Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimica et Biophysica Acta 1655 (2004) 133–139
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Review

The evolutionary development of the protein complement of Photosystem 2

Jason Raymond*, Robert E. Blankenship

Blakenship Laboratory, Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-1604, USA

Received 27 May 2003; accepted 21 October 2003

Abstract

During the transition from anoxygenic to oxygenic photosynthesis, the Type 2 reaction center underwent many changes, none so dramatic as the remarkable increase in complexity at the protein level, from only three or four subunits in the anoxygenic reaction center to possibly more than 25 in Photosystem 2 (PS2). The evolutionary source of most of these proteins is enigmatic, as they have no apparent homology to any other proteins in existing databases. However, some of the proteins in PS2 have apparent homologies to each other, suggesting ancient gene duplications have played an important role in the development of the complex. These homologies include the well-known examples of the D1 and D2 reaction center core proteins and the CP43 and CP47 core antenna proteins. In addition, PsbE and PsbF, the two subunits comprising cytochrome *b*-559, show homology to each other, suggesting that a homodimeric cytochrome preceded the heterodimeric one. Other potential homologies that appear to be statistically significant include PsbV with the N-terminal part of D1 and PsbT with PsbI. Most of the proteins that make up the photosynthetic apparatus bear no relation to any other proteins from any source. This suggests that a period of remarkable evolutionary innovation took place when the ability to make oxygen was invented. This was probably a response to the production of highly toxic oxygen and these new proteins served to protect and repair the photosynthetic apparatus from the harmful effects of oxygen.

Keywords: Photosystem 2; Evolution; Oxygenic photosynthesis

1. Introduction

The origin and evolution of oxygenic photosynthesis is one of the most important innovations in the history of life on Earth, yet remains very poorly understood. There are many approaches that can be used to address this problem, including using the geological record to date the appearance of oxygen in the atmosphere and oceans, microfossil and biomarker analysis, molecular evolution analysis of existing organisms, search for and characterization of possible transitional forms, analysis of pigment biosynthetic pathways and complete genome analysis of photosynthetic organisms [1–5]. None of these approaches could itself solve this most difficult and important problem, but a multidisciplinary approach in which evidence is assembled from a variety of sources is the only possible way to unravel this issue. Fig. 1 shows a conceptual picture of this evolutionary transition from anoxygenic to oxygenic photosynthesis. Changes that needed to take place before oxygenic photosynthesis could work include an alteration in the energetics of pigments and redox reactions, the genesis of the oxygen evolution complex itself, and the development of the ability to protect against the oxygen generated in the complex [2,6,7].

Perhaps the most striking difference between the anoxygenic bacterial reaction centers (RCs) and Photosystem 2 (PS2) is the dramatically increased complexity of the PS2 RC, exemplified by the much larger number of protein subunits in PS2. Experimental evidence suggests that, collectively, there are at least 18 and possibly more than 25 distinct subunits in PS2 from cyanobacteria and plants, as compared to three or four in the bacterial RCs [8–14]. Table 1 lists the known PS2 proteins that have been found in cyanobacterial genomes along with their proposed or elucidated functions. These include several extrinsic subunits that comprise or closely interact with the oxygen evolving complex (OEC), as well as a suite of low molecular weight single transmembrane peptides peripheral to the core RC. In many cases, the functions

* Corresponding author. Tel.: +1-408-965-1437.
E-mail address: jasonraymond@asu.edu (J. Raymond).

0005-2728/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbabio.2003.10.015

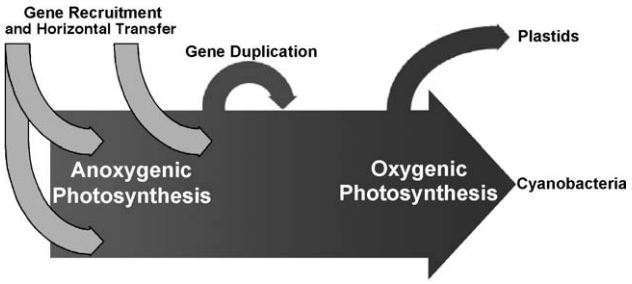

Fig. 1. Schematic of the transition from anoxygenic to oxygenic photosynthesis and associated crucial genome-level events, as discussed in recent literature and highlighted in the text.

