Review

The evolution of photosystem I in light of phage-encoded reaction centres

Yuval Mazor¹, Ilanit Greenberg¹, Hila Toporik¹, Oded Beja² and Nathan Nelson¹,*

¹Department of Biochemistry and Molecular Biology, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel
²Faculty of Biology, Technion - Israel Institute of Technology, Haifa 32000, Israel

Recent structural determinations and metagenomic studies shed light on the evolution of photosystem I (PSI) from the homodimeric reaction centre of primitive bacteria to plant PSI at the top of the evolutionary development. The evolutionary scenario of over 3.5 billion years reveals an increase in the complexity of PSI. This phenomenon of ever-increasing complexity is common to all evolutionary processes that in their advanced stages are highly dependent on fine-tuning of regulatory processes. On the other hand, the recently discovered virus-encoded PSI complexes contain a minimal number of subunits. This may reflect the unique selection scenarios associated with viral replication. It may be beneficial for future engineering of productive processes to utilize ‘primitive’ complexes that disregard the cellular regulatory processes and to avoid those regulatory constraints when our goal is to divert the process from its original route. In this article, we discuss the evolutionary forces that act on viral reaction centres and the role of the virus-carried photosynthetic genes in the evolution of photosynthesis.

Keywords: photosynthesis; reaction centres; evolution; marine viruses; lateral gene transfer

1. INTRODUCTION

Over 4 billion years ago, our planet was under anaerobic conditions and most of the elements were present in their reduced-valence states [1,2]. Thus, the atmosphere was rich in nitrogen in the form of ammonia (NH₃) or N₂, carbon as CO or CO₂, and oxygen as H₂O [3], and the surface was covered by water, which contained not only minerals but also organic molecules. Some of the components absorbed sunlight, and thus were engaged in photochemical reactions, some of which destabilized the organic material, whereas others were crucial for increasing the repertoire of components. The usage of sunlight for harnessing energy was more complicated and required the action of several elements in concert. For this to happen, the single-molecule photochemical reaction had to evolve, and with no exception evolution increases complexity [4]. Thus, photochemical reaction centres evolved, and today some of them contain up to 50 000 atoms. Remarkably, the quantum yield of the single-molecule ‘reaction centres’ was maintained in some of the large complexes and is close to 1 in the modern photosystem I (PSI).

The initial involvement of proteins in photochemical reaction should have been quite simple because every protein that harbours pigments such as flavin, carotene or chlorophyll has the potential of light harvesting coupled to chemical reaction. One of today’s challenges is to understand how, in a relatively short time, oxygenic photosynthesis could evolve with a complexity comparable to the current process. PSI appeared to have evolved to its present form, which is considerably more complicated than the simple bacterial reaction centres, quite rapidly and we do not know of any obvious intermediate steps in this evolutionary process. The organisms that exist today in anaerobic niches and the DNA present in the ocean in unknown organisms and viruses may give us insights into the onset and evolution of photosynthesis as well as the increase in complexity of the reaction centres that drive this process [5,6].

2. FROM HOMODIMERIC TO HETERODIMERIC STRUCTURE OF PHOTOSYNTHETIC REACTION CENTRES

Evidence derived from molecular biology and recent phylogenetic studies suggest that PSI, which probably emerged over 3.5 billion years ago, preceded the appearance of cyanobacteria and it may have evolved from organisms resembling today’s green and gliding bacteria [7–10]. Gene duplication followed by the evolution of a heterodimeric structure may be the most crucial step in the evolution of advanced reaction centres [11,12]. Cloning of the gene encoding the main subunit of Chlorobium RC together with polypeptide sequences
of the large subunit revealed that only one gene encodes the dimer of subunit I [7]. A year later, similar organization was reported for *Heliobacillus*, which is quite distant from the Chlorobiaceae [8]. Recently, another example of homodimeric reaction centre was reported in a novel kind of bacteria representing a new phylum [13]. Currently, a homodimeric, PSI-like reaction centre has been discovered in three bacterial phyla: Chlorobi, heliobacteria (Firmicutes) and Acidobacteria. Do those reaction centres represent the primordial photosynthetic reaction centres? What is their relation to the current PSI present in cyanobacteria and green plants? Where does the ‘original’, type I, PSI-like heterodimer hide on our planet? To what extent do the genes encoding those reaction centres represent the original genes that were present in evolutionary junctions occurring 1.5 or even over 3.5 billion years ago?

