Photosynthesis Research **82:** 339–349, 2004.

Regular paper

Structure, circadian regulation and bioinformatic analysis of the unique sigma factor gene in *Chlamydomonas reinhardtii*

Matthew L. Carter[1], Annette C. Smith[2], Hirokazu Kobayashi[3], Saul Purton[2] & David L. Herrin[1,*]

[1]*Section Molecular Cell and Developmental Biology and Institute for Cellular and Molecular Biology, University of Texas at Austin, 1 University Station A6700 Austin, TX 78712, USA;* [2]*Department of Biology, University College London, Gower Street, London WC1E 6BT, UK;* [3]*Laboratory of Plant Cell Technology, Graduate School of Nutritional and Environmental Sciences, University of Shizuoka, 52-1 Yada, Shizuoka 422-8526, Japan; *Author for correspondence (e-mail: herrin@mail.utexas.edu; fax: +1512-471-3843)*

Received 24 December 2003; accepted in revised form 19 August 2004

Key words: Chlamydomonas, chloroplast transcription, circadian clock, gene regulation, nuclear–chloroplast interactions, *RPOD* gene, sigma factor

Abstract

In higher plants, the transcription of plastid genes is mediated by at least two types of RNA polymerase (RNAP); a plastid-encoded bacterial RNAP in which promoter specificity is conferred by nuclear-encoded sigma factors, and a nuclear-encoded phage-like RNAP. Green algae, however, appear to possess only the bacterial enzyme. Since transcription of much, if not most, of the chloroplast genome in *Chlamydomonas reinhardtii* is regulated by the circadian clock and the nucleus, we sought to identify sigma factor genes that might be responsible for this regulation. We describe a nuclear gene (*RPOD*) that is predicted to encode an 80 kDa protein that, in addition to a predicted chloroplast transit peptide at the N-terminus, has the conserved motifs (2.1–4.2) diagnostic of bacterial sigma-70 factors. We also identified two motifs not previously recognized for sigma factors, adjacent PEST sequences and a leucine zipper, both suggested to be involved in protein–protein interactions. PEST sequences were also found in ~40% of sigma factors examined, indicating they may be of general significance. Southern blot hybridization and BLAST searches of the genome and EST databases suggest that *RPOD* may be the only sigma factor gene in *C. reinhardtii*. The levels of *RPOD* mRNA increased 2–3-fold in the mid-to-late dark period of light-dark cycling cells, just prior to, or coincident with, the peak in chloroplast transcription. Also, the dark-period peak in *RPOD* mRNA persisted in cells shifted to continuous light or continuous dark for at least one cycle, indicating that *RPOD* is under circadian clock control. These results suggest that regulation of *RPOD* expression contributes to the circadian clock's control of chloroplast transcription.

Introduction

The plastid organelle of plants, algae, and apicomplexan parasites contains its own genetic system. In higher plants, the transcription of plastid genes is carried out by at least two different types of RNA polymerase (RNAP). The 'plastid-encoded polymerase' (PEP) is a multisubunit enzyme homologous to the eubacterial RNAP, and consists of a catalytic core encoded by the plastid genes *rpoA*, *B*, *C1* and *C2*. The 'nuclear-encoded polymerase' (NEP) is a single-subunit bacteriophage-type RNAP encoded in the nucleus. One or more different NEPs are targeted to the plant plastid where they are believed to be involved primarily in the transcription of housekeeping

genes (Hess and Borner 1999). In eukaryotic algae and apicomplexa, there is no evidence for the presence of NEPs in the plastid (Smith and Purton 2002). It therefore appears that all plastid genes in these organisms are transcribed by the PEP enzyme.

