

Article

Genes Involved in the Endoplasmic Reticulum *N*-Glycosylation Pathway of the Red Microalga *Porphyridium* sp.: A Bioinformatic Study

Oshrat Levy-Ontman^{1,2,†,*}, Merav Fisher^{1,†}, Yoram Shotland², Yacob Weinstein³, Yoram Tekoah^{1,‡} and Shoshana Malis Arad¹

¹ Department of Biotechnology, Rager Ave., Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel; E-Mails: meravfish@gmail.com (M.F.); yoram.tekoah@protalix.com (Y.T.); arad@bgu.ac.il (S.M.A.)

² Department of Chemical Engineering, Sami Shamoon College of Engineering, Basel/Bialik sts., Beer-Sheva 8410001, Israel; E-Mail: yshotlan@sce.ac.il

³ Department of Microbiology and Immunology, Rager Ave., Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel; E-Mail: yacob@bgu.ac.il

† These authors contributed equally to this work.

‡ Current address: Protalix Biotherapeutics, 2 Snunit st., Carmiel 2161401, Israel.

* Author to whom correspondence should be addressed; E-Mail: oshrale@sce.ac.il; Tel.: +972-8-647-5732; Fax: +972-8-647-5654.

Received: 25 November 2013; in revised form: 13 January 2014 / Accepted: 23 January 2014 / Published: 7 February 2014

Abstract: *N*-glycosylation is one of the most important post-translational modifications that influence protein polymorphism, including protein structures and their functions. Although this important biological process has been extensively studied in mammals, only limited knowledge exists regarding glycosylation in algae. The current research is focused on the red microalga *Porphyridium* sp., which is a potentially valuable source for various applications, such as skin therapy, food, and pharmaceuticals. The enzymes involved in the biosynthesis and processing of *N*-glycans remain undefined in this species, and the mechanism(s) of their genetic regulation is completely unknown. In this study, we describe our pioneering attempt to understand the endoplasmic reticulum *N*-Glycosylation pathway in *Porphyridium* sp., using a bioinformatic approach. Homology searches, based on sequence similarities with genes encoding proteins involved in the ER *N*-glycosylation

pathway (including their conserved parts) were conducted using the TBLASTN function on the algae DNA scaffold contigs database. This approach led to the identification of 24 encoded-genes implicated with the ER *N*-glycosylation pathway in *Porphyridium* sp. Homologs were found for almost all known *N*-glycosylation protein sequences in the ER pathway of *Porphyridium* sp.; thus, suggesting that the ER-pathway is conserved; as it is in other organisms (animals, plants, yeasts, *etc.*).

Keywords: bioinformatics; contigs; microalgae; *N*-glycosylation; *Porphyridium* sp.; red algae

Abbreviations: ER, Endoplasmic reticulum, SCF, Scaffold; Annotation and abbreviations of proteins involved in *N*-glycosylation is detailed in Table S1.

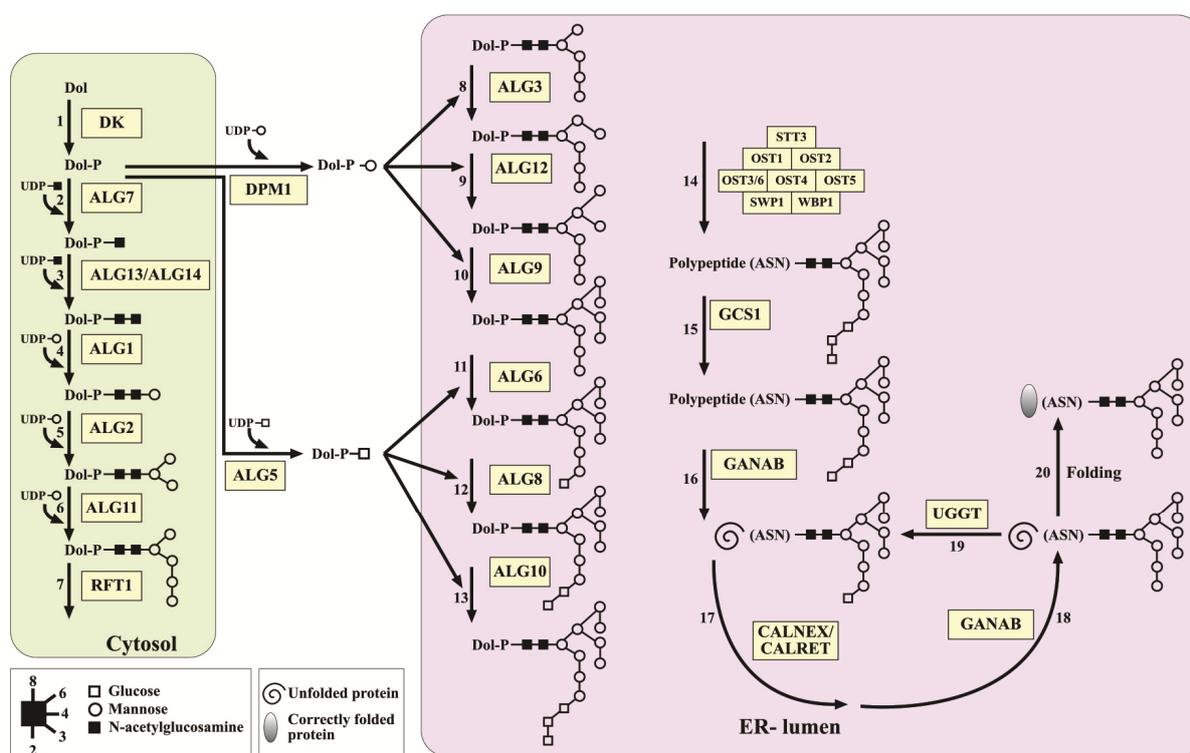
1. Introduction

Glycosylation is one of the most fundamental post-translational protein modifications in eukaryotes. The sugars that are added to the protein during this process affect the physicochemical properties and polymorphism of proteins- e.g., their stabilization, protection, targeting, and direct activity [1–4].

Protein *N*-glycosylation in eukaryotes takes place along the endoplasmic reticulum (ER)-Golgi pathway, beginning with the production of a precursor. In this process, a $\text{Man}_5\text{GlcNAc}_2$ core-oligosaccharide attached to the lipid carrier dolichyl pyrophosphate ($\text{Man}_5\text{GlcNAc}_2\text{-PP-dolichol}$ lipid-linked precursor intermediate) is assembled by the stepwise addition of monosaccharides to dolichol pyrophosphate on the cytosolic side of the ER [5] (stages 1–7, Figure 1). This intermediate precursor is then extended in the lumen of the ER until a $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-dolichol}$ lipid-linked precursor is completed [5] (stages 8–13, Figure 1). Later, the oligosaccharide $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ is transferred from the dolichol phosphate to the growing, nascent polypeptide chain via the nitrogen atom of an asparagine amino acid residue by Dolichyldiphosphoryl oligosaccharide-protein, or oligosaccharyltransferase (OST) in the lumen of the rough ER (stage 14, Figure 1) [6,7]. The asparagine must be part of the consensus sequence, Asparagine-X-Serine/Threonine (where X is any amino acid except Proline) [8]. Following the oligosaccharide transfer, membrane bound Mannosyl-oligosaccharide glucosidase I (GCS1) [9] and the soluble Alpha 1,3-glucosidase II (GANAB), which is composed of two subunits α and β [10], remove the α 1,2-glucose and α 1,3-glucose residues from the oligosaccharide, respectively, generating monoglucosylated *N*-glycan $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ (stages 15–16, Figure 1). The monoglucosylated glycan is required for productive cycle interactions with the ER-resident chaperones calnexin (CALNEX) or/and calreticulin (CALRET) [11–13]. These interactions, associated with the CALNEX/CALRET cycle, facilitate folding of newly-formed glycoproteins in the ER (stages 17–20, Figure 1) [14,15]. If GANAB trims the last glucose residue it prevents further association with CALNEX/CALRET, allowing correctly folded proteins to proceed to the secretory pathway. In contrast, incorrectly folded glycoproteins can be reglucosylated by UDP-glucose:glycoprotein glucosyltransferase (UGGT) (stage 19, Figure 1) to ensure its interaction with the ER-resident chaperones (CALNEX/CALRET), allowing another cycle of CALNEX/CALRET interaction. This enables unfolded substrates to go through multiple rounds of

interaction with the chaperons of the cycle until the native conformation is reached, when recognition by GANAB (but no longer by UGGT) allows exit from the cycle and the ER (stage 20, Figure 1).

