously when performing a gapped BLAST analysis with several type II reaction center polypeptides. For instance, when using *Rhodobacter capsulatus* PufM as query, the BLAST search generated a large number of medium to low-scored hits with short sequence alignments between the reaction center polypeptide and cytochrome *b* sequences from purple bacteria or mitochondria. The *E*-values were not the best hits after the reaction center hits. However, the sheer number of hits (303 out of 399 non-reaction center hits with *E*-values ranging from 1.5 to 9.6) warranted further investigation. A profile-based iterative PSI-BLAST[18] search was carried out, which generated more statistically significant hits. When using full-length PufL from *Chloroflexus aurantiacus* as query, the search generated a significant match with mitochondrial cytochrome *b* (*E*-value 8×10^{-12}) of *Felis catus*. The alignment contained 254 positions with an identity of 13.5% and similarity 31.3% (alignment not shown).

More refined alignment between cytochrome *b* and type II reaction center polypeptide sequences was carried out using CLUSTAL X[19] and DIALIGN[20] followed by manual adjustment. This analysis revealed a region of significant similarity, which includes three contiguous transmembrane helices of cytochrome *b* that bind two hemes (helices B–D) and three contiguous transmembrane helices of type II reaction centers (helices B–D) that bind cofactors such as (bacterio)chlorophylls, (bacterio)pheophytin, quinone and a non-heme iron. Figure 1 shows a refined alignment of the purple bacterial cytochrome *b* subunit of the bc_1 complex with the L subunit of anoxygenic bacterial reaction centers and the D1 subunit of oxygenic cyanobacterial reaction centers. The aligned region includes 44–48% of the full-length reaction center protein sequences. The average pairwise sequence identity of the aligned segments of cytochrome *b* with reaction center polypeptides is 21.5%, and average sequence similarity is 49.4% (Figure 1, Table 1). In particular, the pairwise comparison between the L subunit and cytochrome *b* shows that their sequence identity can be as high as 33.1% (Table 1).

The pairwise sequence similarity for sequences presented in Figure 1 was further assessed for

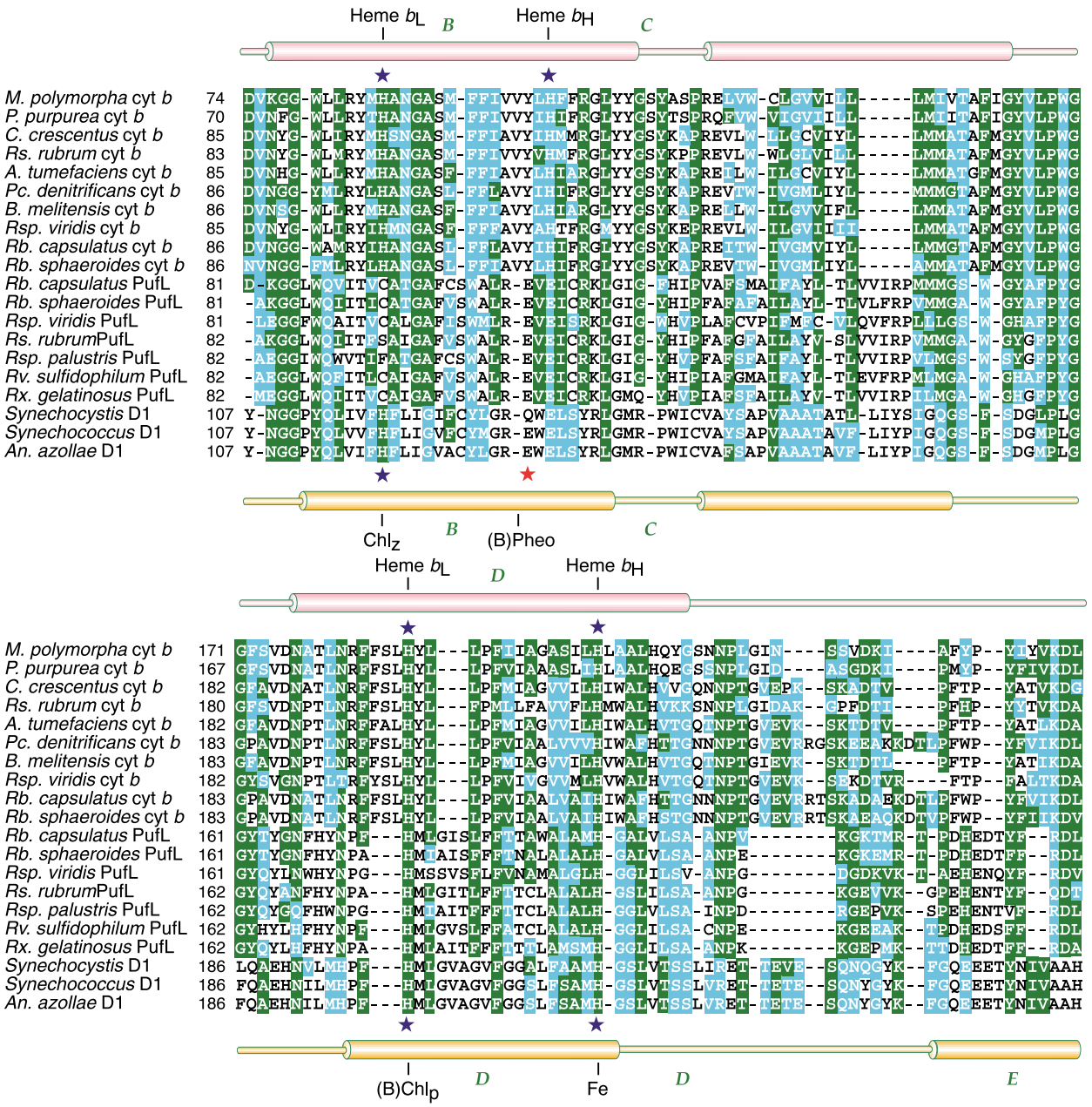

Figure 1. Sequence comparison of cytochrome *b* with type II reaction center core polypeptides. Multiple amino acid sequence alignment of purple bacterial cytochrome (cyt) *b* with type II reaction center core polypeptides from purple bacteria and cyanobacteria is shown with identical and similar residues highlighted by green and cyan, respectively. The alignment was done using CLUSTAL X, DIALIGN 2 followed by manual refinement based on the BLOSUM62 amino acid substitution matrix. The highlighting was also on the basis of the BLOSUM62 matrix. Sequence comparison was made only between residues of different subfamilies, as similarities within the same subfamily are not a concern of this study. Thus, similar residues within the same subfamily are not highlighted. For instance, at position 3, there are K in some PufL sequences and N in some D1 sequences which match identical K and N present in the cytochrome *b* sequences (green), whereas E present in some other PufL sequences matches N in cytochrome *b* as a similar residue (cyan). Conserved transmembrane helices B–D of cytochrome *b* and of the type II reaction centers are shown by cylinders above and below the sequence alignment, respectively. Three of the four heme ligands are aligned with ligands for accessory chlorophyll$_z$ (Chl_z), special pair (bacterio)chlorophyll ((B)Chlp), and the non-heme iron (Fe) of type II reaction center proteins, which are denoted by blue stars and labeled on helices. One of the heme ligands corresponds to a residue in the reaction centers, which is two residues downstream from the (bacterio)pheophytin ((B)Pheo) binding residue in the type II reaction centers (indicated by a red star on helix B). Abbreviations for genus names: *A.*, *Agrobacterium*; *An.*, *Anabaena*; *B.*, *Brucella*; *C.*, *Caulobacter*; *M.*, *Marchantia*; *P.*, *Porphyra*, *Pc.*, *Paracoccus*; *Rb.*, *Rhodobacter*; *Rsp.*, *Rhodopseudomonas*; *Rv*, *Rhodovulum*; *Rx.*, *Rubrivivax*. Certain poorly aligned residues between helices C and D for both cytochrome *b* and reaction center apopolypeptides were omitted.