of these subunits have not been established, and for the most part their evolutionary origin is a complete mystery. None of the peripheral proteins (herein referring to RC proteins other than D1, D2, CP43 and CP47) have unambiguous homologs outside of the cyanobacterial domain, though interestingly a small number are distantly similar to one another. In this paper, we focus on the protein composition of the Type 2 RCs found in certain anoxygenic phototrophs and PS2 of cyanobacteria and try to gain an understanding of this aspect of the evolution of oxygenic photosynthesis (Fig. 2).

Because of their small size, the significance of several putative homologs to PS2 proteins is difficult to assess statistically as local alignments, e.g. as performed by NCBI's BLAST software. Any protein alignment can be expected to produce some matches simply due to chance, and this becomes particularly problematic when aligning short proteins where the chance alignment may cover a significant extent of the total length of the alignment. BLAST and other local alignment tools are normalized against this effect [15], but the problem nonetheless renders difficult the detection and annotation of small proteins and is one of the primary arguments for sequencing genomes from closely related organisms. This scenario is evident for the numerous small, single transmembrane (TM) proteins thought to be associated with PS2, which are—more frequently than might be expected—either overlooked or not properly annotated in public databases. Though the best medicine for this malady is undoubtedly more sequence data, this can be in part overcome by performing global alignments (comparing entire sequences, rather than searching for 'windows' of high identity as is done in local alignment-based database searches). Pairwise and multiple alignment are both based on NP-hard algorithms and so applying these approaches to entire genomes worth of data

Table 1
Cyanobacterial PS2 proteins, gene names, cofactors, and known functions

Subunit name	Gene	Mass (kDa)	Cofactors	Function	Number of invariant sites/ multiple alignment length	Average number of differences
D1	psbA	39	chlorophyll, pheophytin, β-carotene, Fe	core RC	183/371	49.9
D2	psbD	39	chlorophyll, pheophytin, β-carotene, Fe	core RC	235/370	37.8
CP43	psbC	51	chlorophyll, β-carotene	core antenna	214/499	117
CP47	psbB	56	chlorophyll, β-carotene	core antenna	226/556	129
Cyt *b*-559a	psbE	9	heme	core RC	35/87	27.5
Cyt *b*-559b	psbF	4	heme	core RC	19/51	13
H	psbH	8	phosphate	photoprotection, Q_a to Q_b regulation	21/79	23.2
I	psbI	4		core RC	18/42	12.4
J	psbJ	4		PS2 assembly	12/67	15.4
K	psbK	4		?	16/48	14.6
L	psbL	4		role in Q_a binding	16/41	9.07
M	psbM	4		?	14/50	15.6
N	psbN	5		role in PS2 stability	10/58	19.9
OEC33	psbO	27		Mn cluster stability, Ca^{2+} and Cl^- binding	29/320	141
OEC26	psbP	20		Ca^{2+} and Cl^- binding (? in cyanobacteria)	16/258	104
Tc	psbT	3		role in PS2 stability	8/40	12.9
U	psbU			role in O_2 evolution	15/177	72.6
Cyt *c*-550	psbV	15	heme	role in O_2 evolution	23/188	86.6
W	psbW/28	6			18/132	53.4
X	psbX	4		role in Q_a function	5/60	19.8
Y	psbY	6			7/45	21.6
Z	psbZ/27	9		antenna–RC interaction	21/149	72.1

Also given are the number of invariant sites in the entire alignment and the number of differences (amino acid substitutions) averaged over all sequences in the multiple alignment (adapted and modified from Refs. [8,9,14]).