Figure 1. Schematic representation of the pathway for *N*-linked glycoprotein biosynthesis. In this process, a $\text{Man}_5\text{GlcNAc}_2\text{-PP-dolichol}$ lipid-linked precursor intermediate is assembled. This intermediate is then extended in the lumen of the ER until $\text{Glc}_3\text{Man}_5\text{GlcNAc}_2\text{-PP-dolichol}$ lipid-linked precursor is completed. The $\text{Glc}_3\text{Man}_5\text{GlcNAc}_2$ oligosaccharide is then transferred onto the target nascent polypeptide to form a protein precursor. This protein precursor is then deglycosylated/reglycosylated to ensure quality control of the neosynthesized protein. The proteins involved are listed in the yellow squares and Dol annotated in Table S1.



Following the trimming of all three glucose from the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ core oligosaccharide attached to the polypeptide, ER, and Golgi $\alpha 1,2$ mannosidases (ManI) collectively cleave the $\alpha 1,2$ -linked mannose residues from the oligosaccharide precursor and, thus, provide the substrates required for the formation of hybrid and complex glycans in the Golgi of eukaryotes cells. The ER members of $\alpha 1,2$ mannosidase in various organisms play an important role in targeting misfolded glycoproteins for degradation by proteasomes [16]. ER *N*-glycosylation events are crucial for the proper folding of the secreted proteins and are highly conserved in the eukaryotes investigated thus far [17].

To date, *N*-glycosylation patterns and *N*-glycan structures have been studied mainly in mammals, insects, yeasts, and plants [18], with seaweeds and microalgae receiving very little attention. Among the scant research conducted on glycosylation in microalgae, most studies on *N*-glycan structures were performed on green microalgae (Chlorophyta) [18–25]. The studies generally revealed the presence of glycans similar to those found in other more heavily researched species, mainly oligomannosides or mature *N*-glycans having a xylose core residue. Whereas, two reports concerning the investigation

of *N*-glycosylation of the green microalga *Chlamydomonas reinhardtii* describe two different findings [24,25]; Mathieu-Rivet *et al.* 2013, revealed that the predominant *N*-glycans attached to *Chlamydomonas reinhardtii* endogenous soluble and membrane proteins, are of oligomannose type [25]. In addition, minor *N*-linked glycans were identified as being composed of mannose, methylated mannose and xylose residues [25]. However, Mamedov and Yusibov, 2011 [24], reported that the *N*-linked oligosaccharides released from total extracts of *Chlamydomonas reinhardtii* carried mammalian-like sialylated *N*-linked oligosaccharides [24]. It is also noteworthy that the *N*-glycosylation pathway of the diatom *Phaeodactylum tricornutum* photosynthetic microalga was investigated, demonstrating that *Phaeodactylum tricornutum* proteins carry mainly high mannose type *N*-glycans (Man-5 to Man-9) and a minor glycan population carrying paucimannose type [26]. It was also suggested the *Phaeodactylum tricornutum* possesses the ER machinery required for glycoprotein quality control that is normally found in other eukaryotes [26].

Red microalgae seem to have glycosylation pathways that are different from those of other known organisms, as was been concluded in a recent study by Levy-Ontman *et al.* 2011 [27]. This study described, for the first time, the structural determination of the *N*-linked glycans in a 66-kDa glycoprotein, which is a part of the unique sulfated complex cell wall of polysaccharide from the red microalga *Porphyridium* sp. *N*-glycans were found to be of the high-mannose type (8–9 residues), with unique modifications that included two non-characteristic xylose residues (one attached to the core and the other to the non-reducing end) and an additional methylation modification on the sixth carbon of three mannose residues attached to the chitobiose core.

The work presented herein is focused on the red microalga *Porphyridium* sp. This organism is a photosynthetic unicell found in marine environments. One of the characteristics of red microalgae is their cell-wall that is composed of sulfated polysaccharide capsules. During growth, the external parts of the polysaccharides are released to the surrounding aqueous medium where they accumulate, increasing the medium's viscosity [28–30]. These polysaccharides have been shown to possess a variety of bioactivities, with potential applications in different industries, e.g., cosmetics, pharmaceuticals, and nutrition [31,32]. Our group has undertaken the challenge of exploiting the potential of red microalgae sulfated polysaccharides for biotechnological applications and the development of large-scale production technologies [31–36].

In recent years, a great deal of scientific work is being directed at creating a novel assortment of pharmaceutical products using algae as cell factories [37–40]. However, although they are well suited for the large-scale production of recombinant proteins, the full potential of algae as protein-producing cell factories is far from being fulfilled [40–45]. Large-scale cultivation of algae for the production of therapeutic proteins has several advantages. Algae are simple to grow, and have relatively fast growth rate. In addition, algae are able to use sunlight as an energy source, hence they are energy efficient, have a minimal negative impact on the environment, and are easy to collect and purify. To date, the use of red microalgae as cell factories for therapeutic proteins has been limited by the lack of molecular genetics tools. A stable chloroplast transformation system [46] and a nuclear transformation system have been developed for *Porphyridium* sp. [47], the latter of which has paved the way for the expression of foreign genes in red algae, which has far-reaching biotechnological implications. However, the application of this platform cannot reach its full potential without the study of glycosylation. The differences in glycosylation patterns between different organisms may have

influence on the activity of the recombinant protein or may influence its immunogenicity. It is therefore most important to evaluate the glycans attached to any recombinant protein expressed in any specific system.

There is very limited knowledge about red algal genomes; the sequencing of genomes of the unicellular red microalgae extremophiles, Cyanidiophyceae *Cyanidioschyzon merolae* and *Galdieria sulphuraria* have been completed [48,49]. In addition, only recently, the nuclear genome sequence of *Porphyridium purpureum* (referred to as *Porphyridium cruentum*) has been completed [50] and is the first genome sequence from a mesophilic, unicellular red alga that has been reported thus far. An analysis of the *Porphyridium purpureum* genome suggests that ancestral lineages of red algae acted as mediators of horizontal gene transfer between prokaryotes and photosynthetic eukaryotes, thereby significantly enriching genomes across the tree of photosynthetic life [50]. Moreover, based on the genome database it was suggested that red algae mediate cyanobacterial gene transfer into chromalveolates [51]. In addition, our group have made significant progress in the field of red microalgal genomics by the establishment of EST databases of two species of red microalgae, *Porphyridium* sp. (sea water) and *Dixoniella grisea* (brackish water) [32,52]. Non-normalized unidirectional cDNA libraries constructed from *Porphyridium* sp. grown under various physiological conditions generated 7210 expressed sequence tags (ESTs), which gave 2062 non-redundant sequences, containing 635 contigs and 1427 singlets [32]. Some genes derived from the EST database were analyzed and compared to other ortholog genes that exist in other organisms [32,52,53].

In this paper we describe our attempt to better understand the *N*-glycosylation mechanism that takes place in the ER within the red microalga *Porphyridium* sp. Our DNA scaffold (SCF) database of *Porphyridium* sp. was used to search for sequence similarity to algae gene products potentially involved in *N*-glycosylation pathways. Such a study can provide a basis for understanding *N*-glycosylation pathways in red microalgae, and lay the foundations for future gene cloning and characterization.

2. Results and Discussion

2.1. DNA Sequencing of *Porphyridium* sp.

DNA was divided into sections of 330 bases (on average) and 38 bases were sequenced from each end of each section (Pair-end). The total reads identified were 38,537,782 sections, constituted of 1,464,435,716 bases.

Assembly of all reads was completed using VELVET; the best assembly results of the reads was obtained with a hash (or k-mer) of 23. Longer k-mers bestow more specificity (*i.e.*, less spurious overlaps), but lower coverage. A k-mer of 23 indicates more specificity on account of coverage. Nevertheless, we were able to obtain an impressive length of contigs, with N50 of 41,031 bp for the SCF (Table 1). The assembly results were also validated (Section 3.2.3). There were two types of assemblies (Table 1): (1) contigs containing sequences of the DNA reads only; (2) Scaffold (SCF), which consists of close contigs that are adjacent to each other using a number of unknown bases (Ns). Some contigs were joined by stretches of “N” when VELVET, through the paired end information, identifies a link between contigs but cannot determine the sequence. The length of the “N” was calculated from the average insert length. The quality of the sequencing results was high: 96.1% of all

sections were mapped to contigs, while 83.7% of them were used for the contigs database and 89.1% were used to form the SCF database. The adjacent contigs were successfully attached to each other. Each base was sequenced, on average, approximately 70 times. The sequencing results indicated that the genome size of the algae (including its chloroplast and mitochondria) is approximately 20 MB. The genome size is in accordance with former results [50] and again demonstrates that red algal genomes are reduced in comparison to mammalian genomes [48–50,54]. Comparison of the genome size of *Porphyridium* sp. found in this study to that of some other previously reported microalgal genomes was found to be similar; e.g., the diatom *Thalassiosira pseudonana* (genome size 32.4 MB), *Phaeodactylum tricornutum* (genome size 27.4 MB), the green algae *Ostreococcus tauri* (genome size 12.6 MB), *Ostreococcus lucimarinus* (genome size 13.2 MB), and *Micromonas pussila* (genome size 21 MB) [55].