If simpler, homodimeric, reaction centres preceded the more complex heterodimers of both the type I and type II centres, then the present day reaction centres present in Chlorobi, heliobacteria and Acidobacteria can be perceived as a living fossil of the ancestral reaction centre. However, the extremely long evolutionary paths that each homodimeric gene endured leave very little guidance in terms of amino acid conservation. As shown in figure 1, the amino acid sequences of those reaction centres are poorly preserved, and the conservation among them is similar to that obtained by comparison with PsaA or PsaB of cyanobacteria and higher plants. Among the available sequences present in GenBank, photosystem I P700 chlorophyll a apoprotein A2 of *Chromera velia* [15] seems to fit the bill of having poor enough similarity to the homodimeric reaction centres and the current heterodimeric reaction centres in cyanobacteria and plants. In fact, PsaA1 is more closely related to the bacterial homodimeric reaction centre protein than to its PsaA relatives (figure 1). This demonstrates that protein phylogenetic relationships can deviate from the phylogenetic relationships between organisms. However, it is not clear whether this sequence is related at all to the primordial heterodimeric reaction centre. Where can we find genes encoding the most primitive heterodimeric reaction centres? It seems that metagenomic studies conducted in the oceans and in other aquatic environments may provide the clue for this missing link in the evolution of photosynthetic reaction centres.

### 3. OCEANIC OXYGENIC PHOTOSYNTHESIS

Oceans cover almost 70 per cent of the surface of Earth. Within the ocean, cyanobacteria and other photosynthetic bacteria [16] can be found in significant numbers at depths approaching 200 m; hence, this group’s contribution to the Earths’ carbon, nitrogen and oxygen cycles is very significant. Most of the oceanic cyanobacterial species carry out ‘regular’ oxygenic photosynthesis. However, the genome sequence of UCYN-A, presently an unculturable cyanobacterium, revealed that this species lost all of its photosystem II (PSII) genes [17]. The existence of a modern PSI complex in UCYN-A probably means that its PSII genes were lost and this limits its importance concerning the history of photosynthesis. It does, however, demonstrate the kind of diversity that can be found in the oceans.

A significant portion of this diversity can be found in the viruses that infect marine microbes [6,18]. Oceanic bacteriophages outnumber ocean bacteria by 10 to 1 [18]. However, it was found that the numbers of cyanophages capable of infecting the two most abundant marine cyanobacterial families, *Prochlorococcus* and *Synechococcus*, are significantly outnumbered by their bacterial hosts [19]. Given the much faster evolutionary rates seen in viruses in general and in marine phages in particular, it is clear that a major portion of the genetic diversity on Earth will be found in the genomes of these phages.

It has been almost a decade since the surprising discovery of a phage-encoded D1 protein [20]. Since this initial discovery, phage-encoded photosynthetic
genes were found to be both abundant and diverse [21]. It is clear now that a very large portion of the cyanophages carry a D1 gene, and many contain additional genes involved in various aspect of the photosynthetic and respiratory electron transport chain [6]. Cyanophage D1 and D2 genes can be readily distinguished from their cyanobacterial homologues, suggesting that they are subjected to distinct evolutionary forces [22]. Furthermore, phages of the Myoviridae family can infect both Prochlorococcus and Synechococcus species and thus can potentially be very effective agents of horizontal gene transfer [19]. This large pool of photosynthetic genes may be a major player in the evolution of photosynthesis.

In addition to the PSII genes, a large operon containing a minimal PSI complex was found on a marine virus [21]. Very much like the PSII genes, the phage PSI genes differ considerably from their current cyanobacterial homologues.

If we look at the amino acid sequence of the phage PsaA, two facts are readily apparent. First, sequence divergence between the phage and its cyanobacterial hosts is considerable [6,21,23]. Second, if we examine the Prochlorococcus or Synechococcus consensus sequence at the positions where such a consensus exists, the phage sequence is a mix of both Prochlorococcus and Synechococcus sequences (figure 2). The relatively high divergence between the phage PSI sequence and its cyanobacterial counterparts enables the detection of this unique situation. A total of sixty-six positions unique to Prochlorococcus, Synechococcus or the phage PsaA proteins can be detected (figure 2). This variability can stem from very high rates of evolutionary change or, more probably, represent recombination between the Prochlorococcus and Synechococcus consensus sequences within the phage population [22].

These viral genomes thus offer a unique opportunity to understand some of the functional characteristics and evolution of PSI. One of the most noticeable differences between these two families of cyanobacteria is the divinyl chlorophyll, which can only be found in Prochlorococcus. The variable positions within the phage PSI genes may be related to this very unusual functional requirement, the ability to function with two different types of chlorophyll molecules. Synechocystis PCC 6803 strains designed to express divinyl chlorophyll are light-sensitive, and their sensitivity can be partially rescued by expressing a D1 gene containing some Prochlorococcus-related mutations [24,25]. The ability to adapt to divinyl chlorophyll may be a major evolutionary force in the ocean.

The larger form of the phage PSI operon contains seven genes, which constitute the smallest fully sequenced PSI complex. However, an even more compact organization of PSI genes was uncovered on a different marine virus. In this new operon, which is still not fully sequenced, only four genes (PsaD, PsaG, PsaA and PsaB) are found [23]. According to our understanding, the protein products of these genes may be sufficient to assemble a functional PSI complex. Further studies are required to establish the functionality of these small PSI complexes. However, they seem to represent a much simpler and possibly more ancient form of PSI [2].