By analogy with the bacterial enzyme, it is generally assumed that PEP promoter specificity is achieved by sigma factors, and that these factors provide a mechanism for the nuclear regulation of PEP, since sigma factors have not been found on any plastid genomes sequenced to date. This idea is supported by the discovery in red algae, and subsequently in higher and lower green plants, of nuclear gene families encoding homologues of the sigma-70 class of principal bacterial factors (reviewed in Allison 2000). These sigma-like factors (SLFs) have been shown to be targeted to the plastid (e.g. Isono et al. 1997; Tan and Troxler 1999) where they appear to regulate transcription of subsets of plastid genes in response to different environmental and developmental signals (Kanamaru et al. 2001). The expression of many SLF genes has been shown to be up-regulated by light and to be tissue-specific, with maximal expression in green tissue (e.g. Isono et al. 1997; Tozawa et al. 1998; Lahiri et al. 1999). In addition, there is evidence of circadian regulation of expression for some members of the sigma factor gene families (Kanamaru et al. 1999; Morikawa et al. 1999; Harmer et al. 2000; Oikawa et al. 2000). Thus, the presence of multiple sigma factors allows for the regulation of plastid transcription in response to various environmental and endogenous signals.

Previous studies of chloroplast gene expression in cultures of the green alga *Chlamydomonas reinhardtii* grown under a light–dark cycle showed that the transcription of several major chloroplast-encoded genes, as well as total chloroplast transcription, peaks at the start of the light phase (Herrin et al. 1986; Leu et al. 1990; Salvador et al. 1993), and that this rhythm is circadian (Hwang et al. 1996). Recently, it was shown that the circadian peak of transcription of chloroplast genes requires continued cytoplasmic protein synthesis, suggesting that clock control is mediated by one or more nuclear-encoded proteins (Kawazoe et al. 2000). Obvious candidates for such proteins are sigma factors, although no sigma factor genes have yet been described for *C. reinhardtii*, nor any other green alga. Here we report on the characterization of a SLF gene (*RPOD*) from *C. reinhardtii*, and show that its expression is under circadian control. Interestingly, genomic Southern blots and searches of ESTs and the genome sequence (JGI version 2.0) suggest that *RPOD* may be the only bona-fide SLF gene in this organism.

Materials and methods

Cell culture

The wild-type strain 2137 mt+ (CC-1021) of *C. reinhardtii* was grown in flasks of Tris/acetate/phosphate (TAP) medium (Harris 1989) with shaking (~125 rpm) at 23 °C. To obtain synchronous cultures, cells were grown for at least three days in 12 light:12 h dark (LD) cycles; the light flux was ~60 $\mu Em^{-2}\ s^{-1}$. Cell synchrony was monitored microscopically, and by counting cells with a hemacytometer. Under these conditions, the cells divide once each day into 2, 4, or 8 daughter cells with an average increase of 2–4 fold each 24 h (during log phase). For circadian experiments, log-phase synchronous cultures were diluted with fresh TAP medium to a concentration of $\sim 0.5 \times 10^6$ cells/ml, and then placed into either continuous light (LL) or continuous darkness (DD). In DD, cell division slows drastically for at least 48 h (Hwang and Herrin 1994). The LL cultures were diluted back to log phase numbers (0.3–2.0×10^6 cells/ml) with fresh medium when cell counts exceeded this range. Care was taken to minimize the exposure of dark-phase cells to photosynthetically active light by working quickly and by using a dim ($< 1\ \mu Em^{-2}\ s^{-1}$) green light.

Isolation and sequencing of cDNA and genomic clones

Three different cDNA clones representing two 5′-truncated cDNAs from *RPOD* (EST accession numbers AV392956 and AV390342) and a near full-length cDNA (AV641965) were obtained from the Kazusa DNA Research Institute (Asamizu et al. 1999). Sequencing revealed that clone AV392956 actually contains two different ~1.6-kb cDNA inserts, only one of which encodes *RPOD*; this insert was recloned into pBluescript SK- to form plasmid pMatE. The pMatE insert and the ~1.4-kb insert of AV390342 were completely

sequenced (the Genbank accession numbers are AY512563 and AY512562), and the AV641956 insert was sequenced until it overlapped with AV392956 (Figure 1).