Table 1. DNA sequencing results using high-throughput technology by Solexa, produced from the red microalga *Porphyridium* sp.

Assembly	CONTIG	SCF
Total length	18,613,981	18,925,597
Number of contigs	9,653	3,002
N50	4,218	41,031
Undetermined base	0	280,103
Average length of contigs	1,928	6,304
Maximum size	37,208	204,033
Reads mapped	37,023,682	37,034,742
% of all reads	96.1	96.1
Reads paired	30,970,611	32,980,567
% of all mapped	83.7	89.1

2.2. Identifying N-Glycosylation Protein-Encoding Genes in *Porphyridium* sp.

Homology searches based on sequence similarities with genes encoding proteins involved in ER N-glycosylation pathway were conducted by using the TBLASTN function on the algae DNA SCF contigs engine database (Section 3.3.3). TBLASTN was selected because there is very little evidence for introns in *Porphyridium* sp. (based on our in house DNA sequence, unpublished results). In order to identify *Porphyridium* sp. gene products involved in the ER N-glycosylation pathway, homology-based searches of *Saccharomyces cerevisiae* (*S. cerevisiae*) N-glycosylation genes against the *Porphyridium* sp. DNA SCF contigs engine were conducted. The identification of the calreticulin encoding-gene in *Porphyridium* sp. was based on homology-based searches against the *Chlamydomonas reinhardtii* ortholog gene, and that of UGGT encoding-gene was based against *Galdieria sulphuraria* ortholog gene. Searches for encoding-genes of OST 3/6/5 and SWP1 were based on homology-based searches against the ortholog genes in *S. cerevisiae*, *Chlamydomonas reinhardtii*, *Galdieria sulphuraria*, *Cyanidioschyzon merolae* and *Arabidopsis thaliana*. Homologs were found for almost all algal N-glycosylation protein sequences in the ER pathway because all of them exhibit similarity values of above 43% under good sequence coverage calculated as compared to the entire gene sequences (above 60%), with one exception (sequence coverage of UGGT was

only 26%) (Table 2). In addition the conserved domains that are essential for enzymatic activity were identified in all our ER *N*-glycosylation pathway homologues (Table 3). The homology was also verified by the GO values that were obtained by Blast2go program (data not shown). The predicted amino acid sequence for each gene was identified (Table S2).

All the genes encoding proteins involved in the biosynthesis of dolichol pyrophosphate-linked oligosaccharide on the cytosolic side of the ER were identified in the genome of *Porphyridium* sp. (Table 2). The sequences of these predicted proteins (Table S2) are highly similar to the corresponding asparagine-linked glycosylation (ALG) orthologs of *S. cerevisiae* (above 47% similarity, Table 2). Putative transferases, which are able to catalyze the formation of dolichol-activated mannose and glucose required for the biosynthetic steps arising in the ER lumen, were also found (dolichol-phosphate mannosyltransferase (DPM1), dolichyl-phosphate beta-glucosyltransferase (ALG5), Table 2). Almost all the genes involving the ER lumen biosynthesis exhibited above 43% similarity to ortholog genes. However, the subunits OST3/6/5 and SWP1 of the OST, did not exhibit homology to the related *S. cerevisiae*, *Chlamydomonas reinhardtii*, *Galdieria sulphuraria*, *Cyanidioschyzon merolae*, and *Arabidopsis thaliana* subunits. The STT3 protein (that was also found in the *Porphyridium* sp. genome), accounts alone for OST activity in some organisms [56]. In complex organisms the OST works as a multi-protein complex, built as an extension of the STT3 core [57]. Each subunit in this complex has its role in the fine-tune glycosylation process. For example, OST1 acts as a chaperon to promote glycosylation [58–60]; OST3/6 exhibits oxidoreductase activity and binds to specific proteins [61]; and, OST4 was found to be involved in OST3/6 attachment to the OST complex [62]. However, in some organisms, it is known that some of the subunits are not crucial to the OST enzyme function [56,57]. For example, the OST of some protists is composed only from WBP1, STT3, OST2, OST1 [56], or only from STT3 homologs [63,64], bacterial and archeal OSTs are composed only from STT3 homologs [65–68]. Indeed, based on these reports it is possible that the OST enzyme of *Porphyridium* sp. functions without OST5, OST3/6, and SWP1 subunits. It is also important to note that we identified two STT3 copies in the *Porphyridium* sp. genome (Table 2). These multi-spanned sequences have similarity of 66% and 67% respectively with the *S. Cervisea* STT3 subunit (Table 2). It is known that some eukaryotes, bacteria and archea extend their glycosylation ability by the duplication of the STT3 gene and diversification of STT3 specificity [56].

Genes encoding for proteins involved in the quality control of proteins in the ER were also found in the *Porphyridium* sp. genome (Tables 2 and 3). Indeed, Glucosidase I, as well as the subunits of α and β of glucosidase II, were identified (Tables 2 and 3). A putative UGGT and the two chaperons: calnexin and calreticulin, three molecules ensuring the quality control of the glycoproteins in the ER, also exhibit high similarity to ortholog genes (above 48% similarity, Table 2). In addition, three homologs for ManI enzyme were found, all belonging to glycosylhydrolase family 47. The three homologs of ManI were analyzed with InterProScan. The results of this analysis clearly suggest that ManIa (Table S2) is an ER enzyme while ManIb and ManIc are Golgi enzymes, harboring signal peptides at their *N* termini targeting them to the Golgi (Tables 2 and 3). This ManIa gene, that is known to be conserved throughout eukaryotic evolution [69–71], probably plays an important role in targeting misfolded glycoproteins for degradation by proteasomes in *Porphyridium* sp.

Taken together, these results suggest that the ER *N*-glycosylation pathway is conserved in *Porphyridium* sp., as in other organisms (animals, plants, yeasts, etc.).

Table 2. Similarity of *Porphyridium* sp. proteins involved in ER *N*-glycosylation to ortholog proteins of *S. cerevisiae/Chlamydomonas reinhardtii/Galdieria sulphuraria*. The similarity analysis was performed by TBLASTN algorithm. * The coverage calculated as compared to the entire gene sequence.

Abbreviation	Enzyme/Protein	Min. <i>E</i> Value	Mean Similarity Percentage	Coverage *
DK	Dolichol kinase	5.90E-20	55	62
ALG7	UDP- <i>N</i> -acetylglucosamine—dolichyl-phosphate <i>N</i> -acetylglucosaminephosphotransferase	7.09E-91	62	91
ALG13	UDP-GlcNAc:dolichyl-pyrophosphoryl-GlcNAc GlcNAc transferase	4.56E-32	61	93
ALG14	UDP-GlcNAc:dolichyl-pyrophosphoryl-GlcNAc GlcNAc transferase	1.54E-36	65	69
ALG1	Chitobiosyldiphosphodolichol beta-mannosyltransferase	3.43E-68	47	92
ALG2	Glycolipid 3-alpha-mannosyltransferase	1.24E-77	54	91
ALG11	GDP-mannose:glycolipid 1,2-alpha-D-mannosyltransferase	1.40E-95	63	83
ALG3	Dolichol phosphomannose-oligosaccharide-lipid mannosyltransferase	1.38E-90	64	87
ALG9	Dolichol phosphomannose-oligosaccharide-lipid mannosyltransferase	9.39E-102	55	87
ALG12	Alpha-1,6-mannosyltransferase	1.48E-73	57	86
ALG6	Alpha-1,3-glucosyltransferase	9.10E-80	60	70
ALG8	Alpha-1,3-glucosyltransferase	5.89E-68	51	83
ALG10	Alpha-1,2 glucosyltransferase	8.49E-28	43	93
ALG5	Dolichyl-phosphate beta-glucosyltransferase	2.15E-69	69	70
DPM1	Dolichol-phosphate mannosyltransferase	9.25E-80	71	99
RFT1	Flippase	4.69E-31	47	79
STT3		0	66	98
		0	67	99
OST1		6.00E-39	47	70
OST2		2.45E-27	72	80
OST3/6	OST-dolichyldiphosphoryloligosaccharide-protein	-	-	-
OST4		-	-	-
OST5		-	-	-
WBP1		1.45E-62	53	83
SWP1		-	-	-
GCS1	Mannosyl-oligosaccharide glucosidase I	7.55E-93	45	93
GANAB	Alpha 1,3-glucosidase II	0	62	71
GANABb	Alpha 1,3-glucosidase II,beta subunit	1.36E-33	43	98
UGGT	UDP-glucose:glycoprotein glucosyltransferase	1.16E-94	67	26
MAN1a		1.41E-82	55	72
MAN1b	Mannosyl-oligosaccharide	5.83E-66	49	78
MAN1c	alpha-1,2-mannosidase	1.56E-62	49	82
CALNEX	Calnexin	3.46E-92	53	88
CALRET	Calreticulin	1.91E-58	48	94

Table 3. Existence of the conserved domain of *Porphyridium* sp. proteins involved in ER *N*-glycosylation. Separately, sequence detection of conserved domains was done directly (Blast2Go versus Interpro scan and then added GO's (Gene ontology)).