Table 1. Pairwise sequence identities (I) and similarities (S) for the sequence alignment shown in Figure 1

	Rb. sphaeroides cyt *b*			*Rb. capsulatus* cyt *b*			*Rsp. viridis* cyt *b*			*B. melitensis* cyt *b*			*Pc. denitrificans* cyt *b*		
	I (%)	S (%)	P	I (%)	S (%)	P	I (%)	S (%)	P	I (%)	S (%)	P	I (%)	S (%)	P
Rb. capsulatus PufL	33.1	58.6	8.61×10^{-4}	32.3	57.1	7.57×10^{-6}	26.2	55.4	8.73×10^{-5}	28.5	56.9	1.11×10^{-4}	33.0	55.6	2.26×10^{-4}
Rb. sphaeroides PufL	29.3	59.4	2.67×10^{-3}	31.6	58.6	6.51×10^{-5}	23.1	56.9	5.14×10^{-4}	27.6	56.9	4.20×10^{-4}	27.8	57.2	1.21×10^{-3}
Rsp. viridis PufL	24.1	52.6	0.035	25.6	53.4	0.014	18.5	53.8	0.125	21.5	53.8	0.082	27.8	50.0	0.207
Rs. rubrum PufL	26.3	54.1	0.010	30.0	54.1	1.37×10^{-3}	24.6	50.0	0.012	28.5	54.6	3.74×10^{-3}	27.8	51.1	4.99×10^{-3}
Rsp. palustris PufL	23.3	50.3	0.028	25.6	51.9	3.32×10^{-3}	21.5	50.8	0.048	20.0	50.7	0.011	24.8	48.1	0.032
Rv. sulfidohilum PufL	27.1	54.9	0.011	26.1	52.6	1.94×10^{-3}	20.8	48.5	0.017	23.8	54.6	4.53×10^{-3}	29.3	55.6	9.45×10^{-3}
Rx. gelatinosus PufL	21.8	51.9	0.229	24.1	48.9	0.013	20.0	49.2	0.090	22.3	51.5	0.064	22.5	50.3	0.219
Synechocystis D1	18.2	49.6	0.061	18.2	48.9	0.085	14.2	41.0	0.086	18.7	44.8	0.404	20.4	49.6	0.172
Synechococcus D1	16.1	48.1	0.015	15.3	47.4	0.337	11.9	45.5	0.013	15.7	41.8	0.044	17.5	44.5	0.362
An. azollae D1	16.8	42.0	0.038	16.8	43.8	0.581	11.9	42.5	0.014	14.9	43.3	0.049	16.0	45.2	0.560

	A. tumefaciens cyt *b*			*Rs. rubrum* cyt *b*			*C. crescentus* cyt *b*			*P. purpurea* cyt *b*			*M. polymorpha* cyt *b*		
	I (%)	S (%)	P	I (%)	S (%)	P	I (%)	S (%)	P	I (%)	S (%)	P	I (%)	S (%)	P
Rb. capsulatus PufL	27.7	56.2	6.41×10^{-5}	23.8	53.0	2.16×10^{-5}	26.9	53.1	5.15×10^{-4}	29.5	55.8	1.85×10^{-5}	25.6	53.5	3.44×10^{-5}
Rb. sphaeroides PufL	26.1	55.4	9.49×10^{-4}	25.3	51.5	5.52×10^{-5}	25.4	54.6	8.21×10^{-4}	26.4	55.0	1.19×10^{-4}	25.6	53.4	3.70×10^{-4}
Rsp. viridis PufL	23.0	50.8	0.036	22.3	50.8	0.089	22.3	47.8	0.123	24.0	52.3	4.90×10^{-3}	24.0	49.6	7.38×10^{-3}
Rs. rubrum PufL	26.9	52.3	5.44×10^{-3}	26.2	50.0	9.29×10^{-5}	29.2	50.0	5.75×10^{-3}	27.1	51.9	7.23×10^{-3}	27.1	48.8	2.16×10^{-3}
Rsp. palustris PufL	22.3	50.0	0.023	21.5	47.7	0.014	21.5	48.5	0.026	20.9	51.2	2.77×10^{-3}	19.4	49.6	0.017
Rv. sulfidohilum PufL	23.8	48.5	2.85×10^{-3}	22.3	51.5	3.13×10^{-4}	22.3	50.8	0.012	24.0	51.9	1.83×10^{-3}	23.3	48.8	0.016
Rx. gelatinosus PufL	23.8	49.2	0.097	21.5	50.0	7.51×10^{-3}	20.8	48.5	0.022	21.7	51.9	9.03×10^{-3}	20.9	49.6	0.084
Synechocystis D1	19.4	43.2	0.081	12.7	38.0	0.272	15.6	39.6	0.100	16.5	43.6	0.054	15.0	42.1	0.394
Synechococcus D1	15.6	41.8	8.79×10^{-3}	11.2	36.6	0.383	14.6	37.3	5.39×10^{-3}	15.0	45.9	0.133	13.5	43.6	0.072
An. azollae D1	13.4	41.8	0.017	10.4	34.3	0.307	14.6	36.6	0.047	17.1	36.1	0.108	12.8	33.8	0.081

The identity and similarity values were calculated according the formulas: $I\% = [(L_i \times 2)/(L_a + L_b)]$ and $S\% = [(L_s \times 2)/(L_a + L_b)]$. Statistical significance (P-values) for the pairwise sequence comparison was analyzed by an alignment-independent PRSS test.[21]

statistical significance using the program PRSS.[21] Statistical support (P values) for each pairwise comparison is shown in Table 1. The entropy-based analysis without influence from our alignment shows that 70% of the pairwise comparisons have P values smaller than 0.05. In particular, the purple bacterial reaction center L subunit shows more significant similarity with cytochrome b with 57% of the pairwise comparisons having P smaller than 0.01, with a P value for maximum similarity being 7.57×10^{-6}. These values provide decent support for a homologous relationship between the reaction center L subunit and cytochrome b. It needs to be pointed out that the above statistical analysis was carried out for the partial sequence regions of both protein subfamilies (including helices B–D). Although the evidence presented here may be sufficient to establish a distant or remote homology relationship between the two types of membrane proteins, it is not sufficient to establish a close homology relationship because the full-length sequence alignment with high similarity cannot be reliably obtained.