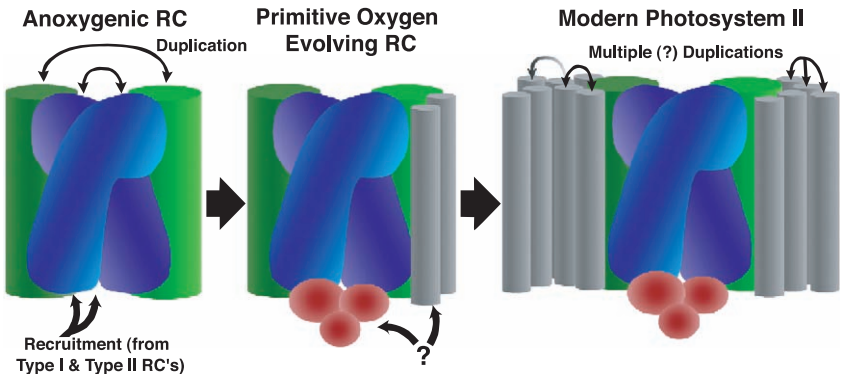

Fig. 2. Successive steps in the evolution of cyanobacterial PS2, from most primitive to modern (left to right). The reaction center (D1 and D2, blue) and antenna proteins (CP43 and CP47, green) clearly evolved via gene duplication and, respectively, share a common ancestor with the type 2 RC and the antenna domain of the type 1 RC. The evolutionary history of the extrinsic OEC components (red) is still unclear. Also not known is the origin of the low molecular weight intrinsic subunits (gray) that, though largely functionally uncharacterized, show some signs of being diverged duplicates of one another. This supports the notion of a primitive cyanobacterial PS2 that was genetically as well as structurally a much simpler complex.

quickly becomes computationally intractable [16]. Additionally, score distributions and statistical metrics for global sequence comparisons are not rigorously known [15]. Here we attempt to circumvent computation costs and gather at least some sense of the significance of alignments by performing global alignments on a reduced set of smaller (approximating the size distribution of photosynthetic proteins) set of curated Swiss-Prot sequences.

One important caveat in comparing sequences whose divergence from one another was an ancient event (as is the case for these proteins, as discussed below) is that some proteins that are in fact related may have changed to such an extent that no homology remains that can be unambiguously detected. This is one factor thought to contribute to the preponderance of so-called "orphan" genes in whole genome sequences. Without exception, orphan genes have constituted roughly one-quarter to one-half of putative genes in every complete genome to date, have no clear family or genetic relatives to which they are related and usually no known or predictable function [17]. Though the mystery of orphan genes may in part be the result of sampling bias (e.g. we have only examined the tip of the protein universe iceberg), it is also likely due to the prowess with which natural selection acts to promulgate beneficial genes and eradicate the riff-raff and the redundant. Models based on this balancing act between birth and death of new genes have been successfully used to predict protein family size distribution as well as a constituency of orphan genes [18–20]. These models establish that many protein families have likely diverged too rapidly for us to be able to gauge homology between once closely related gene duplicates. By analogy, homology detection might be thought of as a phylogenetic signal-to-noise issue, with signal corresponding to conserved nucleotides, amino acids, or protein folds, for example, and noise being the mutations that continually accrue after a pair of sequences diverge. All else being equal, two relatively long sequences will begin with a greater phylogenetic signal at the time of their divergence as compared to two short sequences. Returning to the RC complex, the fact that only very distant identity is observed, for example, between the large D1 and D2 subunits of PS2 and their L and M bacterial homologs suggests that the origin and evolution of many of the peripheral RC subunits will need to be deciphered within the context of high resolution structural data from a diverse group of photosynthetic organisms.

2. Methods

All proteins between 30 and 60 amino acids in size in the curated Swiss-Prot database (108,251 sequences at the time of download) were retrieved and aligned to each of the PS2 proteins with every protein in this reduced set. Because many of the peripheral PS2 proteins are single transmembrane spanning peptides (which can skew alignments because there are effectively fewer possible amino acid substitutions over short stretches of TM proteins), the Swiss-Prot subset was further divided into proteins with predicted TM spanning domains in order to determine if these alignments showed any overall increase in percent identity [21]. As a final test of significance, the program PRSS was used to calculate optimal alignment scores as well as alignment scores of 1000 sequence-shuffled realignments for every pair of proteins in our set [22].