Abbreviation	Definition	Domain Identification	Name	E Value
DK	Dolichol kinase	PTHR13205:SF8	transmembrane protein 15	6.70E−29
ALG7	UDP- <i>N</i> -acetylglucosamine—dolichyl-phosphate <i>N</i> -acetylglucosaminephosphotransferase	PF00953	Glyco_transf_4	1.20E−49
ALG13	UDP-GlcNAc:dolichyl-pyrophosphoryl-GlcNAc GlcNAc transferase	PF04101	Glyco_transf_28_C	1.10E−26
ALG14	UDP-GlcNAc:dolichyl-pyrophosphoryl-GlcNAc GlcNAc transferase	PF08660	Alg14	2.00E−72
ALG1	chitobiosyldiphosphodolichol beta-mannosyltransferase	PF00534	Glyco_transf_1	9.80E−11
ALG2	glycolipid 3-alpha-mannosyltransferase	PF00534	Glyco_transf_1	5.40E−32
ALG11	GDP-mannose:glycolipid 1,2-alpha-D-mannosyltransferase	PF00534	Glyco_transf_1	8.30E−23
ALG3	Dolichol phosphomannose-oligosaccharide-lipid mannosyltransferase	PF05208	ALG3	4.10E−143
ALG9	Dolichol phosphomannose-oligosaccharide-lipid mannosyltransferase	PF03901	Glyco_transf_22	8.50E−97
ALG12	Alpha-1,6-mannosyltransferase	PF03901	Glyco_transf_22	1.30E−19
ALG6	Alpha-1,3-glucosyltransferase	PF03155	ALG6_ALG8	1.90E−107
ALG8	Alpha-1,3-glucosyltransferase	PF03155	ALG6_ALG8	3.80E−85
ALG10	Alpha-1,2 glucosyltransferase	PF04922	DIE2_ALG10	1.30E−15
ALG5	Dolichyl-phosphate beta-glucosyltransferase	PF00535	glyco_transf_2	7.50E−23
DPM1	Dolichol-phosphate mannosyltransferase	PF00535	Glyco_transf_2	8.30E−34
	Dol-P-Glc phosphodiesterase	-	-	-
STT3		PF02516	STT3	1.40E−142
		PF02517	STT3	3.30E−133
OST1		PF04597	Ribophorin I	2.50E−19
OST 2		PF02109	DAD	1.20E−41
OST 3/6	Dolichyldiphosphoryl oligosaccharide-protein (OST)	-	-	-
OST 4		PF10215	Ost4	7.80E−06
OST 5		-	-	-
WBP1		PF03345	DDOST_48kD	1.20E−87
SWP1		-	-	-
RFT1	Flippase	PF04506	Rft-1	9.20E−11
GCS1	Mannosyl-oligosaccharide glucosidase I	PF03200	Glyco_Hydro 63	1.50E−87
GANAB	Alpha 1,3-glucosidase II	PF01055	Glyco_Hydro 31	5.90E−282
GANABb	Alpha 1,3-glucosidase II, beta subunit	PTHR12630:SF1	Glucosidase II β subunit	2.70E−28
MAN1a	Mannosyl-oligosaccharide		Mannosyl-oligosaccharide	3.80E−131
MAN1b	alpha-1,2-mannosidase	PTHR11742	alpha-1,2-mannosidase	7.80E−95
MAN1c			related UDP—glucose	1.40E−98
UGGT	UDP-glucose:glycoprotein glucosyltransferase	PTHR11226	glycoprotein: glucosyltransferase	2.30E−168
CALNEX	Calnexin	PF00262	Calreticulin	1.00E−147
CALRET	Calreticulin	PF00262	Calreticulin	2.90E−63

2.3. Bioinformatic Comparative Study of *Porphyridium* sp. Protein Sequences Involved in N-Glycosylation, with Ortholog Sequences of Various Organisms

TBLASTN, a bioinformatics-based similarity tool, was used to compare between protein sequences involving the ER N-glycosylation pathway of *Porphyridium* sp. and ortholog protein sequences of other organisms. The various organisms tested included red microalgae (*Galdieria sulphuraria*, *Cyanidischyzon merolae*), green microalgae (*Chlamydononas reinhardtii*, *Osterococcus lucimarinus*, *Micromonas* sp. RCC229, *Micromonas pusilla*), diatoms (*Phaeodactylum tricornutum*, *Fragilariopsis cylindru*, *Thalassiosira pseudonana*), mammals (*Human*, *Mus musculus*), and the yeast *S. cerevisiae*. It appears that most of the genes involved in the ER N-Glycosylation pathway in *Porphyridium* sp. also exist in other red algae, green algae, diatoms, yeasts, and mammals (Table 4). Most of *Porphyridium* sp. protein sequences that were studied, presented more than 40% similarity to ortholog sequences of various organisms (Figure 2). It is noteworthy that the similarity between *Porphyridium* sp. N-glycosylation protein sequences and ortholog sequences in other red algae is not significantly higher than the similarity found with other organisms. This can be explained by the theory that the red microalga *Porphyridium* sp. is an ancient organism that conserved its N-glycosylation genes. In a previous report, general EST-derived protein sequences of *Porphyridium* sp. were compared to ESTs of other organisms and the best homology was found to be the red microalgae *Cyanidischyzon merolae*, followed by the green plant *Arabidopsis thaliana* [32]. Looking at N-glycosylation genes; some genes that were found in *Porphyridium* sp., were missing in other green and red algae and in diatoms (Table 4). The genomes of several species of green algae, and all the diatoms, lack two enzyme sequences, α -1,2 glucosyltransferase ALG10 and GCS1. It is most likely that the genes are indeed missing from the genomes of these organisms—since their genomes are well understood. Further strengthening for this notion comes from the fact that these genes encode for enzymes which essentially act together—they are responsible for the addition and the removal, respectively, of the third glucose residue found in the ER N-glycans: ALG10 is responsible for the addition of the Glc α (1–2) to the N-glycan and GCS1 is responsible for trimming the Glc α (1–2) residue after the N-glycan is transferred to the nascent protein. This assumption was also verified by comparing between the similarity of the *Porphyridium* sp. STT3 sequence to ortholog genes of organisms that did or did not contain the ALG10 and GCS1 genes. Since the subunit STT3 is accepted as a substrate of the ALG10 enzyme, paucity in glucose residues on the substrate can change the connection to STT3. Indeed, the similarity between *Porphyridium* sp. STTs gene to organisms that were found to have ALG10 and GCS1 was higher compared to organisms that lacked the ALG10 and GCS1, indicated by changes in STT3 subunit affinity.

Based on the strong similarity of *Porphyridium* sp. encoded-genes to ortholog genes of complex eukaryotes, and the resemblance of *Porphyridium* sp. OST complex to that found in lower organisms (including prokaryotes) it appears that the red alga retained genes from both partners, thus, bringing together mutual elements in the red alga genome. Indeed, the N-glycan structures of the cell-wall glycoprotein within the *Porphyridium* sp. polysaccharide were also found to be composed of prokaryote to multicellular organism elements [27].

Table 4. Existence of homological protein sequences involved in *N*-glycosylation of different organisms by searching for the similarity level to the sequences of ortholog proteins in *Porphyridium* sp. ● Homolog sequence exist. Gray squares: homolog was not found.