Detailed inspection of the alignment revealed matching of key cofactor ligands (Figure 1) in which the first His residue of helix D of cytochrome b that forms a ligand to heme b_L matches the first His residue in helix D of type II reaction center polypeptides. This His residue, which is conserved in all type II reaction center polypeptides, forms a ligand to the special pair (bacterio)-chlorophylls that are responsible for the primary charge separation. A second fully conserved His residue near the end of helix D of cytochrome b, which forms a ligand to heme b_H, matches a non-heme iron His ligand in the reaction center polypeptides. This His ligand in the reaction center polypeptides plays an important role in electron transfer between quinones Q_A and Q_B. Indeed, if the reaction center and cytochrome b evolved from a common ancestor, the non-heme iron can be perceived as a remnant of heme b_H.

In helix B of cytochrome b, the first His residue which forms a ligand to heme b_L matches with the first His in helix B of the oxygenic reaction center D1 and D2 polypeptides only. In D1 or D2, this His provides a ligand to a unique distal accessory chlorophyll (chlorophyll$_z$) that may play an exciton-connecting role to mediate energy transfer between CP43/CP47 and the reaction center core pigments.[11] The second His ligand to heme b_H in helix B of cytochrome b is not conserved in the reaction center polypeptides. The corresponding residue in the reaction center is a Glu which is located two residues downstream from the Glu/Gln that is used for (bacterio)pheophytin binding (indicated by a red star, Figure 1). Overall, there is a matching of most of the crucial cofactor ligands, which, along with the PRSS analysis and overall sequence similarity (49.4%), strengthens the hypothesis that cytochrome b and photosynthetic reaction centers share a common evolutionary origin.

To probe structural relationships, we compared three-dimensional structures of the cofactor-binding transmembrane helices of cytochrome b and corresponding regions of the type II reaction center. In this comparison, we used the crystal structure of cytochrome b of the mitochondrial bc_1 complex[22] with the assumption that this structure is largely similar to that of the purple bacterial bc_1 complex due to the high level of sequence identity. We also examined the crystal structure of the L and M subunit of the purple bacterial reaction center.[23] Despite the conservation of key cofactor binding residues between cytochrome b and the reaction center proteins, the overall membrane helical arrangement outside helices B–D is actually not notable. The heme binding domain of cytochrome b is a four-helix bundle with the respective helices B, C, and D being adjacent to each other, whereas in the reaction center structures, the helix E is positioned between helices C and D. In addition, the spacing between the two His ligands in the reaction center polypeptides is three residues longer than that in cytochrome b (16 residues instead of 13). This accounts for roughly an extra turn in the helix. However, the overall helical structure within helices B–D of the two groups of proteins remains roughly similar (Figure 2), though the reaction center helices may have "opened up" with the space between the His ligands elongated during the course of evolution to accommodate the placement of large chromophores such as bacteriopheophytins and quinones in performing the new function, the photosynthetic electron transfer.

Somewhat less intuitive but also noticeable is the fact that, in cytochrome b, the hemes are embedded in a four-helix bundle (helices A–D), whereas the corresponding cofactors in the reaction center, the non-heme iron and the special pair (bacterio)chlorophylls, are also surrounded by four helices composed of helices D and E from both L and M subunits, which together resemble a tilted helix bundle (see Supplementary Material). The resemblance is striking, though the quaternary structural similarity does not correlate with the sequence similarity.

Phylogenetic analysis of cytochrome *b* and photosynthetic reaction centers

To seek phylogenetic evidence on the relationship of cytochrome b and photosynthetic reaction center proteins, we constructed a Bayesian phylogenetic tree using the refined multiple sequence alignment of the core regions of cytochrome b with type II reaction center apoproteins shown in Figure 1. The maximum likelihood-based Bayesian method results in an extremely robust phylogeny for each single node with the reaction center polypeptides forming a monophyletic group to the exclusion of cytochrome b sequences (Figure 3(a)). The Bayesian method has an advantage over traditional likelihood analysis, in that it is able to handle large phylogeny and able to consider all

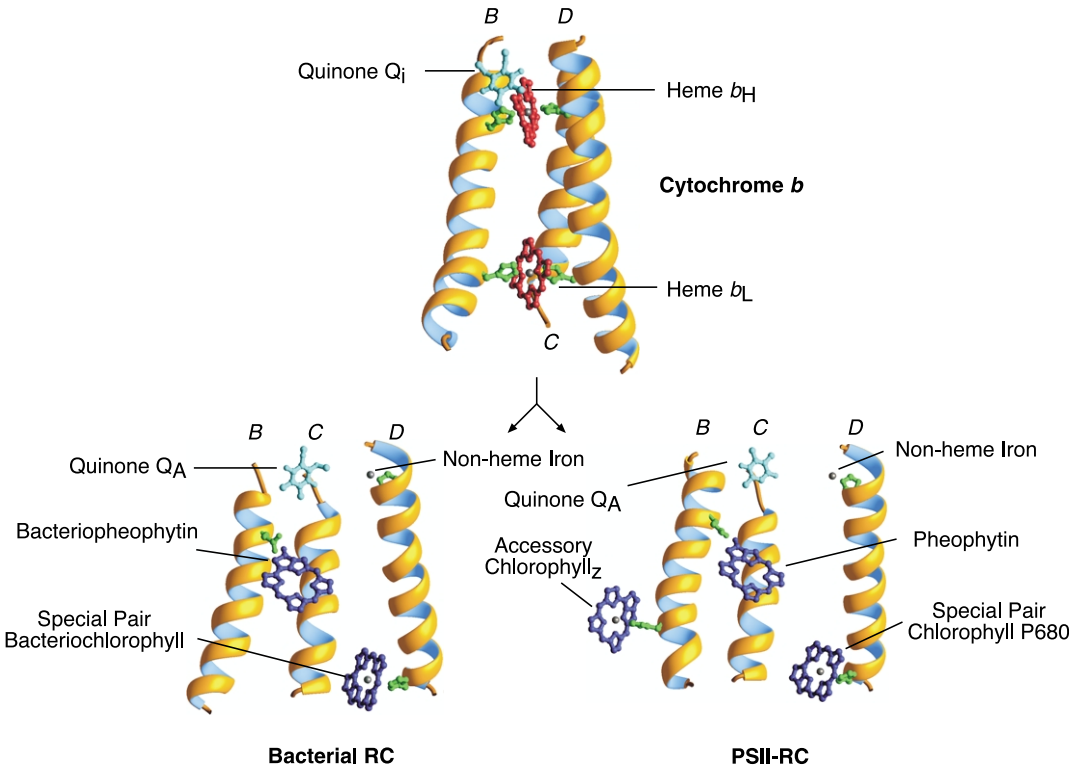

Figure 2. Comparison of the cofactor-binding core structures (helices B–D) of cytochrome *b* (1BCC, from chicken mitochondria[22]) and of the type II reaction centers (2RCR, from *Rb. sphaeroides*[23] and a modeled structure by Xiong *et al.*[44]). For clarity, the side-chains of the porphyrin molecules are omitted. All structures were rendered using the RIBBONS program.[50]

potential trees by sampling using a Metropolis-coupled Markov chain Monte Carlo (MCMCMC) approach.[24] This method recovers a tree with a maximum posterior probability with resulting Bayesian probability values representing true credibility of each clade. In addition, it uses explicit probabilistic models of sequence evolution and corrects multiple substitutions at individual sites, thus essentially eliminating the long branch attraction effect.[24] Consequently, the excessive long branch for the D1 clade (Figure 3(a)) can be considered correctly rooted and truly proportional to the divergence of the sequences.