Whole genome comparisons were made through a local protein database comprised of complete or nearly complete microbial genomes (106 total, including 10 cyanobacteria and 7 anoxygenic phototrophs) available through GenBank. Local database searching was carried out using NCBI's BLASTP program with a threshold expectation value of $1e-4$ considered significant (though hits were extended out to an expectation value of 10). Database searching was done using cyanobacterial query proteins from the 'photosynthesis' subset of the KEGG database [23], which includes Photosystem 1 and 2 and associated subunits, phycobilipro-

teins and light harvesting complexes, and several soluble and integral membrane cytochromes such as the subunits of the cytochrome b_6f complex. Each of the PS-2 proteins was then used as a BLAST query against available cyanobacterial genomes [*Synechocystis PCC6803*, *Prochlorococcus* strains *MED4, MIT9313,* and *SS120, Nostoc PCC7120, Nostoc punctiforme*, *Trichodesmium erythraeum*, *Thermosynechococcus elongatus*, *Synechococcus WH8102*, and *Gloeobacter violaceus*]. All putative homologs with a BLAST e-value below 10 to one of the KEGG photosynthesis proteins were retained. Subsequently, these putative homologs were manually refined and removed based on multiple alignment using CLUSTALW (default parameters) as well as additional information such as the known approximate size of the proteins, likely number of transmembrane spanning regions, and conserved sequence motifs and cofactor-binding sites. All multiple alignments are available upon request.

3. Results

Comparison of known RC proteins to a large (~100 genomes) database of fully sequenced microbial genomes illustrates both the utility and limitations of primary sequence comparison. For example, for the suite of Psb proteins comprising PS2, only the D1 and D2 proteins have detectable homologs in organisms other than cyanobacteria (the L and M reaction center subunits found in *Chloroflexus aurantiacus* and photosynthetic proteobacteria). Arguments for a common evolutionary origin of type 2 RC's have been bolstered by recent structural analysis and extensive literature is available on this subject [24–28], so this will not be discussed further here. Though CP43 and CP47 (PsbC and PsbB, respectively) are clearly related to one another, they have no counterpart either in the type 2 RC of *Chloroflexus* or of purple bacteria [29]. They are, however, distantly similar to one another—suggesting a gene duplication event—and also to the N-terminal (antenna) domain of the type 1 RCs found in green sulfur bacteria and heliobacteria as well as the pcb and isiA antenna complexes [29–31]. The available evidence strongly suggests a common evolutionary origin of type 1 and type 2 RC's, as well as a common ancestor for all photosynthetic RCs [6,25,32,33].

Of the extrinsic subunits, including the multiple proteins comprising the OEC, similarity is found only between the cyanobacterial cytochrome *c*-550 (PsbV) and *c*-type cytochromes (especially cytochrome c_6/*c*-553) from other anoxygenic as well as non-photosynthetic bacteria. Based on conserved structural folds and 3-D topology, it has been postulated that the cytochrome *c* superfamily stems from a common ancestor [34]. The other cyanobacterial OEC proteins (PsbO and PsbU) exhibit no significant similarity to any other known proteins. Interestingly, all cyanobacterial genomes analyzed to date have a clear PsbP homolog, which has been proposed only to function in the OEC of green algal and plant photosystems. Furthermore, the related highly reduced marine *Prochlorococcus* MED4 and SS120 genomes both appear to be missing PsbU and PsbV [35]. Though the biochemical relevance of these observed variations are not yet clear, they do suggest that some degree of plasticity may exist in the composition of the OEC.

Employing the metrics described in Methods, several pairwise alignments of PS2 proteins are statistically significant. PsbE and PsbF, the alpha and beta subunits of cytochrome *b*-559, show sequence similarity to each other that is significant ($P<0.05$) based on PRSS shuffled alignments. These two cytochrome subunits show particular conservation of residues around their heme-ligating histidines. Schematically, an origin from a homodimeric cytochrome seems plausible and is consistent with the tentative assignments of the alpha helical (heme-binding) region of these two proteins. Surprisingly, PRSS alignments also were significant for the cytochrome *c*-550 (PsbV) protein and the D1 (PsbA) protein from *T. elongatus*. Aligned residues are concentrated particularly around several tryptophans at the N-terminus of each of these proteins. Given the tentative function of cytochrome *c*-550, it is difficult to extrapolate the meaning of this alignment.