<i>N</i> -glycosylation protein	Red Microalgae		Diatom					Green Microalgae			Mammals	Yeast	
	<i>Galdieria sulphuraria</i>	<i>Cyanidischyzon merolae</i>	<i>Phaeodactylum tricornutum</i>	<i>Thalassiosira pseudonana</i>	<i>Fragilariopsis cylindrus</i>	<i>Chlamydomonas reinhardtii</i>	<i>Micromonas</i> sp. RCC229	<i>Micromonas pusilla</i>	<i>Osterococcus lucimarinus</i>	<i>Chlorella</i>	Human	<i>Mus musculus</i>	<i>Saccharomyces cerevisiae</i>
DPM1	●	●	●	●	●	●	●	●	●	●	●	●	●
ALG7	●	●	●	●	●	●	●	●	●	●	●	●	●
ALG13	●	●	●	●	●	●	●	●	●	●	●	●	●
ALG14	●	●	●	●	●	●	●	●	●	●	●	●	●
ALG1	●	●	●	●	●	●	●	●	●	●	●	●	●
ALG2	●	●	●	●	●	●	●	●	●	●	●	●	●
ALG11	●	●	●	●	●	●	●	●	●	●	●	●	●
RFT1	●	●	●	●	●	●	●	●	●	●	●	●	●
ALG3	●	●	●	●	●	■	●	●	●	●	●	●	●
ALG9	●	●	●	●	●	■	●	●	●	●	●	●	●
ALG12	●	■	●	●	●	■	●	●	●	●	●	●	●
ALG6	●	●	●	●	●	●	●	■	●	●	●	●	●
ALG8	●	●	●	●	●	●	●	■	●	●	●	●	●
ALG10	●	●	■	■	■	■	■	■	■	●	●	●	●
ALG5	●	●	●	●	●	●	●	●	●	●	●	●	●
OST-STT3	●	●	●	●	●	●	●	●	●	●	●	●	●
GCS1	●	●	■	■	■	■	■	■	■	●	●	●	●
GANAB	●	●	●	●	●	●	●	●	●	●	●	●	●
ER-ManI	●	●	●	●	●	●	●	●	●	●	●	●	●
UGGT	●	●	●	■	■	■	●	●	●	●	●	●	●
CALNEX	●	●	●	●	●	●	●	●	●	●	●	●	●
CALRET	●	■	●	●	■	●	■	■	■	●	●	●	●

The sequence of the conserved part of three *N*-glycosylation enzymes of *Porphyridium* sp. (a result of sequence integration of ALG1, ALG7, and GANAB) was compared to ortholog parts of other organisms—other algae, plant and non-photosynthetic organisms—and a phylogenetic analysis was performed (Figure 3). As expected, it seems that *Porphyridium* sp. *N*-glycosylation enzymes have the same ancestral origin as other red microalgae as was previously presented [32]. In addition, it seems that the red alga has the same ancestral origin as the diatom group. This finding is not surprising; indeed the diatoms, are believed to be derived from a secondary endosymbiotic process that took place around one billion years ago between a red alga and a heterotrophic eukaryote [72]. Green algae and higher plants also show a common ancestral origin (Figure 3). These findings are in line with the idea that the red algae is a sister group to green plants [73,74]. Moreover, according to phylogenetic

analysis (Figure 3) all photosynthetic organisms have the same ancestral origin, supporting the theory that endosymbiosis took place between eukaryotes and a cyanobacterium, giving rise to all photosynthetic organisms [75–78].

Figure 2. Similarity coverage (% positives blast coverage/query) of protein sequences involved in *N*-glycosylation from various organisms to the sequences of the ortholog protein from the alga *Porphyridium* sp. The results were obtained by BLAST 2 seq. Red—Red algae, Purple—diatoms, Green—Green algae, Blue—mammals, Black—yeast.

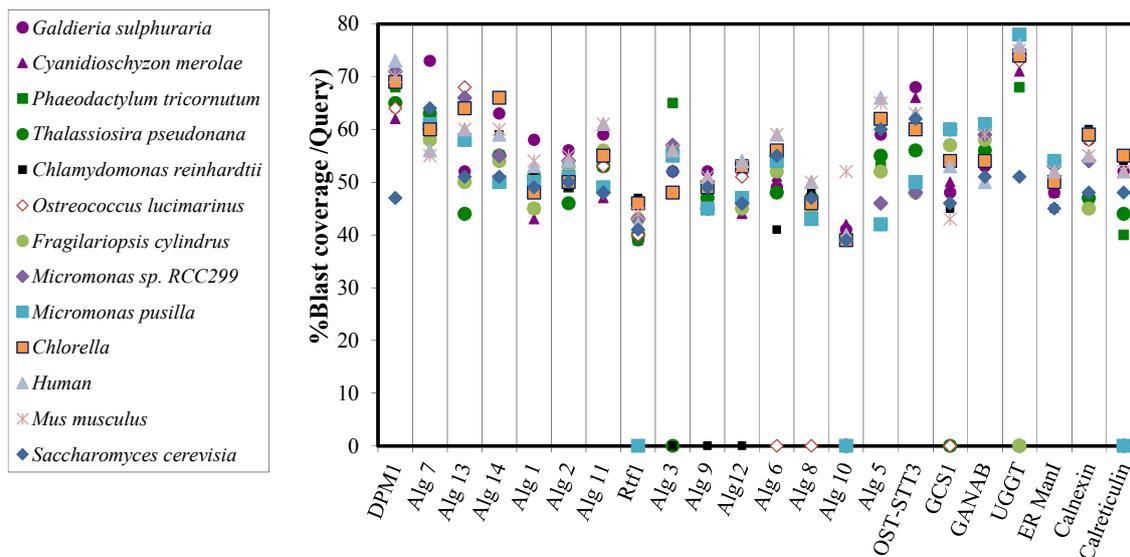
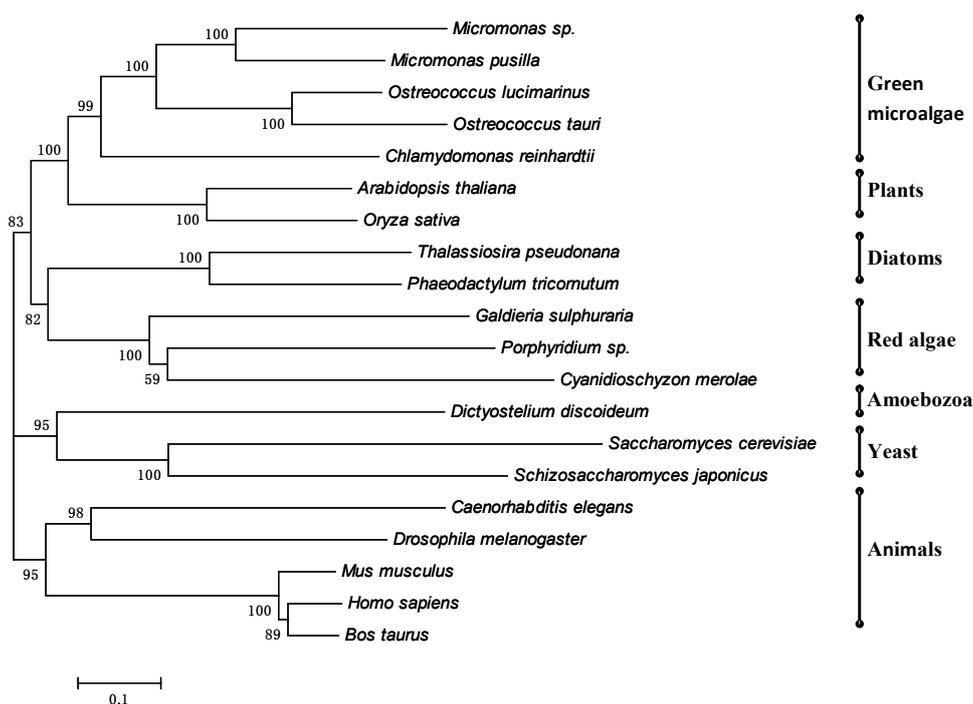


Figure 3. Phylogenetic tree of *N*-glycosylation enzymes (ALG1, ALG7, GANAB) of *Porphyridium* sp. The tree was built by the Neighbor-Joining method with 500 repeated bootstrap tests, using MEGA software, version 4.0. The conserved domain of the related enzyme sequences were tracked by Interproscan (1369 sites).



3. Experimental Section

3.1. *Porphyridium sp.* and Growth Conditions

The microalga *Porphyridium sp.* (UTEX 637), obtained from University of Texas Culture Collection, was grown in 250 mL Erlenmeyer flasks, each containing 100 mL of artificial seawater (ASW) [79]. The algae cultures were grown at a shaking speed of 100 rpm and aerated with sterile air containing 2%–3% CO₂. The growth temperature was maintained at 25 ± 3 °C, and illumination was supplied continuously from above with white fluorescent lamps at a photon flux density of 90 μmol photons m⁻²s⁻¹. Cells were counted with a haemocytometer using a Zeiss light microscope (Carl Zeiss, Oberkochen, Germany).