In the phylogenetic tree of Figure 3(a), the monophyletic separation of cytochrome *b* and reaction center sequences cannot be automatically interpreted as a result of ancient gene duplication because even non-homologous sequences can be recovered as monophyly. What clearly defines homology in this case between the two groups of sequences is the statistical test shown in Table 1. The fact that most of the sequences used in the analysis have identity levels above 15% (Table 1) suggests that the data are well within the divergence range that a likelihood-based method can handle.[25] Thus, with the support of sequence-based analysis, we can now establish an ancestor–descendant relationship between the two proteins.

The phylogenetic tree in Figure 3(a) is, strictly speaking, unrooted. However, the fact that the root can be determined by external knowledge, which is, in this case, the "respiration early" hypothesis (see Discussion), allowed us to tentatively root the reaction center clade with the cytochrome *b* clade. On the basis of this assumption, we further inferred an ancestral amino acid sequence for the common ancestor of photosynthetic reaction center polypeptides using the Bayesian method with the James, Taylor and Thornton (JTT) model of amino acid substitutions. Its position is located at the basal node for both anoxygenic and oxygenic reaction center proteins (indicated by a filled circle in Figure 3(a)). Not surprisingly, the inferred ancestral protein sequence (Figure 3(b), 151 amino acid residues in length) with posterior probability values indicated at each position is shown to be intermediate between cytochrome *b* and the L subunit of the purple bacterial reaction center, with conservation of three of the four His ligands in cytochrome *b*. It also has higher overall similarity to cytochrome *b* than the extant reaction center proteins (pairwise sequence identity 45.7% and similarity 65.7% with cytochrome *b* of *R. capsulatus*; the *P* value from the PRSS test is 1.70×10^{-16}).

Through a BLAST search, the inferred sequence was identified as an L subunit-like protein. In our additional phylogenetic analysis that includes L and M, D1 and D2 subunits (see Supplementary Material), the results confirmed that an L-subunit-like protein served as the last common ancester to all type II reaction-center polypeptides and that

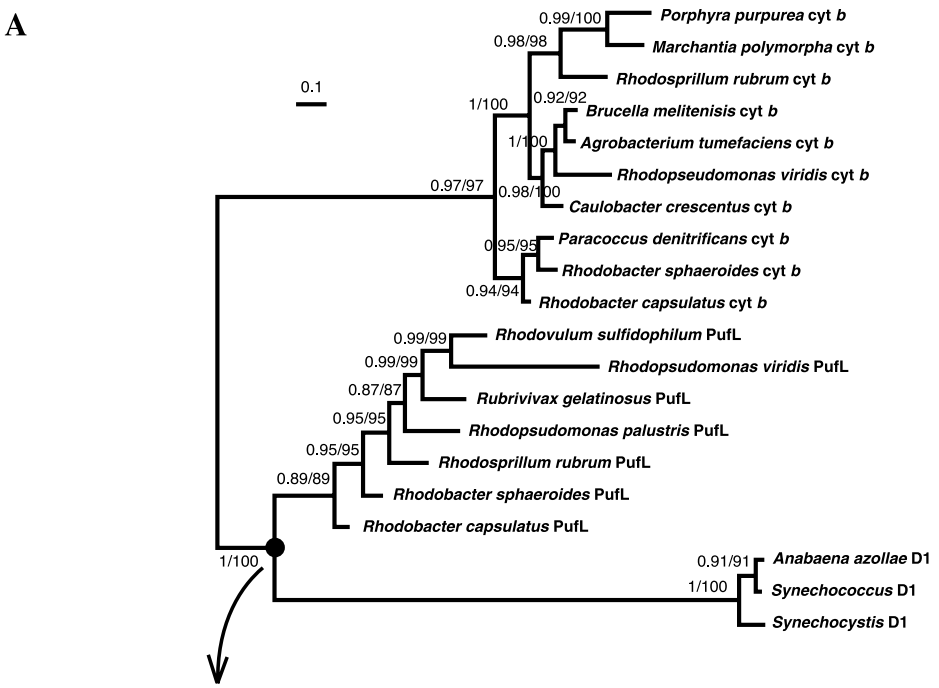

Figure 3. (a) Bayesian phylogenetic tree for cytochrome *b* (Cyt *b*) and type II reaction center apopolypeptides on the basis of amino acid alignment in Figure 1 (left). The best tree is recovered from 4000 MCMCMC sampled trees using the MRBAYES program after exclusion of the burn-in. The final tree has a maximum likelihood value $-\ln \times L = 9448.01$. The Bayesian posterior probability values (on the left of the slash) and maximum likelihood non-parametric bootstrap values (on the right of the slash) are indicated near the base of each node. Branch lengths represent mean values of the sampled trees with the scale bar corresponding to 0.1 amino acid substitution per site. The node for the type II reaction center common ancestor is indicated by a filled circle. (b) Inferred ancestral sequence of type II reaction centers using the Bayesian likelihood approach. The posterior probability of each position is indicated below the sequence.

the D1 and D2 gene duplication as well as the L and M gene duplication are results of ancient independent events.[6,26] This suggests that the extant oxygenic and anoxygenic forms of type II reaction centers may have evolved twice from a gene for a cytochrome *b*-like protein. Given the fact that there is strong tertiary structural similarity in the two different type II reaction centers,[11] it appears that there had to be strong selective pressure to maintain an overall topological organization and cofactor arrangement in order to drive a similar photochemistry.

Common evolutionary origin for type I and II reaction centers

The above analysis has provided evidence on the origin of the type II photosynthetic reaction center. The type I reaction center has been shown to share a common evolutionary origin with the type II reaction center through manual structural comparisons[9–11] despite the lack of detectable sequence similarity. However, no quantitative analysis to support such a claim has ever been made. Here, we explore the structural relationship of the two types of reaction centers with a protein structural fold recognition method that evaluates how well a protein sequence fits into a known structural fold, on the basis of a number of criteria such as sequence profiles, secondary structure, and energy functions. This process is known as threading. It often provides a good indication for close structural similarity that results from common ancestry.

In this analysis, we used a threading program GenTHREADER,[27] which uses a neural network algorithm to predict protein folds on the basis of alignments of sequence profiles with a known fold library and predicted secondary structures. As shown in Table 2, when a type I reaction center protein sequence is used as query without prior

Table 2. Structural homology between type I and type II reaction cener apoproteins detected by a threading program GenTHREADER

Query	Significant hit (PDB code)	*E*-value
PsaA (*S. elongatus*)	PufL (6PRC)	8×10^{-3}
	PufM (6PRC)	0.01
PsaB (*S. elongatus*)	PufL (6PRC)	0.01
	PufM (6PRC)	0.01
PufL (*C. aurantiacus*)	PsaA (1JB0)	0.02
PufL (*R. capsulatus*)	PsaA (1JB0)	0.01
PufM (*R. capsulatus*)	PsaA (1JB0)	0.03

The query sequences for the type I reaction center include only the reaction center domain (PsaA, residues 431–755; and PsaB, residues 412–740). For the type II reaction center proteins, full-length sequences were used in the search analysis. Only significant hits for the non-self type reaction center ($E < 0.05$) are shown.

knowledge of its actual structure, a protein structure for a type II reaction center can often be identified with a significant statistical score ($E < 0.05$). Similar fold recognition analysis for a type II reaction center would result in a significant hit for a type I reaction center (at present there is only one high-resolution type I reaction center structure available, 1JB0 from *Synechococcus elongatus*).