Several other pairwise alignments stand out from the background noise of the Swiss-Prot random alignments. PsbT and PsbI share over one-third identical residues on alignment, though the former is an extrinsic protein that co-purifies with PsbO [36], while the latter is a single transmembrane protein. PsbT also shows distant identity with PsbE above the 30% threshold, concentrated mostly in a small phenylalanine-rich segment of the peptide sequence but is otherwise interspersed with multiple gaps. Several additional PS2 proteins show very distant identity to one another, for example pairwise combinations of PsbI, PsbH, and PsbJ are on the order of 25–30% identical on alignment and, though worth noting, these identities are not statistically significant based on our random Swiss-Prot alignments (Fig. 3).

Table 1 also lists the degree of conservation of the known and putative cyanobacterial OEC proteins from the 10 presently available cyanobacterial genomes. The D1 and D2 proteins are the least variable of this group with roughly 64% and 49%, respectively, of the total residues absolutely conserved across multiple duplicates from each of the 10 genomes. With the exception of the reduced *Prochlorococcus* strains MED4 and SS120, D1 is found in multiple copies in all cyanobacterial genomes (D2 is duplicated in about half of the genomes). Among *Synechocystis PCC6803*, these multiple copies are regulated in response to changing light intensity and redox conditions and all appear to be at least transcriptionally active under normal and Fe-deficient growth conditions [37–39]. Based on protein phylogeny, most of the D1 paralogs and all of the D2 paralogs have occurred since the

divergence of these 10 organisms, so that the closest genetic relative for any given gene is the paralog within that same genome (one notable exception involves *N. punctiforme* and *Nostoc PCC7120*, which appear to have had two copies of D1 at the time of their divergence that have survived to the present, only one of which was extensively duplicated in both organisms subsequent to speciation). This suggests that, supposing niche adaptation is the primary selective pressure for RC duplication and maintenance, cyanobacterial core RC genes may have experienced a persistent duplication-and-loss turnover process that, coupled to the fine tuning of regulation and expression, could facilitate their adaptation to such a wide variety of environments.

As mentioned above, the marine cyanobacterial strains analyzed herein, including the three *Prochlorococcus* strains for which genomes are now available, show some deviation in their Psb gene content. These organisms exhibit highly reduced genomes, presumed to be optimized for survival in the pelagic desert in which they exist, and so it seems most plausible that these genes have been lost during the course of evolution [35]. Additionally, the genome of *G. violaceus* has recently become available and, though fairly typical with respect to Psb gene content (with the exception of missing PsbY and PsbZ), sequence analysis and phylogeny support the notion that this organism is an early branching cyanobacterium. The Psb proteins analyzed herein almost invariably segregate on phylogenetic trees into two clades: one comprised of the three *Prochlorococcus* strains and *Synechococcus WH8102*, and the second including mostly non-marine cyanobacteria. However, the photosystem proteins from *Gloeobacter* are routinely distinct and in several cases strikingly different from counterparts found in the other cyanobacterial genomes (Table 2).

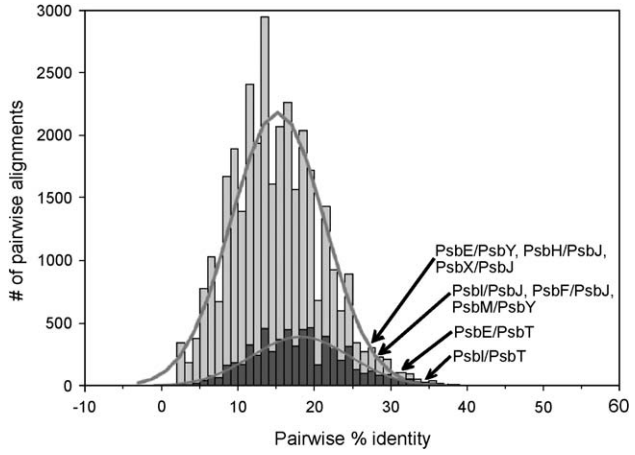

Fig. 3. (Light gray bars) Combined distribution of pairwise identity scores for randomized alignments of the low molecular weight intrinsic subunits of PS2 versus all protein sequences between 30 and 60 residues in the Swiss-Prot repository, and a Gaussian distribution (thick line) of the same mean and standard deviation. Dark gray bars show pairwise identity scores of PS2 subunits versus the single transmembrane Swiss-Prot subset, with the corresponding Gaussian shown as a thin line. Also indicated is the bin for several high scoring pairwise alignments of intrinsic proteins.