Genomic clones containing *RPOD* were obtained by screening a cosmid library of the *C. reinhardtii* genome (Purton and Rochaix 1995) with the AV392956/pMatE insert. A 4.6-kb HindIII fragment containing most of the gene (see Figure 1) was isolated from cosmid clone c8, and cloned into pBluescript SK- to give plasmid pSK.rpoDg. The sequence of the HindIII insert was determined. The sequence of the remainder of the gene (exon 1 and upstream region) was obtained by sequencing directly from cosmid clone c8. The assembled gene sequence was deposited in Genbank as accession number AF525691. DNA sequencing was carried out enzymatically using fluorescent dye-terminator nucleotides in conjunction with an ABI 377 DNA sequencer.

Southern and northern blot analyses

Southern blot analysis was carried out as described by Sambrook et al. (1989) using *C. reinhardtii* genomic DNA prepared as per Rochaix et al. (1988). The hybridization probes were the inserts from cDNA clones AV390342 and AV392956, respectively, labeled with ^{32}P-dCTP to $\sim 1 \times 10^9$ dpm/μg. Following hybridization, the filter was washed at 55 °C in 0.1 × SSC, 0.1% SDS. Northern blotting was performed as described previously (Hwang and Herrin 1994); the formaldehyde/agarose gels were loaded with 10 or 20 μg RNA per lane, and the blots were stained with methylene blue (Herrin and Schmidt 1988) to check loading and transfer. The hybridization probe was the 1.6-kb *RPOD* cDNA insert from AV393956/pMatE labeled with ^{32}P-dCTP to $\sim 1 \times 10^9$ dpm/μg. After hybridization, the membranes were washed at high stringency (0.1 × SSPE, 0.5% SDS at 65 °C), and exposed to X-ray film (BioMax MS, Kodak) at −70 °C with an HE Tran-screen (Kodak). Exposures within the linear range of the film were scanned and quantified with image analysis software (Kodak 1D v3.6). The methylene blue stained membranes were also scanned, quantified, and used to correct the data from the autoradiographs for uneven loading of the gels. The graphs were generated with Excel (Microsoft) using the curve fitting features. The LD and circadian analyses of *RPOD* mRNA levels were repeated at least twice with different cultures, and representative data are shown.

Bioinformatics

Homology searches of the *C. reinhardtii* genome (http://genome.jgi-psf.org/chlre2/chlre2.home) and EST databases (www.biology.duke.edu/chlamy_\genome; www.kazusa.or.jp/en/plant/chlamy/EST/) were carried out on-line using the BLAST suite of software (Altschul et al. 1990). Motifs within the *RPOD* protein sequence were identified using the PredictProtein, PESTfind and 2ZIP programs, among others (Rechsteiner and Rogers 1996; Rost 1996; Bornberg-Bauer et al. 1998; www.expasy.org), and targeting analysis was performed with at least three separate prediction programs: ChloroP (Emanuelsson et al. 1999), PCLR (Schein et al. 2001) and TargetP (Emanuelsson et al. 2000). Phylogenetic analysis was carried out using ClustalW (Thompson et al. 1994) and MacVector 7.2 (Accelrys, Inc.). Trees were built using amino acid

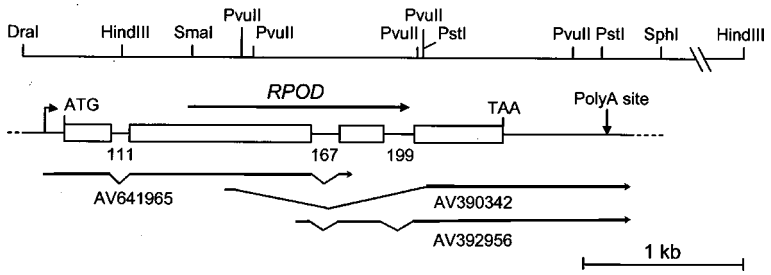

Figure 1. Physical map of the *RPOD* locus. The coding region is shown as open boxes disrupted by three introns of sizes 11, 167 and 199 bp. The putative start of transcription is shown as a forward arrow upstream of the first exon, and the site of polyadenylation (PolyA site) as a downward arrow. The sequenced regions of three cDNA clones are indicated below the gene; the thinner lines below the horizontal indicate internal regions that are missing from the cDNAs.