We also compared three-dimensional structures of both types of reaction centers using the DALI program,[28] which performs optimal protein structural alignment based on graph theory that maximizes the extent of contact patterns between two structures. The major parameters for the structural comparisons are shown in Table 3. Structure pairs with *Z*-scores in the range of 2 to 8 are considered minimally related, whereas those between 8 and 16 indicate a medium level of structural homology.[28] Thus, the two types of reaction centers are either close to or within the medium homology level. The other three statistical parameters, root-mean-square deviation (RMSD), alignment length and local sequence identity, have empirical thresholds that represent 75% probability of identifying protein homologs.[29] The fact that two out of three parameters exceeded the thresholds gives a good indication that the characteristics shared by the two types of reaction centers may be a result of common ancestry.

Table 3. Three-dimensional structural comparison between type I and type II reaction centers using the DALI program

	Z score	RMSD	Alignment length (%)	Local identity (%)
PsaA × PufL	8.7	2.8	91.3	6.0
PsaA × PufM	7.9	2.6	92.6	7.0
PsaB × PufL	6.9	2.7	88.7	8.0
PsaB × PufM	7.3	2.6	93.5	8.0
Threshold	>2	<3.1	>50	>10

The major statistical parameters from the structural alignment of the two types of proteins are shown, along with empirical threshold levels for the two proteins to be considered as related. Only the helical regions of the reaction center domain of reaction center polypeptides were used in the structural comparison. PsaA and PsaB structures (PDB code 1JB0) were derived from *S. elongatus*; PufL and PufM (PDB code 2RCR) structures were from *R. sphaeroides*. *Z* score indicates the significance of the similarity in standard deviations above a database average. RMSD is the square root of the average squared Euclidean distance over all topologically equivalent pairs of alpha carbon positions.

This is the first time that an independent analysis with evolutionary information has verified the previous conclusion that both type I and type II reaction center proteins adopt the same structural fold, giving support for the common origin of the two types of reaction centers. The fact that nearly no sequence similarity between the two types of reaction centers is detectable means that traces of evolution have been mainly left at the tertiary structural level. If both types of reaction centers indeed evolved through a divergent pathway, a putative origin identified for the type II reaction center (in this case, cytochrome *b*), in theory, should apply equally well for the type I reaction center. On the basis of degrees of sequence similarity, the type II reaction centers are clearly closer to cytochrome *b*, and are thus considered earlier evolving than the type I reaction centers. An evolutionary pathway involving both type I and type II reaction centers has recently been proposed.[8]

Discussion

An "apoprotein early model" for the evolution of photosynthetic reaction centers

Our recent comprehensive analysis of both the reaction center polypeptides and Mg–tetrapyrrole biosynthesis genes[8] has clearly indicated that evolutionary pathways for reaction center apoproteins and for Mg–tetrapyrrole biosynthesis enzymes are significantly different, which indicates that the two sets of genes did not coevolve. This raises the question of which set of genes, the reaction center genes or Mg–tetrapyrrole biosynthesis genes, would have evolved first.

We propose here that reaction center proteins may have evolved early and that Mg–tetrapyrrole biosynthesis enzymes may have subsequently evolved in organisms containing pre-existing reaction center-like apoproteins that may have still functioned as cytochromes. This hypothesis is based on the argument that if Mg–tetrapyrroles had arrived alone without polypeptides to bind Mg–tetrapyrroles that assist in photochemistry, then there would be a waste of cellular resources for biosynthesis and, more importantly, generation of damaging free radicals by photoexcitation, which would be highly detrimental to the cell's survival, a clearly unfavorable and unlikely evolutionary outcome.