Table 2
Pairwise percentage identity between low molecular weight extrinsic PS2 proteins

psbF	psbL	psbJ	psbK	psbM	psbN	psbH	psbT	psbI	psbX	psbY	
(25.0)	13.5	25.6	23.9	19.4	21.4	16.9	*31.3*	26.3	22.0	(27.9)	psbE
	24.3	28.2	25.6	20.6	9.5	15.6	21.9	19.4	17.8	11.9	psbF
		22.2	8.1	21.2	19.4	21.6	13.8	19.4	21.6	25.0	psbL
			18.9	13.9	16.2	**27.5**	15.6	28.6	*27.5*	16.2	psbJ
				17.1	20.9	21.7	21.9	7.9	23.9	26.2	psbK
					13.9	22.2	6.3	13.9	16.7	**28.6**	psbM
						23.3	18.8	18.4	18.6	11.9	psbN
							15.6	18.4	16.0	18.6	psbH
								34.4	18.8	12.5	psbT
									24.3	20.0	psbI
										16.3	psbX

Significance at $P<0.1$ in italics, $P<0.05$ in bold, and significance based on PRSS shuffled alignments in parentheses. Note that significance is based on each Psb protein aligned against each protein from the entire Swiss-Prot set and so may vary from protein to protein.

4. Discussion

Remarkably, simply the presence in a genome of a sequence even distantly related to one of the dozens of characterized RC or light-harvesting proteins is sufficient to be diagnostic of that organism being capable of carrying out photosynthesis. This sort of litmus test is quite astonishing from a bioinformatical perspective. The concept of protein families has been used with incredible success to categorize a substantial fraction of known protein sequences into discrete sets of homologs, for example, into groups of related membrane-spanning dehydrogenases or soluble cytochromes. The utility of this approach is a testament to the central tenet of Darwinian evolution, that is, the idea of vertical descent with modification. However, the components of the photosynthetic apparatus have by and large been recalcitrant to this sort of categorization and, with few exceptions, each photosynthetic protein comprises a singular, homogeneous protein family. Phylum cyanobacteria is particularly enigmatic in this regard, for one because of the explosion in complexity of their photosynthetic machinery, something mirrored nowhere else in the bacterial domain. Exemplifying this, a recent whole genome comparison delineated 181 genes common to eight species of cyanobacteria with no clear homologs in any other bacteria [40]. The vast majority of the 43 genes that have been characterized in this signature set indeed are either constituents of or directly involved with the photosynthetic apparatus, as are a dozen or so of the remaining 138 uncharacterized genes whose functions have been elucidated since the original analysis (G.E. Fox, personal communication). Clearly, oxygenic photosynthesis was the dominant force in early cyanobacterial diversification and remains the hallmark of extant species.

There are several possible explanations for the paucity of cyanobacterial protein homologs in non-photosynthetic organisms, ranging from lack of sensitivity (or of available information) in the comparison of distant sequences to missing data due to limited sampling of microbial diversity. It is also very likely that, spurred by the development of oxygenic photosynthesis, cyanobacteria underwent a period of dramatically accelerated evolution. The toxicity of molecular oxygen, especially exacerbated in the presence of light-excited pigments, must have been nothing short of calamitous to organisms in a hitherto anaerobic world. Cyanobacteria, and in particular PS2, represent ground zero in the invention of this Earth-transforming process and this, along with the unprecedented availability to use H_2O as a reductant, likely provided an enormous impetus for evolutionary leaps in the complexity of the photosynthetic apparatus.