sequences registered in GenBank (www.ncbi.nlm.nih.gov/) and CyanoBase (www.kazusa.or.jp/cyano/) databases. Only one independently registered sequence was employed when the sequences were for the same gene from the same organism, and only the sequences in the conserved 2.1 to 4.2 regions of sigma-70 factors were used to calculate the tree.

Results

Isolation and characterization of RPOD

In order to identify *C. reinhardtii* genes for sigma factors we carried out a BLAST search of a *C. reinhardtii* EST database (Asamizu et al. 1999) using sequences from cyanobacterial RNAP sigma factors. We identified several cDNA clones encoding a protein showing strong homology (~46% identity) in regions to the cyanobacterial proteins. The complete sequencing of one such clone (AV392956) confirmed that the protein contained the conserved regions 2.1–4.2 that are diagnostic of the sigma-70 class of principal sigma factors (Tanaka et al. 1988). The cDNA clone was used to isolate the corresponding gene from a *C. reinhardtii* cosmid library (Purton and Rochaix 1995). Sequencing of the genomic gene (*RPOD*) indicated that AV392956 contains a truncated cDNA lacking the 5′ end, however a further search of the EST database using the genomic sequence allowed the identification of additional ESTs (e.g. AV641965) representing near full-length cDNA clones for *RPOD* (Figure 1).

The structure of *RPOD* is typical of *C. reinhardtii* nuclear genes in that it has a high GC content (63%) and a marked codon bias for G or C in the third position (Silflow 1998). The gene is split by three introns of 111, 167, and 199 bp, whose sizes and boundary sequences match the consensus for *C. reinhardtii* introns (Silflow 1998). The splice sites were confirmed from the cDNA sequences (AV641965 and AV392956 in Figure 1). It should be noted that the insert of AV390342 contains a fragment of exon 2 fused to a site 93 bp into exon 4, and containing the rest of the *RPOD* mRNA sequence (Figure 1). This is most likely an artifact of reverse transcription, since the deleted segment does not have typical intron-terminal nucleotides, and there is a short sequence (GCTGGC) that is repeated at the fusion sites in exon 2 and exon 4. A conserved sequence (TGTAA) that is a putative polyadenylation signal is located 639 bp downstream of the stop codon and 10 bp upstream of the polyadenylation site. The transcriptional start was not determined experimentally; however, analysis of the 5′-most *RPOD* sequences present in the *C. reinhardtii* EST databases suggests that transcription starts ~132 bp upstream of the first AUG used for translation start (Figure 1).

To determine if *RPOD* was part of a gene family within the haploid genome of *C. reinhardtii*, Southern blot analysis of total genomic DNA was performed under moderate stringency conditions using as probe the cDNA insert from AV390342 (Figure 1), which contains conserved regions 3.1, 3.2, 4.1, and 4.2 of sigma-70 factors (Paget and Helmann 2003). As shown in Figure 2, only a single strongly hybridizing band is detected for each restriction digest; also, the size of the major band corresponds to the predicted gene map, at least for those fragments that lie within the sequenced region (lanes 2, 3 and 5; and see Figure 1). This result suggests that *RPOD* is present as a single copy within the genome. Hybridization of a similar blot with the original AV392956 clone, which contained the *RPOD* insert shown in Figure 1 plus that from another *C. reinhardtii* gene, produced the same set of fragments as in Figure 2 plus an additional one in each lane (data not shown), consistent with both cDNAs being from single-copy genes.