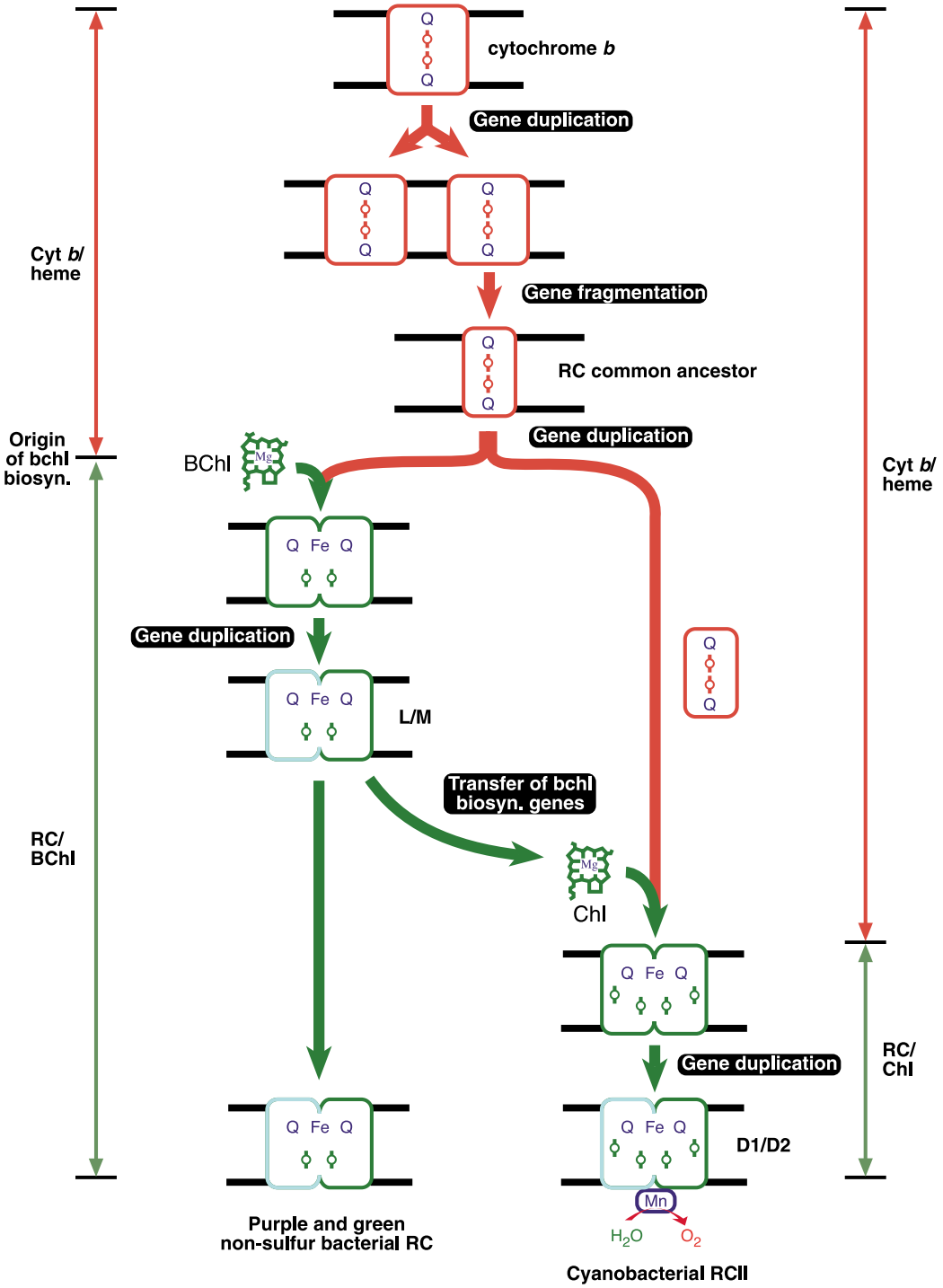

Figure 4. An aproprotein early hypothesis evolutionary pathway for type II reaction center holoproteins. By combining information derived from different evolutionary pathways for reaction center apoproteins and for Mg–tetrapyrrole biosynthesis, we propose that a gene duplication event for cytochrome *b* (with eight transmembrane helices) gave rise to an intermediate form between cytochrome *b* and a type II reaction center protein, which after a loss of three transmembrane helices became a primitive form of a reaction center apoprotein that most likely resembles the extant L subunit of type II reaction centers. The protein later diverged separately into purple or green non-sulfur bacterial lineages and into a cyanobacterial lineage. Purple bacteria, which may be the first organism to synthesize Mg–tetrapyrroles, would become the first lineage to evolve a reaction center holoprotein after incorporating Mg–tetrapyrrole into the ancestral reaction center, which may have first existed as homodimer form and later became a heterodimer after a paralogous gene duplication event. For cyanobacteria, which may have received Mg–tetrapyrrole biosynthesis genes at a later stage *via* acquisition from an anoxygenic phototrophic organism, the precursor form for D1/D2 proteins may have existed as a heme-binding cytochrome *b*-like protein for a much longer period of time before Mg–tetrapyrroles became available (reaction center apoprotein early scenario) for the protein to be converted for photosynthesis. Arrowed bars at the two sides are hypothetical time lines for the development of components of the photosynthetic reaction centers.

The "reaction center apoprotein early" hypothesis thus provides a unified view on the basis of phylogenetic information from both reaction center components. In this hypothesis (Figure 4), we suggest that an initial gene duplication event for cytochrome *b* (presumably with eight transmembrane helices) gave rise to a cytochrome that was a precursor of a type II reaction center apoprotein. After a loss of three transmembrane helices, an intermediate cytochrome with five transmembrane helices became a primitive form of a reaction center protein, which most likely resembled the extant L subunit of type II reaction centers. This protein later diverged separately into an anoxygenic purple/green non-sulfur bacterial lineage and into an oxygenic cyanobacterial lineage. As the ancestral purple bacteria may be the first to synthesize Mg–tetrapyrroles,[12,30] the incorporation of the photosynthetic pigment in this lineage would convert a cytochrome into a primitive reaction center and allow the purple bacterial lineage to become the first photosynthetic form. The reaction center may have first existed as a homodimer form and later became a heterodimer following a paralogous gene duplication event.

In the case of the oxygenic cyanobacterial lineage, we also apply the apoprotein early scenario, as this lineage may have received Mg–tetrapyrrole biosynthesis genes at a much later stage *via* a lateral gene transfer from an anoxygenic phototrophic organism.[12,30] Consequently, the precursor of the oxygenic reaction center apoproteins may have existed as a heme-binding cytochrome *b* protein for a much longer period of time due to the late arrival of Mg–tetrapyrrole biosynthesis genes. Thus, the excessive long branch for the D1 clade (Figure 3(a)) may not have been in the photosynthetic form throughout its developmental history. Instead, only a lateral transfer of Mg–tetrapyrrole biosynthesis genes at a later stage of development would have provided conditions to convert the cytochrome *b*-like protein into a Mg–tetrapyrrole-binding photosynthetic reaction center. A subsequent gene duplication would convert a homodimer oxygenic reaction center to the extant D1/D2 heterodimer form.

The above evolutionary scenario also explains an apparent paradox regarding the late-evolving cyanobacterial reaction center holoprotein having a more conserved His in helix B of D1/D2 proteins, which contrasts the lack of the first His residue in the anoxygenic bacterial reaction center (Figure 1) despite the higher overall level of sequence conservation. According to the reaction center apoprotein early hypothesis, the ancestral form of the cyanobacterial type II reaction center apoproteins would have existed as a *b*-type cytochrome for a much longer period of time before being converted to the photosynthetic form. This would provide selective pressure to conserve the first His in helix B of the apoprotein, since it would have been needed as a heme ligand. In conversion to the photosynthetic form, this His would still be preserved as a ligand to a peripheral accessory antenna pigment (chlorophyll$_z$) which serves as a conduit for excitation energy transfer from the CP43/CP47 core antenna system to the primary donor P680.[11,31] This type of core antenna system is, however, not present in the purple bacterial reaction center, which uses a different type of light-harvesting system.[32] Thus, this His is not under selective pressure to be preserved in the bacterial reaction center.

An implication for the above reaction center apoprotein early hypothesis is that the evolution of Mg–tetrapyrrole biosynthesis is the limiting or determining step for the evolution of photosynthesis. Thus, our recent phylogenetic analysis of Mg–tetrapyrrole biosynthesis genes[12] provides a better indication of the overall evolutionary history of photosynthesis, in which purple bacteria are the earliest emerging photosynthetic lineage.

Additional support for the link between respiration and photosynthesis

Numerous studies have demonstrated a ubiquitous existence of cytochrome *b* and the Rieske iron–sulfur protein, essential components of the cytochrome *bc* complex, in Archaea, Bacteria and Eukarya, indicating that many respiratory components including the cytochrome *bc* complex may have existed in the last common ancestor of Archaea and Bacteria.[33–35] In contrast, Mg–tetrapyrrole-based photosystems are found in Bacteria only, inclusive of the chloroplast lineage, and are less likely to have existed in the common ancestor of both Archaea and Bacteria. This led to the proposal that the photosynthetic metabolism must have appeared after the advent of the respiratory metabolism.[36] The evidence for the existence of the respiratory components prior to the photosynthetic ones is contrary to a common belief that oxygenic photosynthesis must precede respiration, since the respiratory process requires oxygen as substrate. However, the early-evolving cytochrome *bc* complex may have performed either an anaerobic type of respiration for energy conversion or, as recently suggested, an aerobic type of respiration in the presence of an extremely low level of oxygen in a primitive Earth environment.[37,38] The oxygen is considered to have resulted from photolysis of water molecules. Thus, there was ample opportunity and evidence for the respiratory components to have existed prior to the development of photosynthesis. This respiration early hypothesis forms the foundation for considering cytochrome *b* as root to photosynthetic reaction centers between respiratory and photosynthetic electron transfer proteins.