What is known about the timing and Precambrian biogeochemical milieu circa 2 billion years ago, just preceding the so-called age of cyanobacteria? The paleobiochemistry of these organisms has been well studied, and several independent mileposts exist in the form of micro- and macrofossils (e.g. stromatolites, [1]), diagnostic biomarkers [3], and environmental redox states as established through geology and isotope geochemistry [41]. This complementary and often convergent evidence suggests that the cyanobacteria had emerged by 2.7 billion years ago and oxygenic photosynthesis was invented shortly thereafter, almost certainly by 2.0 billion years ago [41]. The oldest microfossils of what are purported to be macroscopic algae belong to *Grypania* sp., and have been dated between 1.7 and 2.1 billion years old [1]. This is significant in that, based on comparative genomics and studies of microbial, plant and algal photosystems [42,43], nearly the entire spectrum of PS2 proteins had originated by the time the first plastid endosymbiosis occurred.

The intriguing conclusion is that the oxygenic photosystem evolved within a remarkably narrow window of geologic time and, moreover, has remained largely unchanged with respect to subunit composition for perhaps 2 billion years. As yet to be clarified and central to our understanding of the early evolution of oxygenic photosynthesis is whether the modern complexity of the PS2 reaction center—most notably the extrinsic and peripheral subunits—preceded and thereby possibly facilitated the water oxidation reaction or, conversely, developed in conjunction with or subsequent to water oxidation, where the function might have been within the context of protection, regulation, or repair.

Acknowledgements

The NASA Exobiology program and the NASA Astrobiology Institute at ASU are gratefully acknowledged for support. The genomes of *N. punctiforme* and *T. erythraeum* were provided freely by the US DOE Joint Genome Institute. The genome of *G. violaceus* is freely available from the Kazusa DNA Research Institute web site (http://www.kazusa.or.jp/). This is publication no. 583 from the ASU Center for the Study of Early Events in Photosynthesis.

References

[1] J.P. Grotzinger, Controls on Carbonate Platform and Basin Development, Soc. of Econ. Paleontol. and Mineral, Tulsa, OK, 1989, p. 44. Special Publication.
[2] R.E. Blankenship, H. Hartman, Trends Biochem. Sci. 23 (1998) 94–97.
[3] J.J. Brocks, G.A. Logan, R. Buick, R.E. Summons, Science 285 (1999) 1033–1036.
[4] J. Xiong, W.M. Fischer, K. Inoue, M. Nakahara, C.E. Bauer, Science 289 (2000) 1724–1730.
[5] J. Raymond, O. Zhaxybayeva, J.P. Gogarten, S.Y. Gerdes, R.E. Blankenship, Science 298 (2002) 1616–1620.
[6] R.E. Blankenship, Photosynth. Res. 33 (1992) 91–111.
[7] G.C. Dismukes, V.V. Klimov, S.V. Baranov, Y.N. Kozlov, J. DasGupta, A. Tyryshkin, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 2170–2175.
[8] J. Barber, J. Nield, E.P. Morris, D. Zheleva, B. Hankamer, Physiol. Plant. 100 (1997) 817–827.
[9] R.E. Blankenship, Molecular Mechanisms of Photosynthesis, Blackwell, Oxford, 2002.
[10] K.H. Rhee, E.P. Morris, J. Barber, W. Kuhlbrandt, Nature 396 (1998) 283–286.
[11] B. Hankamer, E.P. Morris, J. Barber, Nat. Struct. Biol. 6 (1999) 560–564.
[12] Y. Kashino, H. Koike, M. Yoshio, H. Egashira, M. Ikeuchi, H.B. Pakrasi, K. Satoh, Plant Cell Physiol. 43 (2002) 1366–1373.
[13] Y. Kashino, W.M. Lauber, J.A. Carroll, Q. Wang, J. Whitmarsh, K. Satoh, H.B. Pakrasi, Biochemistry 41 (2002) 8004–8012.
[14] J. Barber, Q. Rev. Biophys. 36 (2003) 71–89.
[15] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. 215 (1990) 403–410.
[16] L. Wang, T. Jiang, J. Comput. Biol. 1 (1994) 337–348.
[17] I. Iliopoulos, S. Tsoka, M.A. Andrade, P. Janssen, B. Audit, A. Tramontano, A. Valencia, C. Leroy, C. Sander, C.A. Ouzounis, Genome Biol. 2 (2001) 1.1–1.3 (INTERACTIONS0001).
[18] E.V. Koonin, Y.I. Wolf, G.P. Karev, Nature 420 (2002) 218–223.
[19] B. Snel, P. Bork, M.A. Huynen, Genome Res. 12 (2002) 17–25.
[20] M.A. Huynen, P. Bork, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 5849–5856.
[21] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, J. Mol. Biol. 305 (2001) 567–580.
[22] W.R. Pearson, D.J. Lipman, Proc. Natl. Acad. Sci. 85 (1988) 2444–2448.
[23] M. Kanehisa, Novartis Found. Symp. 247 (2002) 91–101 (discussion 101-3, 119-28, 244-52).
[24] J. Deisenhofer, O. Epp, K. Miki, R. Huber, H. Michel, J. Mol. Biol. 180 (1984) 385–398.
[25] W.D. Schubert, O. Klukas, W. Saenger, H.T. Witt, P. Fromme, N. Krauss, J. Mol. Biol. 280 (1998) 297–314.
[26] A. Zouni, H.T. Witt, J. Kern, P. Fromme, N. Krauss, W. Saenger, P. Orth, Nature 409 (2001) 739–743.
[27] P. Jordan, P. Fromme, H.T. Witt, O. Klukas, W. Saenger, N. Krauss, Nature 411 (2001) 909–917.
[28] N. Kamiya, J.R. Shen, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 98–103.
[29] P.K. Fyfe, M.R. Jones, P. Heathcote, FEBS Lett. 530 (2002) 117–123.
[30] W.F. Vermaas, Photosynth. Res. 41 (1994) 285–294.
[31] J. La Roche, G.W. van der Staay, F. Partensky, A. Ducret, R.