The presence of some additional weakly hybridizing bands on the genomic blot could be attributed to related SLF genes. We therefore carried out BLAST searches (expect value of 1) with *RPOD*, and other sigma factor genes, of the *C. reinhardtii* genome sequence (JGI version 2.0) as well as the EST databases (Asamizu et al. 1999; Shrager et al. 2003). No other genes were found, other than *RPOD* (which is JGI gene C_30171), that contain the conserved sigma factor regions 2.1–4.2. However, we did detect a gene (JGI gene C_30152) with weak similarity (~25%) to the sigma-70 regions 2.1–2.4 and 3.1, but lacking conserved regions 3.2, 4.1, and 4.2, and having no other similarities to *RPOD*. Predicted localization analysis suggested that this pseudo-sigma-70 factor protein is likely targeted to mitochondria. These results indicate that *C. reinhardtii* may possess only one bona fide sigma-70 factor gene.

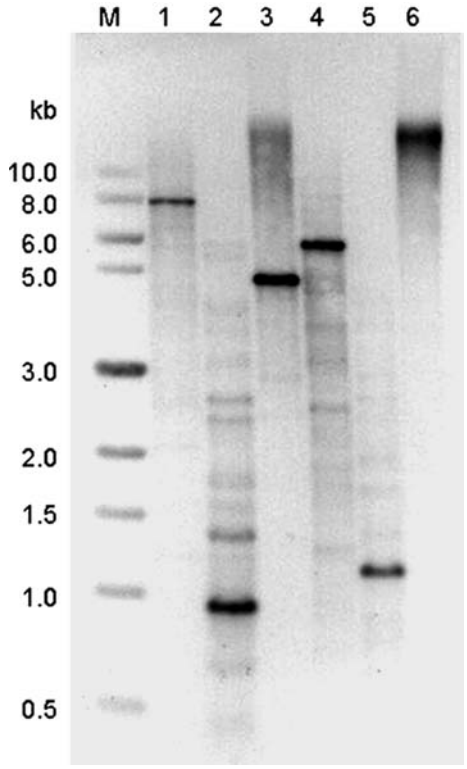

Figure 2. Genomic southern blot analysis of *RPOD*. Total genomic DNA was digested with: 1. *Sma*I, 2. *Pvu*II, 3. *Hin*dIII, 4. *Sph*I, 5. *Pst*I, and 6. *Hpa*I, and probed with the cDNA insert from clone AV390342 (Figure 1). The filter was washed in 0.1 × SSC at 55 °C and exposed to X-ray film for ~48 h.

Motifs in the RPOD protein

The predicted sequence of the RPOD protein suggests an acidic (pI 5.8) precursor polypeptide of 752 amino acids (80 kDa). As indicated in Figure 3, the conserved sigma-70 regions 2.1–4.2 lie in the C-terminal half of the protein. At the N terminus is a putative chloroplast targeting sequence of 23 amino acids. Other interesting protein motifs were found in the non-conserved region upstream of the sigma-70 domain. The PESTfind program predicted two good PEST sequences (scores of +11 and +10, respectively) very close together (aa 283–310 and 315–352 of the precursor). PEST sequences are short peptide sequences rich in proline (P), glutamate (E), serine (S), and threonine (T), and can cause increased susceptibility to proteolytic degradation. They are also likely sites of protein-protein interaction (Rechsteiner and Rogers 1996). Since PEST sequences had not been reported previously for chloroplast sigma factors, we also examined most of the known plastid sigma factors for the presence of such sequences, and found that ~40% contain at least one good (i.e., score < +5) PEST sequence. A large number of the prokaryotic sigma factors also contain good PEST sequences, and these, as well as the plant factors, are indicated with an asterisk in Figure 4.