The proposal that a cytochrome *b*-like protein is ancestral to reaction center proteins requires that a membrane-bound cytochrome be capable of substituting hemes with (bacterio)chlorophylls. Interestingly, Bartsch *et al.*[39] observed that aerobically grown *R. capsulatus* contains a *b*-type cytochrome

(cytochrome b_{561}) with a midpoint potential of −65 mV, which is actually comprised of a homodimer of the light-harvesting II-β (LHII-β) polypeptide containing a bound heme. This indicates that, under aerobic conditions before bacteriochlorophyll is synthesized, a dimer of LHII-β forms a ligand to heme Fe at the His30 residue and functions as a cytochrome. Under anaerobic (i.e. phototrophic) conditions where bacteriochlorophyll is synthesized, the LHII-β polypeptide complexes with an LHII-α polypeptide to form a heterodimer light-harvesting complex. In this case, the His30 residue of the β polypeptide forms a ligand to the Mg in bacteriochlorophyll. Thus, the LHII-β polypeptide appears capable of binding either heme or bacteriochlorophyll. This observation indicates that, under certain conditions, Mg–tetrapyrroles and Fe–tetrapyrroles can be readily exchanged among tetrapyrrole-binding membrane-spanning proteins.

Fisher and Rich[40] observed an interesting similarity for quinone binding sites between cytochrome *b* and type II reaction centers. A histidine triad with an L–X(3)–H–X(2–3)–S/T motif is shared by both types of proteins. In addition, there is a striking similarity at the three-dimensional level. The common feature for the quinone binding sites may indicate a certain evolutionary relatedness between the respiratory and photosynthetic electron transfer proteins.

A recent functional classification of transmembrane transporter proteins places the respiratory and photosynthetic transmembrane proteins as two families within the "primary active transporter" superfamily.[41] In a protein classification system on the basis of tertiary structures (SCOP classification[42]), the cytochrome bc_1 complex and photosynthetic reaction centers, along with a number of other transmembrane proteins, are placed within the "membrane all-alpha" superfamily with presumed distant homology.[43] Our identification of a putative evolutionary link between the respiratory and photosynthetic systems supports the above classification schemes with more detailed evidence, which pinpoints a closer relationship between two members of the same superfamily involved in two different metabolic pathways.

Materials and Methods

Sequence analysis for reaction center polypeptides was conducted using a gapped BLAST[18] search against a non-redundant protein database in GenBank. The BLAST search first performed in June 1999 was done by using the BLOSUM62 matrix for amino acid sequences and by choosing the Expect threshold level as 10, Descriptions 1000 and Alignments 1000 without the Low complexity filter. This search analysis was later repeated using a non-redundant protein database at San Diego Supercomputer Center by using the Biology Workbench web server†. In this BLAST search, we used the BLOSUM45 amino acid substitution matrix, with a gap opening penalty 12, extension penalty 2, expectation value 10, l-line description 1000 and alignments 1000. PSI-BLAST (position-specific iterative BLAST[18]), which is a more sensitive search method using iterative searches to detect weak homologies, was also carried out for type II reaction center core polypeptides at NCBI BLAST server‡. Multiple sequence alignment of full-length cytochrome *b* and type II reaction center core polypeptides was conducted with the aid of the CLUSTAL X program (version 1.63b),[19] the result of which was further modified on the basis of alignment generated by the DIALIGN 2 program§.[20] The alignment was further refined manually. Statistical significance of pairwise sequence similarities was evaluated by an alignment-independent program PRSS, which calculates the probability of similarities of randomly shuffled and unshuffled sequences using a distance matrix Monte Carlo procedure.[21] PRSS was accessed through the Biology Workbench. The analysis was carried out by setting the gap-opening penalty as 12 and gap-extending penalty as 2, and by performing 1000 global shuffles using the BLOSUM50 scoring matrix.

Detailed structural comparison between the two types of transmembrane proteins was done using a UNIX Silicon Graphics O2 workstation in a QUANTA 3-D graphical environment (Accelrys, San Diego, CA). X-ray crystal structures of a mitochondrial cytochrome *b* (PDB code: 1BCC[22]) and a purple bacterial reaction center (2RCR[23]) were retrieved from the Protein Data Bank. Due to the lack of high-resolution crystal structures for the photosystem II (PSII) reaction center, we used a modeled structure[44] with slight modifications.

Overall three-dimensional structural comparison was also done between the two types of photosynthetic reaction centers using the DALI server‖. The pairwise comparison was done between the helical regions of the two types of photosynthetic reaction centers, which include residues 34–53, 85–107, 117–137, 172–198 and 227–250 for PufL, residues 53–73, 112–135, 143–159, 199–223 and 263–283 for PufM, residues 435–464, 533–559, 592–620, 670–691, and 725–750 for PsaA and residues 416–448, 520–545, 578–605, 650–670 and 711–737 for PsaB.

To probe the close structural similarity of the two types of reaction centers, we used a threading method, which performs structural fold recognition on the basis of sequence profile alignments, secondary structure prediction, and energy calculations. The fold recognition of reaction center core apoproteins was done using the GenTHREADER program¶[27] with profile and secondary structure input.

Phylogenetic trees were reconstructed using the Bayesian inference (BI) method,[24] which finds trees by calculating posterior probabilities using a MCMCMC sampling approach. The analysis was performed using the MRBAYES program[45] during which four simultaneous Markov chains were run under the Jones, Taylor and Thornton (JTT) amino acid substitution matrix.[46] The starting *a priori* tree was estimated using the distance neighbor-joining method[47] by CLUSTAL X with 1000 bootstrap replications. The Bayesian program was allowed to run for 50,000 generations with a tree

† http://workbench.sdsc.edu/

‡ http://www.ncbi.nlm.nih.gov/BLAST/

§ http://bibiserv.techfak.uni-bielefeld.de/cgi-bin/dialign_submit

‖ http://www.ebi.ac.uk/dali/Interactive.html

¶ http://bioinf.cs.ucl.ac.uk/psiform.html

sampling frequency every ten generations. The "burn-in" period when the likelihood values had not reached convergence was determined approximately to be 5000 generations. Thus, the first 10,000 generations representing the first 1000 trees were excluded from analysis of posterior probability distribution. The remaining 4000 trees were used to determine a consensus BI tree. Maximum likelihood non-parametric bootstrap values of the final tree were generated by MRBAYES. The posterior probability values of the final tree were derived with a consensus building feature of the PAUP* program (version 4.0b3a)[48] using a 50% majority rule.

Ancestral amino acid sequences of reaction center proteins at internal nodes of phylogenetic trees were also inferred with a likelihood-based Bayesian method using the CODEML program in the PAML package,[49] which performed inference using the JTT model.

Acknowledgements

We thank Yuichi Fujita and Howard Gest for valuable suggestions during preparation of the manuscript. We thank Terry Meyer for helpful suggestions at the proof stage. This work is supported by an NIH grant GM53940 (to C.E.B.) and an Arizona State University NASA astrobiology grant (to C.E.B.) as well as a faculty startup fund (to J.X.) from Texas A&M University.