Aebersold, R. Li, S.S. Golden, R.G. Hiller, P.M. Wrench, A.W. Larkum, B.R. Green, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 15244–15248.
[32] W. Nitschke, A.W. Rutherford, Trends Biochem. Sci. 16 (1991) 241–245.
[33] P. Heathcote, P.K. Fyfe, M.R. Jones, Trends Biochem. Sci. 27 (2002) 79–87.
[34] O.B. Ptitsyn, J. Mol. Biol. 278 (1998) 655–666.
[35] G. Rocap, F.W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N.A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W.R. Hess, Z.I. Johnson, M. Land, D. Lindell, A.F. Post, W. Regala, M. Shah, S.L. Shaw, C. Steglich, M.B. Sullivan, C.S. Ting, A. Tolonen, E.A. Webb, E.R. Zinser, S.W. Chisholm, Nature 424 (2003) 1042–1047.
[36] U. Ljungberg, T. Henrysson, C.P. Rochester, H.E. Akerlund, B. Andersson, Biochim. Biophys. Acta 849 (1986) 112–120.
[37] S.A. Bustos, M.R. Schaefer, S.S. Golden, J. Bacteriol. 172 (1990) 1998–2004.
[38] M. Alfonso, I. Perewoska, D. Kirilovsky, Plant Physiol. 122 (2000) 505–516.
[39] A.K. Singh, L.M. McIntyre, L.A. Sherman, Plant Physiol. 132 (2003) 1825–1839.
[40] K.A. Martin, J.L. Siefert, S. Yerrapragada, Y. Lu, T.Z. McNeill, P.A. Moreno, G.M. Weinstock, W.R. Widger, G.E. Fox, Photosynth. Res. 75 (2003) 211–221.
[41] D.J. Des Marais, Science 289 (2000) 1703–1705.
[42] W. Martin, T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, D. Penny, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 12246–12251.
[43] C.J. Howe, A.C. Barbrook, V.L. Koumandou, R.E. Nisbet, H.A. Symington, T.F. Wightman, Philos. Trans. R. Soc. Lond., B Biol. Sci. 358 (2003) 99–106 (discussion 106-7).