Another interesting motif in RPOD that is also implicated in protein–protein interactions is a leucine zipper (LZ), which is close to the N-terminus (aa 57–91 of the precursor). Although this is an unusual location for a LZ motif, this region was also predicted to be a coiled-coil (by 2ZIP and COILS), which is required of bonafide LZs that dimerize (Bornberg-Bauer et al. 1998). Also, the LZ motif in RPOD is predicted to be at least four heptads long, well within the typical range (Deppmann et al. 2004). The fourth heptad has a charged amino acid (E) at the *a* location that may prevent homodimerization (Deppmann et al. 2004). Thus, the role of the LZ may be to interact with another protein, possibly a RNAP core subunit (Troxler et al. 1994) or another regulatory

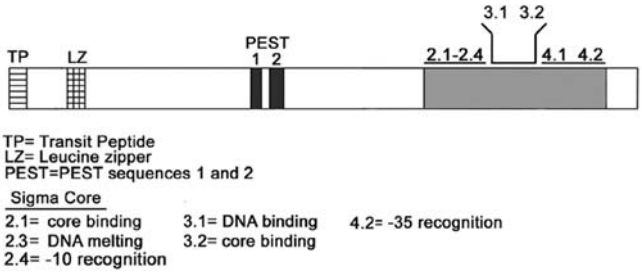

Figure 3. Motifs of the *C. reinhardtii* RPOD protein. The predicted protein structure (752 aa) contains several regions of interest, as indicated in the diagram, which is close to scale. In the non-conserved N-terminal half is the predicted chloroplast transit peptide (TP, aa 1–23), a LZ motif (aa 57–91), and two PEST sequences (aa 283–310 and 315–352, respectively). The conserved sigma factor region (aa 485–715) contains several subregions (2.1–4.2) whose likely functions are as indicated.