References

1. Mauzerall, D. (1992). Light, iron, Sam Granick and the origin of life. *Photosynth. Res.* **33**, 163–170.
2. Hartman, H. (1998). Photosynthesis and the origin of life. *Origins Life Evol. Biosph.* **28**, 515–521.
3. Braiser, M. D., Green, O. R., Jephcoat, A. P., Kleppe, A. K., Van Kranendonk, M. J., Lindsay, J. F. *et al.* (2002). Questioning the evidence for Earth's oldest fossils. *Nature*, **416**, 76–81.
4. Nitschke, W., Mühlenhoff, U. & Liebl, U. (1997). Evolution. In *Photosynthesis: A Comprehensive Treatise* (Raghavendra, A., ed.), pp. 285–304, Cambridge University Press, Cambridge.
5. Olson, J. M. (1999). Early evolution of chlorophyll-based photosystems. *Chemtracts*, **12**, 468–482.
6. Blankenship, R. E. (1992). Origin and early evolution of photosynthesis. *Photosynth. Res.* **33**, 91–111.
7. Blankenship, R. E. (2001). Evidence for the early evolution of photosynthesis. *Trends Plant Sci.* **6**, 4–6.
8. Xiong, J. & Bauer, C. E. (2002). Complex evolution of photosynthesis. *Annu. Rev. Plant. Biol.* **53**, 503–521.
9. Schubert, W.-D., Klukas, O., Saenger, W., Witt, H. T., Fromme, P. & Krauss, N. (1998). A common ancestor for oxygenic and anoxygenic photosynthetic systems: a comparison based on the structural model of photosystem I. *J. Mol. Biol.* **280**, 297–314.
10. Rhee, K. H., Morris, E. P., Barber, J. & Kühlbrandt, W. (1998). Three-dimensional structure of the plant photosystem II reaction centre at 8 Å resolution. *Nature*, **396**, 283–286.
11. Zouni, A., Witt, H.-T., Kern, J., Fromme, P., Krauß, N., Saenger, W. & Orth, P. (2001). Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature*, **409**, 739–743.
12. Xiong, J., Fischer, W., Inoue, K., Nakahara, M. & Bauer, C. E. (2000). Molecular evidence for the early evolution of photosynthesis. *Science*, **289**, 1724–1730.
13. Mulkidjanian, A. Y. & Junge, W. (1997). On the origin of photosynthesis as inferred from sequence analysis. *Photosynth. Res.* **51**, 27–42.
14. Meyer, T. E. (1994). Evolution of photosynthetic reaction center and light harvesting chlorophyll proteins. *BioSystems*, **33**, 167–175.
15. Vermaas, W. F. J. (1994). Evolution of heliobacteria: implications for photosynthetic reaction center complexes. *Photosynth. Res.* **41**, 285–294.
16. Olson, J. M. (2001). Evolution of photosynthesis (1970), re-examined thirty years later. *Photosynth. Res.* **68**, 95–112.
17. Baymann, F., Brugna, M., Mühlenhoff, U. & Nitschke, W. (2001). Daddy, where did (PS)I come from? *Biochim. Biophys. Acta*, **1507**, 291–310.
18. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
19. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
20. Morgenstern, B. (1999). DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics*, **15**, 211–218.
21. Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
22. Zhang, Z., Huang, L., Shulmeister, V. M., Chi, Y. I., Kim, K. K., Hung, L. W. *et al.* (1998). Electron transfer by domain movement in cytochrome *bc*1. *Nature*, **392**, 677–684.
23. Chang, C. H., El-Kabbani, O., Tiede, D., Norris, J. & Schiffer, M. (1991). Structure of the membrane-bound protein photosynthetic reaction center from *Rhodobacter sphaeroides*. *Biochemistry*, **30**, 5352–5360.
24. Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310–2314.
25. Yang, Z., Kumar, S. & Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*, **141**, 1641–1650.
26. Ruffle, S. V., Donnelly, D., Blundell, T. L. & Nugent, H. A. (1992). A three-dimensional model of the Photosystem II reaction center of *Pisum sativum*. *Photosynth. Res.* **34**, 287–300.
27. Jones, D. T. (1999). GenTHREADER: an efficient and reliable protein fold recognition method for genomics sequences. *J. Mol. Biol.* **287**, 797–815.
28. Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science*, **273**, 595–602.
29. Matsuo, Y. & Bryant, S. H. (1999). Identification of homologous core structures. *Proteins: Struct. Funct. Genet.* **35**, 70–79.
30. Xiong, J., Inoue, K. & Bauer, C. E. (1998). Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*. *Proc. Natl Acad. Sci. USA*, **95**, 14851–14856.
31. Wang, J., Gosztola, D., Ruffle, S. V., Hemann, C., Seibert, M., Wasielewski, M. R. *et al.* (2002). Functional

asymmetry of photosystem II D1 and D2 peripheral chlorophyll mutants of *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA*, **99**, 4091–4096.
32. Cogdell, R. J. & Fyfe, P. K. (1996). Purple bacterial antenna complexes. *Curr. Opin. Struct. Biol.* **6**, 467–472.
33. Schütz, M., Brugna, M., Lebrun, E., Baymann, F., Huber, R., Stetter, K.-O. *et al.* (2000). Early evolution of cytochrome *bc* complexes. *J. Mol. Biol.* **300**, 663–675.
34. Castresana, J., Lübben, M., Saraste, M. & Higgins, D. G. (1994). Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. *EMBO J.* **13**, 2516–2525.
35. Castresana, J., Lübben, M. & Saraste, M. (1995). New archaebacterial genes coding for redox proteins: implications for the evolution of aerobic mechanism. *J. Mol. Biol.* **250**, 202–210.
36. Castresana, J. & Saraste, M. (1995). Evolution of energetic metabolism: the respiration-early hypothesis. *Trends Biochem. Sci.* **20**, 443–448.
37. Kasting, J. F. (1993). Earth's early atmosphere. *Science*, **259**, 920–926.
38. Saraste, M., Castresana, J., Higgins, D., Lübben, M. & Wilmanns, M. (1996). *Origin and Evolution of Biological Energy Conversion* (Baltscheffsky, H., ed.), pp. 255–289, Wiley-VCH, New York.
39. Bartsch, R. G., Caffrey, M. S., van Beeumen, J. J., Salamon, Z., Tollin, G., Meyer, T. E. & Cusanovich, M. A. (1993). Purification and properties of an unusual membrane-derived cytochrome *b*-561 from the purple phototrophic bacterium *Rhodobacter capsulatus*, which is structurally related to the bacteriochlorophyll-binding protein, LHII β. *Arch. Biochem. Biophys.* **304**, 117–122.
40. Fisher, N. & Rich, P. R. (2000). A motif for quinone binding sites in respiratory and photosynthetic systems. *J. Mol. Biol.* **296**, 1153–1162.
41. Saier, M. H., Jr (2000). A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.* **64**, 354–411.
42. Murzin, A., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
43. Russell, R. B., Sasieni, P. D. & Sternberg, M. J. (1998). Supersites within superfolds. Binding site similarity in the absence of homology. *J. Mol. Biol.* **282**, 903–918.
44. Xiong, J., Subramaniam, S. & Govindjee (1998). A knowledge-based three-dimensional model of the photosystem II reaction center of Chlamydomonas reinhardtii. *Photosynth. Res.* **56**, 229–254.
45. Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic tree. *Bioinformatics*, **17**, 754–755.
46. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.
47. Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
48. Swofford, D. L. (1999). *PAUP*: Phylogenetic Analysis Using Parsimony and other Methods*, Sinauer, Sunderland, MD
49. Yang, Z. (2000). *Phylogenetic Analysis by Maximum Likelihood (PAML)*, version 3.0, University College London, London, England
50. Carson, M. (1991) Ribbons. J. Appl. Crystallog. 24, 958–961.

Edited by R. Huber

(Received 29 November 2001; received in revised form 29 July 2002; accepted 30 July 2002)

http://www.academicpress.com/jmb

Supplementary Material for this paper comprising three Figures is available on IDEAL