3.2. Sequencing the DNA of *Porphyridium sp.*—The Solexa Technology

3.2.1. Producing Genomic DNA from the Cells

A 48-h logarithmic-scale cell culture (12–15 × 10⁶ cells/mL) was centrifuged (3000× g, 5 min, 4 °C) and the pellet was washed with acidified water (pH 4) to remove cell-bound polysaccharides and then centrifuged (3000× g, 5 min, 4 °C). The remaining pellet (~1 g) was frozen immediately in liquid-nitrogen and minced into powder with a mortar and pestle. The pellet was incubated with gentle rocking for 2 h at 65 °C. Ten milliliters of CTAB buffer (CTAB—3% w/v), NaCl—1.4 M, EDTA (pH = 8)—20 mM, Tris-HCl (pH = 8)—10 mM, 0.1% β-Mercaptoethanol) were then added to the incubated pellet. The mixture was extracted by mixing it with a 10 mL solution of chloroform: isoamyl alcohol (24: 1 v/v) and then incubating the mixture for 5 min at 25 °C. Later, the extraction mixture was centrifuged (8000× g, 10 min, 4 °C) and the supernatant was collected to a clean vial. The extraction procedure was repeated 3 times. DNA precipitation was obtained by adding isopropyl alcohol (0.75 of the total volume) to the collected supernatant, mixing gently and incubating at –20 °C for 20 min, until a precipitate was formed. The DNA precipitate was collected by centrifuging (15 min, 8000× g, 4 °C), washing in 70% ethanol, drying by vacuum centrifuge and resuspending in 300 μL DDW. To remove residual RNA, the sample was incubated with RNaseA (final concentration 150 μg/mL), for 1.5 h at 37 °C. Afterwards the sample was precipitated again by adding 3 M sodium acetate pH 5.2 (1/10 of the total volume) and cold absolute ethanol (2/3 of the total volume). The sample was centrifuged (10,000× g, 10 min, 4 °C), and the resulting precipitate was washed in 70% ethanol, dried by vacuum centrifuge, and resuspended in 300 μL DDW. DNA concentration was determined by nanodrop and gel electrophoresis.

3.2.2. DNA Sequencing

DNA extracted from *Porphyridium sp.* (that was produced from algae cells) was sequenced at FASTERIS, Switzerland by the high-throughput sequencing method of Solexa technology with Illumina's Genome Analyzer EAS210R version-GA IIX (San Diego, CA, USA) using the Pair-end method.

3.2.3. Contigs Assembly

The assembly of the reads to contigs was done using VELVET Version 0.7.54 suitable for *de novo* assembly [80]. Validation of the results was carried out by MAQ-Mapping and Assembly with Qualities software, version 0.7.1 [81].

In order to receive two database sources, two types of assemblies were made: (1) contigs database source containing DNA sequences based solely on reads, and (2) SCF database source, made by close contigs that were adjacent to each other using a number of unknown bases (N's).

3.3. Bioinformatic Tools for Analyzing DNA Sequences

Based on the fact that most algal genes have no introns, the partitioning of the sequences to ORFs was programmed with Glimmer v 3.02 [82], in the contigs database.

3.3.1. Blast2go

Analysis and annotation of the genes and proteins were done by blast2go [83,84].

3.3.2. Translating DNA Sequences into Protein Sequences

The DNA sequences were translated into protein sequences by a program available on the ExPASy (Expert Protein analysis system) [85].

3.3.3. Identification of Sequences by Similarity Comparison Using the Scaffold (SCF) Database

Based on the *Porphyridium* sp. SCF sequences database (Section 3.2.3), a search engine based on homological sequences according to similarity, was created by BLAST using Blast Command Line Application program [86], which is a part of NCBI C++ Toolkit.

3.4. Existence of Gene Sequences Involved in the N-Glycosylation Pathway

S. cerevisiae/Chlamydomonas reinhardtii/Galdieria sulphuraria/Cyanidioschyzon merolae/Arabidopsis thaliana were used as a reference species for sequence-based comparative analysis. Sequences belonging to the N-glycosylation pathway of these reference species were taken from Gene bank (NCBI) and used for searching for homological sequences according to similarity in the sequences of the algae's DNA using the SCF search engine (Section 3.3.3). The sequences identified in this study were translated into protein sequences (Section 3.3.2) in order to set the start and end of each protein sequence (the first methionine to the end codon, respectively). Each sequence was analyzed by BlastP to ensure the appropriate annotation to the related N-glycosylation proteins. In addition, multiple alignments were performed using the ClustalW algorithm [87].

3.5. Locating the Conserved Domain in the Algal Protein Sequences Involved in the N-Glycosylation Pathway

A homology study of conserved domains in various related *N*-glycosylation proteins based on the algae genome was conducted by blast2go. Separately, sequence detection of conserved domains was performed directly by Interpro and the added GO's (Gene ontology) [88–91], which was weighted to the sequence annotation.

3.6. Bioinformatic Comparison of Predicted Proteins Involved in the N-Glycan Biosynthesis with Ortholog Proteins of Various Organisms

Sequences belonging to encoded-genes involved in the *N*-glycosylation pathway of various organisms were taken from Gene bank (NCBI) and used for comparative similarity with the predicted ortholog sequences of the red microalga *Porphyridium* sp. The results were achieved using BLAST 2 seq [83,84].

3.7. Phylogenetic Tree of *Porphyridium* sp. Enzymes Involved in N-Glycosylation

A Phylogenetic tree showing evolutionary relationships of enzymes involved in *N*-glycosylation, was built based on integration of three major enzyme sequences involving *Porphyridium* sp. *N*-glycosylation pathway: ALG1, ALG7, and GANAB, and closely related enzymes from different organisms based on amino acid sequences obtained from NCBI protein database. The conserved domains of *Porphyridium* sp. sequences were detected by Interposcan. The tree was built using the Neighbor-Joining method with 500 repeated bootstrap tests, using MEGA software, version 4.0 [92].

4. Summary and Conclusions

The *N*-glycosylation pathway that occurs in the ER involves a large number of proteins, the majority of which are catalytic enzymes that are relatively well conserved in different organisms. We have identified genes encoding 24 products implicated in the ER *N*-glycan biosynthesis in the red microalga *Porphyridium* sp. A bio-informatic analysis of the *Porphyridium* sp. genome revealed the presence of a complete set of sequences potentially encoding for the following proteins: those involved in the synthesis of the lipid-linked Glc₃Man₉GlcNAc₂-PP-dolichol *N*-glycan; some subunits of the OST complex; ER glucosidases and chaperones required for protein quality control; and finally, the ER α -mannosidase I involved in the trimming of the *N*-glycan precursor into Man-5 *N*-glycan. As all the encoded genes involved in eukaryote *N*-glycosylation (before the *N*-glycan is transferred to the protein) were found in the *Porphyridium* sp. genome, it can be assumed that the final structure of the *N*-glycan in the ER before it is transferred to the protein, is the usual precursor Glc₃Man₉GlcNAc₂ as known in other organisms.

A study of gene conservation in the ER *N*-glycosylation pathway in *Porphyridium* sp. was also undertaken. The similarity of protein sequences to ortholog sequences from red and green algae, diatoms, mammals, and yeasts was studied. The similarity was found to be relatively high (above 40%), indicating the degree of conservation of the *N*-glycosylation pathway, and its importance in eukaryotes in general, and photosynthetic organisms in particular. This study showed that the

genomes of several species of green algae and all the diatoms lack two enzyme sequences, ALG10 and GCS1. Furthermore, it was found that the similarity of the STT3 gene of these organisms (several species of green algae and all the diatoms) in relation to *Porphyridium* sp. ortholog gene was smaller in comparison to the other organisms that were tested. As the ALG10 product is a substrate for STT3, it is likely that these organisms (diatoms and several green algae species) do not express the STT3 products as active enzymes. These findings indicate a close evolutionary relation of red algae to complex eukaryotes. Conversely, the OST encoded-subunits that were missing (that normally exist in higher eukaryotes) and the existence of several copies of the STT3 gene in *Porphyridium* sp. indicate its relation to lower organisms such as diatoms. This finding supports the theory that endosymbiosis took place between a eukaryote and a cyanobacterium as a single event that gave rise to all photosynthetic organisms. In addition, the grouping of *Porphyridium* sp. sequences with those of other red algae confirms that the homologs found in this study are in fact orthologs of the *N*-glycosylation enzymes that are also found in eukaryotes.

In summary, we demonstrated that *Porphyridium* sp. contains the majority of encoded-genes responsible for the *N*-glycosylation pathway in the ER as in eukaryotes organisms. Studies to elucidate the exact mode of action of these encoded-gene products are currently under way.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bhatia, P.K.; Mukhopadhyay, A. Protein glycosylation: Implications for *in vivo* functions and therapeutic applications. *Adv. Biochem. Eng. Biotechnol.* **1998**, *64*, 155–201.
2. Wormald, M.R.; Dwek, R.A. Glycoproteins: Glycan presentation and protein-fold stability. *Structure* **1999**, *7*, R155–R160.
3. Crocker, P.R.; Feizi, T. Carbohydrate recognition systems: Functional triads in cell-cell interactions. *Curr. Opin. Struct. Biol.* **1996**, *6*, 679–691.
4. Lee, J.; Park, J.S.; Moon, J.Y.; Kim, K.Y.; Moon, H.M. The influence of glycosylation on secretion, stability, and immunogenicity of recombinant HBV pre-S antigen synthesized in *Saccharomyces Cerevisiae*. *Biochem. Biophys. Res. Commun.* **2003**, *303*, 427–432.
5. Burda, P.; Aebi, M. The dolichol pathway of *N*-linked glycosylation. *Biochim. Biophys. Acta* **1999**, *1426*, 239–257.
6. Silberstein, S.; Gilmore, R. Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J.* **1996**, *10*, 849–858.
7. Knauer, R.; Lehle, L. The oligosaccharyltransferase complex from *Saccharomyces cerevisiae*. Isolation of the *OST6* gene, its synthetic interaction with OST3, and analysis of the native complex. *J. Biol. Chem.* **1999**, *274*, 17249–17256.
8. Bause, E. Structural requirements of *N*-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **1983**, *209*, 331–336.