The most noticeable feature of the longer phage operon is the PsaJF fusion protein, in which the N-terminus of PsaF is truncated [21] (figure 3). PsaJ and PsaF remained separated throughout the
evolution of the PSI complex in both plants and cyanobacteria. The PsaF subunit of plant PSI contains a positive N-terminal loop that is absent from the cyanobacteria complex [26,27]. This loop was shown to promote the fast phase of the electron transfer reaction between PSI and its electron donors in algae [28]. However, the role of the entire N-terminal domain of PsaF is not as well defined. This domain, together with a PsaA loop (624–633 in the Synechococcus elongatus structure [29]), are the two most prominent features of the rather flat luminal side of PSI. These two structural features have been proposed to coordinate the interaction between cytochrome (cyt) c₆/plastocyanin (PC) and P700 during electron transfer, and both are missing from the phage PSI complex [21](figure 3). In cyanobacteria, the role of PsaF N-terminus appeared to be rather minor since a PsaF deletion strain of Synechocystis sp. PCC6803 is capable of photoautotrophic growth, and a phage mimetic PsaJF fusion in Synechocystis does not confer any growth phenotype [30] (Y. Mazor, H. Toporick, N. Nelson 2010, unpublished observation).

Interestingly, the deletion of the PsaA 624–633 loop is not unique to the phage operon and is also present in all examined Prochlorococcus sequences. The physiological role of this deletion is less defined, but its proximity to the hydrophobic binding site of Table 1. Variability in the isoelectric points (PIs) of cyanobacterial soluble cytochromes. All the small soluble cytochrome sequences from the indicated strains were retrieved from GenBank, and PI values were calculated using the built-in annotation tools in the VECTOR NTI suite. We created a multiple sequence alignment from all the collected sequences and found that all cytM sequences form a well-supported clade within this alignment. The calculated PI for plastocyanin (PC) appears in the first column as a reference. Darker shading indicates higher PI values.
PC/cyt c₆ created by the surface loops of PsA and PsB may point to its significance in the interaction between cyt c₆ and PSI.

The main structural changes that we detect on the phage complex are both located on the luminal side of the complex, and we suggest that both of these changes facilitate electron transfer from a more diverse set of soluble electron donors [31]. The finer details of the interaction between PSI and cyt c₆/PC can be quite variable and were found to be affected by electrostatic interactions between these two moieties [32]. A rough proxy for these electrostatic forces is the isoelectric point (PI) of the protein, which is easily calculated.

Between different species, large variation in the PI of cyt c₆ and PC can be seen; however, as both PC and cyt c₆ donate electrons to the same complexes, within the same species, they both have similar PI. This rule of thumb is far from absolute. In table 1, we show the calculated PI for all the PC and small cytochrome sequences from the set of fully sequenced marine Prochlorococcus and Synechococcus species, as well as from several other cyanobacterial species for reference. The calculated values differ somewhat from experimental measurements where such measurements are available (around six for both cyt c₆ and PC from Synechocystis sp. PCC 6803) [33] but the deviation is quite constant and uniform. Therefore, these values can be used to compare PIs among different proteins.

As can be seen in table 1, there is a considerable variability in the number of genes among species. In some cases only the gene for PC can be detected, and in other cases as many as four genes for different cyt c₆ are seen. The last column in table 1 lists the PI for cytM, a less well-understood cytochrome, which is present in most of the examined species. More interestingly, in several Prochlorococcus strains the cyt c₆ gene appears to have been lost, and where a cyt c₆ gene is present the calculated PIs of PC and of cyt c₆ are widely different. In Synechocystis sp. PCC 6803, these small, soluble proteins can be differentially regulated according to the iron availability in the growth media [34]. However, the different PIs that we detect between these proteins suggest that they are not simply exchanged in response to iron availability but also in response to other conditions such as light intensity or stress.

It can be concluded that the functional requirement involved in the evolution of protein complexes can lead to a pattern that does not follow the phylogenetic patterns of species over long, evolutionary, time scales [35]. Essentially phylogenetic trees are calculated according to current sequences of proteins that function in diverse biochemical and cell biology reactions. Each of them may evolve at different rates. Protein complexes contain several gene products. In the case of reaction centres, they may also have spatial information that holds their various pigments in check. Therefore, a combined approach based on structural, biochemical and bioinformatics analysis is required to fully appreciate the evolution of these complexes.

REFERENCES


16 Beja, O., Suzuki, M. T., Heidelberg, J. F., Nelson, W. C., Preston, C. M., Hamada, T., Eisen, J. A., Fraser, C. M. &


Hippler, M., Drepper, F., Rochaix, J. D. & Muhlenhoff, U. 1999 Insertion of the N-terminal part of Psaf from Ohlamydomonas reinhardtii into photosystem I from Synechococcus elongatus enables efficient binding of algal plastocyanin and cytochrome c6. J. Biol. Chem. 274, 4180–4188. (doi:10.1074/jbc.274.7.4180)