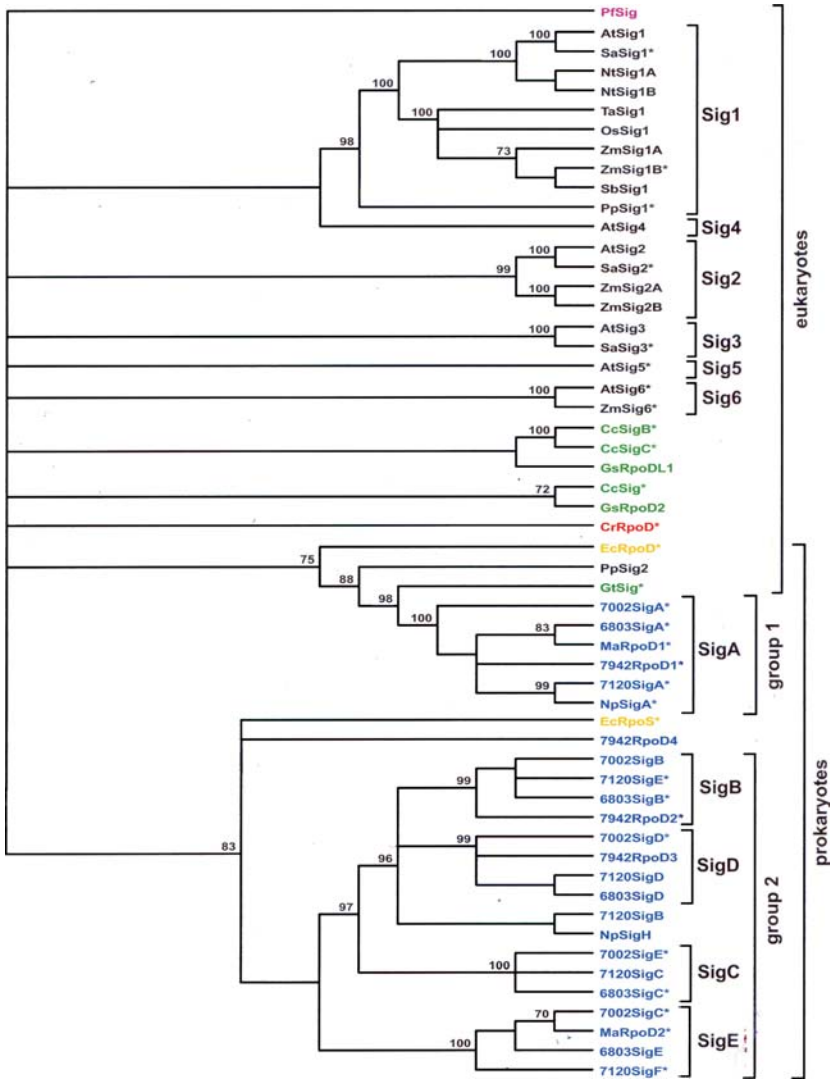

Figure 4. Phylogenetic tree of sigma-70 factors; those with PEST sequences are indicated with an asterisk (*). The tree of sigma-70 factors, from cyanobacteria to higher plants, was built using amino acid sequences for the conserved 2.1 to 4.2 regions (Figure 3). ClustalW was used for the alignments, and then a neighbor-joining tree [with bootstrapping (nodes with $<60\%$ support were collapsed); midpoint rooting] was generated with MacVector 7.2. The numbers in the tree are bootstrap values, and represent the percent occurrences of the nodes that appeared very frequently in a resampling of the data 1000 times. The *C. reinhardtii* gene/protein is indicated in red (CrRPOD*). The other species, gene names, accession numbers and color codes are as follows: land plants (black) – At, *Arabidopsis thaliana* [Sig1, AB004821; Sig2, AB004820; Sig3, AB004822; Sig4, AB021119; Sig5, AB021120; Sig6, AB029916]; Nt, *Nicotiana tabacum* [Sig1A, AB023571; Sig1B, AB023572]; Os, *Oryza sativa* [Sig1, AB005290]; Sa, *Sinapsis alba* [Sig1, Y15899; Sig2, AJ276656; Sig3, AJ276657]; Sb, *Sorghum bicolor* [Sig1, Y14276];Ta, *Triticum aestivum* [Sig1, AJ132658]; Zm, *Zea mays* [Sig1A, AF058708; Sig1B, AF058709; Sig2A, AF099110; Sig2B, AF099111; Sig6, AF099112]; moss: Pp, *Physcomitrella patens* [Sig1, AB059354; Sig2, AB046872]; eukaryotic algae (green, except for *C. reinhardtii*) – Cc, *Cyanidium caldarium* RK-1 [Sig, D83179; SigB, AB006798; SigC, AB006799]; Cr, *Chlamydomonas reinhardtii* [RpoD, this study]; Gs, *Galdieria sulphuraria* (*Cyanidium caldarium*, Allen strain) [RpoDL1, L42639; RpoD2, AF050634]; Gt, *Guillardia theta* [Sig, AAK39807]; protozoa (purple) – *Plasmodium falciparum* 3D7 [Pf, CAD50481]; cyanobacteria (blue) – 6803, *Synechocystis* sp. PCC 6803 [SigA, slr0653; SigB, sll0306; SigC, sll0184; SigD, sll2012; SigE, sll1689]; 7002, *Synechococcus* sp. PCC 7002 [SigA, U15574; SigB, U82435; SigC, U82436; SigD, U82484; SigE, U82485]; 7120, *Anabaena* sp. PCC 7120 [SigA, M60046; SigB, M95760; SigC, M95759; SigD, AF262216; SigE, AF262217; SigF, AF262218]; 7942, *Synechococcus* sp. PCC 7942 [RpoD1, D10973; RpoD2, D78583; RpoD3, AB024709; RpoD4, AB024710]; Ma, *Microcystis aeruginosa* K-81 [RpoD1, D85684; RpoD2, D86575]; Np, *Nostoc punctiforme* PCC 73102 [SigA. AF265349; SigH, AF022822]; *Escherichia coli* (yellow) – Ec, *Escherichia coli* [RpoD, J01687; RpoS, X16400].

protein. To our knowledge, a LZ motif has not previously been reported for a sigma factor. We did find other sigma factors that have predicted coiled-coil domains, but not a classic LZ motif.