9. Hettkamp, H.; Legler, G.; Bause, E. Purification by affinity chromatography of glucosidase I, an endoplasmic reticulum hydrolase involved in the processing of asparagine-linked oligosaccharides. *Eur. J. Biochem.* **1984**, *142*, 85–90.
10. Trombetta, E.S.; Simons, J.F.; Helenius, A. Endoplasmic reticulum glucosidase II is composed of a catalytic subunit, conserved from yeast to mammals, and a tightly bound noncatalytic HDEL-containing subunit. *J. Biol. Chem.* **1996**, *271*, 27509–27516.
11. Hammond, C.; Helenius, A. Quality control in the secretory pathway: Retention of a misfolded viral membrane glycoprotein involves cycling between the ER, intermediate compartment and Golgi apparatus. *J. Cell Biol.* **1994**, *126*, 41–52.
12. Nauseef, W.M.; McCormick, S.J.; Clark, R.A. Calreticulin functions as a molecular chaperone in the biosynthesis of myeloperoxidase. *J. Biol. Chem.* **1995**, *270*, 4741–4747.
13. Peterson, J.R.; Ora, A.; Van, P.N.; Helenius, A. Transient, lectin-like association of calreticulin with folding intermediates of cellular and viral glycoproteins. *Mol. Biol. Cell* **1995**, *6*, 1173–1184.
14. Parodi, A.J. Protein glycosylation and its role in protein folding. *Annu. Rev. Biochem.* **2000**, *69*, 69–93.
15. Ellgaard, L.; Helenius, A. ER quality control: Towards an understanding at the molecular level. *Curr. Opin. Cell Biol.* **2001**, *13*, 431–437.
16. Herscovics, A. Structure and function of class I α 1,2-mannosidases involved in glycoprotein synthesis and endoplasmic reticulum quality control. *Biochimie* **2001**, *83*, 757–762.
17. Helenius, A.; Aebi, M. Intracellular functions of *N*-linked glycans. *Science* **2001**, *291*, 2364–2369.
18. Weerapana, E.; Imperiali, B. Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **2006**, *16*, 91R–101R.
19. Balshüsemann, D.; Jaenicke, L. The oligosaccharides of the glycoprotein pheromone of *Volvox carteri f. nagariensis Iyengar* (Chlorophyceae). *Eur. J. Phycol.* **1990**, *192*, 231–237.
20. Becker, B.; Dreschers, S.; Melkonian, M. Lectin binding of flagellar scale-associated glycoproteins in different strains of *Tetraselmis* (Chlorophyta). *Eur. J. Phycol.* **1995**, *30*, 307–312.
21. Becker, D.; Melkonian, M. *N*-linked glycoproteins associated with flagellar scales in a flagellar green alga: Characterization of interactions. *Eur. J. Cell Biol.* **1992**, *57*, 109–116.
22. Becker, B.; Perasso, L.; Kammann, A.; Salzburg, M.; Melkonian, M. High molecular weight glycoprotein complexes link scales to the flagellar membrane in *Scherffelia dubia* (Chlorophyta). *Planta* **1996**, *199*, 503–510.
23. Gödel, S.; Becker, B.; Melkonian, M. Flagellar membrane proteins of *Tetraselmis striata* Butcher (Chlorophyta). *Protist* **2000**, *151*, 147–159.
24. Mamedov, T.; Yusibov, V. Green algae *Chlamydomonas reinhardtii* possess endogenous sialylated *N*-glycans. *FEBS Open Bio* **2011**, *1*, 15–22.
25. Mathieu-Rivet, E.; Scholz, M.; Arias, C.; Dardelle, F.; Schulze, S.; le Mauff, F.; Teo, G.; Hochmal, A.K.; Blanco-Rivero, A.; Loutelier-Bourhis, C.; *et al.* Exploring the *N*-glycosylation pathway in *Chlamydomonas reinhardtii* unravels novel complex structures. *Mol. Cell. Proteomics* **2013**, *12*, 3160–3183.

26. Baïet, B.; Burel, C.; Saint-Jean, B.; Louvet, R.; Menu-Bouaouiche, L.; Kiefer-Meyer, M.C.; Mathieu-Rivet, E.; Lefebvre, T.; Castel, H.; Carlier, A.; Cadoret, J.P.; *et al.* N-Glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *J. Biol. Chem.* **2011**, *286*, 6152–6164.
27. Levy-Ontman, O.; Arad, S.M.; Harvey, D.J.; Parsons, T.B.; Fairbanks, A.; Tekoah, Y. Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga *Porphyridium* sp. *J. Biol. Chem.* **2011**, *286*, 21340–21352.
28. Ramus, J. The production of extracellular polysaccharide by the unicellular red alga *Porphyridium aeruginum*. *J. Phycol.* **1972**, *8*, 97–111.
29. Ramus, J. Rhodophytes Unicells: Biopolymer Physiology and Production. In *Algal Biomass Technology*; Barclay, W.R., McIntosh, R.P., Eds.; J. Cramer: Berlin-Stuttgart, Germany, 1986; pp. 51–55.
30. Arad (Malis), S. Production of Sulfated Polysaccharides from Red Unicellular Algae. In *Algal Biotechnology—An Interdisciplinary Perspective*; Stadler, T., Mollion, J., Verduset, M.C., Eds.; Elsevier Applied Science: London, UK, 1988; pp. 65–87.
31. Arad (Malis), S.; Levy-Ontman, O. Red microalgal cell-wall polysaccharides: Biotechnological aspects. *Curr. Opin. Biotech.* **2010**, *21*, 358–364.
32. Lapidot, M.; Shrestha, R.P.; Weinstein, Y.; Arad (Malis), S. Red Microalgae: From Basic Know-How to Biotechnology; In *Red Algae in the Genomic Age*; Seckbach, J., Chapman, D.J., Eds.; Springer: Dordrecht, The Netherlands, 2010; pp. 205–225.
33. Cohen, E.; Arad (Malis), S. A closed system for outdoor cultivation of *Porphyridium*. *Biomass* **1989**, *18*, 59–67.
34. Cohen, E.; Koren, A.; Arad, S. A closed system for outdoor cultivation of microalgae. *Biomass Bioenergy* **1991**, *2*, 83–88.
35. Arad (Malis), S.; Richmond, A. Industrial production of microalgal cell-mass and secondary products-species of high potential: *Porphyridium* sp. In *Handbook of Microalgal Culture: Biotechnology and Applied Phycology*; Richmond, A., Ed.; Blackwell Science: Carlton, Australia, 2004; pp. 289–297.
36. Arad (Malis), S.; van Moppes, D. Novel Sulfated Polysaccharides of Red Microalgae: Basics and Applications. In *Handbook of Microalgal Culture: Applied Phycology and Biotechnology*, 2nd ed.; Richmond, A., Hu, Q., Eds.; Wiley: New Delhi, India, 2013; pp. 406–416.
37. Giddings, G.; Allison, G.; Brooks, D.; Carter A. Transgenic plants as factories for biopharmaceuticals. *Nat. Biotechnol.* **2000**, *18*, 1151–1155.
38. Walmsley, A.M.; Arntzen, C.J. Plant cell factories and mucosal vaccines. *Curr. Opin. Biotechnol.* **2003**, *14*, 145–150.
39. Rasala, B.A.; Muto, M.; Lee, P.A.; Jager, M.; Cardoso, R.M.; Behnke, C.A.; Kirk, P.; Hokanson, C.A.; Crea, R.; Mendez, M.; *et al.* Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* **2010**, *6*, 719–733.
40. Potvin, G.; Zhang, Z. Strategies for high-level recombinant protein expression in transgenic microalgae: A review. *Biotech. Adv.* **2010**, *28*, 910–918.

41. Specht, E.; Miyake-Stoner, S.; Mayfield, S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnol. Lett.* **2010**, *32*, 1373–1383.
42. Rasala, B.A.; Mayfield, S.P. The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. *Bioeng. Bugs* **2011**, *2*, 50–54.
43. Tran, M.; Zhou, B.; Pettersson, P.L.; Gonzalez, M.J.; Mayfield, S.P. Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnol. Bioeng.* **2009**, *104*, 663–673.
44. Tran, M.; Van, C.; Barrera, D.J.; Pettersson, P.L.; Peinado, C.D.; Bui, J.; Mayfield, S.P. Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E15–E22.
45. Franklin, S.E.; Mayfield, S.P. Prospects for molecular farming in the green alga *Chlamydomonas*. *Curr. Opin. Plant Biol.* **2004**, *7*, 159–165.
46. Lapidot, M.; Raveh, D.; Sivan, A.; Arad (Malis), S.; Shapira, M. Stable chloroplast transformation of the unicellular red alga *Porphyridium* species. *Plant Physiol.* **2002**, *129*, 7–12.
47. Plesser, E. Molecular Characterization of the Sulfotransferase from the Red Microalga *Porphyridium* sp. Ph.D. Thesis, Ben-Gurion University of the Negev, Beer-Sheva, Israel, 15 July 2009.
48. Matsuzaki, M.; Misumi, O.; Shin-I, T.; Maruyama, S.; Takahara, M.; Miyagishima, S.Y.; Mori, T.; Nishida, K.; Yagisawa, F.; Nishida, K.; *et al.* Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **2004**, *428*, 653–657.
49. Schönknecht, G.; Chen, W.H.; Ternes, C.M.; Barbier, G.G.; Shrestha, R.P.; Stanke, M.; Bräutigam, A.; Baker, B.J.; Banfield, J.F.; Garavito, R.M.; *et al.* Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **2013**, *339*, 1207–1210.
50. Bhattacharya, D.; Price, D.; Chan, C.X.; Qiu, H.; Rose, N.; Ball, S.; Weber, A.P.; Arias, M.C.; Henrissat, B.; Coutinho, P.M.; *et al.* Genome of the red alga *Porphyridium purpureum*. *Nat. Commun.* **2013**, *4*, 1941.
51. Qiu, H.; Yoon, H.S.; Bhattachary, D. Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Plant Sci.* **2013**, *4*, 366.
52. Arad, S.; Weinstein, Y. Novel lubricants from red microalgae: Interplay between genes and products. *Biomedic (Israel)* **2003**, *1*, 32–37.
53. Hoef-Emden, K.; Shrestha, R.P.; Lapidot, M.; Weinstein, Y.; Melkonian, M.; Arad, S. Actin phylogeny and intron distribution in bangiophyte red algae (Rhodoplantae). *J. Mol. Evol.* **2005**, *61*, 360–371.
54. Collé'n, J.; Porcel, B.; Carré, W.; Ballg, S.G.; Chaparroh, C.; Tonona, T.; Barbeyron, T.; Michel, G.; Noel, B.; Valentin, K.; *et al.* Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5247–5252.
55. Parker, M.S.; Mock, T.; Armbrust, E.V. Genomic insights into marine microalgae. *Annu. Rev. Genet.* **2008**, *42*, 619–645.
56. Schwatz, F.; Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **2011**, *21*, 576–582.

57. Kelleher, D.J.; Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **2006**, *16*, 47R–62R.
58. Wilson, C.M.; Kraft, C.; Duggan, C.; Ismail, N.; Crawshaw, S.G.; High, S. Ribophorin I associates with a subset of membrane proteins after their integration at the sec61 translocon. *J. Biol. Chem.* **2005**, *280*, 4195–4206.
59. Wilson, C.M.; High, S. Ribophorin I acts as a substrate-specific facilitator of *N*-glycosylation. *J. Cell Sci.* **2007**, *120*, 648–657.
60. Wilson, C.M.; Roebuck, Q.; High, S. Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9534–9539.
61. Schulz, B.L.; Stirnimann, C.U.; Grimshaw, J.P.; Brozzo, M.S.; Fritsch, F.; Mohorko, E.; Capitani, G.; Glockshuber, R.; Grütter, M.G.; Aebi, M. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11061–11066.
62. Spirig, U.; Bodmer, D.; Wacker, M.; Burda, P.; Aebi, M. The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology* **2005**, *15*, 1396–1406.
63. Nasab, F.P.; Schulz, B.L.; Gamarro, F.; Parodi, A.J.; Aebi, M. All in one: *Leishmania major* STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **2008**, *19*, 3758–3768.
64. Izquierdo, L.; Schulz, B.L.; Rodrigues, J.A.; Güther, M.L.S.; Proctor, J.B.; Barton, G.J.; Aebi, M.; Ferguson, M.A.J. Distinct oligosaccharide donor and peptide acceptor specificities of *Trypanosoma brucei* oligosaccharyltransferases. *EMBO J.* **2009**, *28*, 2650–2661.
65. Feldman, M.F.; Wacker, M.; Hernandez, M.; Hitchen, P.G.; Marolda, C.L.; Kowarik, M.; Morris, H.R.; Dell, A.; Valvano, M.A.; Aebi, M. Engineering *N*-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3016–3021.
66. Glover, K.J.; Weerapana, E.; Numao, S.; Imperiali, B. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **2005**, *12*, 1311–1315.
67. Igura, M.; Maita, N.; Kamishikiryo, J.; Yamada, M.; Obita, T.; Maenaka, K.; Kohda, D. Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **2008**, *27*, 234–243.
68. Abu-Qarn, M.; Yurist-Doutsch, S.; Giordano, A.; Trauner, A.; Morris, H.R.; Hitchen, P.; Medalia, O.; Dell, A.; Eichler, J. *Haloferax volcanii* AglB and AglD are involved in *N*-glycosylation of the *S*-layer glycoprotein and proper assembly of the surface layer. *J. Mol. Biol.* **2007**, *374*, 1224–1236.
69. Parodi, A.J. Role of *N*-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem. J.* **2000**, *348*, 1–13.
70. Ellgaard, L.; Molinari, M.; Helenius, A. Setting the standards: Quality control in the secretory pathway. *Science* **1999**, *286*, 1882–1888.
71. Lehrman, M.A. Oligosaccharide-based information in endoplasmic reticulum quality control and other biological systems. *J. Biol. Chem.* **2001**, *276*, 8623–8626.

72. Bhattacharya, D.; Archibald, J.M.; Weber, A.P.; Reyes-Prieto, A. How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays* **2007**, *29*, 1239–1246.
73. Moreira, D.; le Guyader, H.; Philippe, H. The origin of red algae and the evolution of chloroplasts. *Nature* **2000**, *405*, 69–72.
74. Nozaki, H.; Matsuzaki, M.; Takahara, M.; Misumi, O.; Kuroiwa, H.; Hasegawa, M.; Shin-i, T.; Kohara, Y.; Ogasawara, N.; Kuroiwa, T. The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J. Mol. Evol.* **2003**, *56*, 485–497.
75. Lane, C.E.; Archibald, J.M. The eukaryotic tree of life: Endosymbiosis takes its TOL. *Trends Ecol. Evol.* **2008**, *23*, 268–275.
76. Archibald, J.M. The puzzle of plastid evolution. *Curr. Biol.* **2009**, *19*, 81–88.
77. Burki, F.; Shalchian-Tabrizi, K.; Pawlowski, J. Phylogenomics reveals a new “megagroup” including most photosynthetic eukaryotes. *Biol. Lett.* **2008**, *4*, 366–369.
78. Dorrell, R.G.; Smith, A.G. Do red and green make brown? Perspectives on plastid acquisitions within chromalveolates. *Eukaryot. Cell* **2011**, *10*, 856–868.
79. Jones, R.H.; Speer, H.L.; Kury, W. Studies on the growth of the red alga *Porphyridium cruentum*. *Physiol. Plant* **1963**, *16*, 636–643.
80. Zerbino, D.R.; Birney, E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821.
81. Li, H.; Ruan, J.; Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **2008**, *18*, 1851–1858.
82. Delcher, A.L.; Bratke, K.A.; Powers, E.C.; Salzberg, S.L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **2007**, *23*, 673–679.
83. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676.
84. Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435.
85. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, .I; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788.
86. Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **2000**, *7*, 203–214.
87. Multiple Sequence Alignment. Available online: <http://www.ebi.ac.uk/clustalw> (accessed on 10 January 2011).
88. Hunter, S.; Jones, P.; Mitchell, A.; Apweiler, R.; Attwood, T. K.; Bateman, A.; Bernard, T.; Binns, D.; Bork, P.; Burge, S.; *et al.* InterPro in 2011: New developments in the family and domain prediction database. *Nucleic Acids Res.* **2011**, *40*, D306–D312.
89. Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Birney, E.; Biswas, M.; Bucher, P.; Cerutti, L.; Corpet, F.; Croning, M.D.; *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **2001**, *29*, 37–40.

90. Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Birney, E.; Biswas, M.; Bucher, P.; Cerutti, L.; Corpet, F.; Croning, M.D.R.; *et al.* InterPro-an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* **2000**, *16*, 1145–1150.
91. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; *et al.* Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29.
92. Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. MEGA4: Molecular evolutionary genetics analysis. *Mol. Biol. Evol.* **2007**, *24*, 1596–1599.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).