Finally, there is a low complexity region in RPOD between the LZ and PEST sequences (aa 105–163) that contains several stretches of repeated amino acids (3–5 aa long), particularly alanine and serine. Similar sequences have been seen in nuclear gene products that function in post-transcriptional steps in chloroplast gene expression in *C. reinhardtii* (Vaistij et al. 2000; Rivier et al. 2001), although their significance is not clear. The serine (and threonine)-rich part of this region (aa 153–163) is predicted with high confidence (by NetPhos 2.0 and PROSITE) to be phosphorylated by a CK2-type kinase. This finding is of interest because some higher plant chloroplast sigma factors have been shown to be phosphorylated by a CK2-type enzyme (Baginsky et al. 1997; Ogrzewalla et al. 2002).

Phylogenetic analysis

It is generally accepted that plastids evolved from a cyanobacterial endosymbiont, and that two main lineages of plastid-bearing eukaryotes have emerged from the primary endosymbiotic event; green algae and plants representing one lineage, and red algae the other. Other algal plastids have evolved through the secondary acquisition of plastids from either red or green algae (Delwiche 1999). Phylogenetic analysis of the sigma-70 factor families from higher plant species (black letters in Figure 4) indicate that the factors can be grouped into distinct classes, and that this gene divergence pre-dates the divergence of these species from a common ancestral plant (Fujiwara et al. 2000 and Figure 4). In order to investigate the relationship of the *C. reinhardtii RPOD* to the plant, red algal and cyanobacterial sigma factors we constructed a phylogenic tree using the conserved region of sigma-70 factors (Figure 4); the aa sequence identity (relative to *RPOD*) ranged from 46% to 30%. In general, the prokaryotic and eukaryotic sigma factors formed distinct clades, except for subgroup 1 of prokaryotes, which, interestingly, also has the moss (*Physcomitrella patens*) *sig2* gene, and a *sig* gene from the cryptomonad alga, *Guillardia theta*. This affinity could be explained by recent horizontal gene transfer, which seems especially likely for the moss sigma factor, but poor sequence conservation of a specific type of sigma factor might also account for the results. The tree also indicates that the *C. reinhardtii RPOD* gene (CrRpoD in Figure 4) is not closely related to the other sigma factor genes, possibly because it is the only green algal gene in the tree. The data also suggest that *RPOD* diverged early from the other eukaryotic sigma factors, and that sigma factor genes are ancient eukaryotic genes. The tree does not clearly inform us as to which of the extant cyanobacterial sigma factor lines was the ancestor to the *C. reinhardtii RPOD* gene.

LD and circadian regulation of RPOD *mRNA*

The expression of *RPOD* was examined at the level of mRNA accumulation during a light-dark (LD) cycle. Figure 5 shows a northern blot analysis during a typical 12 h:12 h LD cycle. The *RPOD*-specific DNA probe hybridized mainly to a 3.2-kb RNA, which is close to the 3.1-kb transcript size predicted from the cDNA and genomic sequences. The data also shows that the level of *RPOD* mRNA fluctuated ~2.5-fold during the 24-h LD cycle, peaking late in the dark period. This pattern correlates quite well with the LD patterns of chloroplast gene transcription, except that the amplitude of the *RPOD* mRNA peak is less than the total and specific gene transcription amplitudes analyzed previously (except for *psbA*) (Hwang et al. 1996).

The regulation of *RPOD* mRNA was also examined under circadian conditions, i.e. in LD-entrained cells shifted to continuous darkness (DD). Figure 6 shows a northern blot analysis of *RPOD* mRNA levels during the first, and part of the second, cycle in DD. Levels of *RPOD* mRNA peaked in the subjective mid-dark period of the first cycle in DD, and then increased on schedule during the second. These results indicate that the mostly nocturnal rhythm of *RPOD* mRNA levels is under circadian clock control. It should be noted that some shifting in peak characteristics, such as broadening, decreased amplitudes, and shifts in period length under continuous conditions is common for circadian rhythms, and was also seen with the chloroplast transcription rhythms (Hwang et al. 1996).

Figure 7 shows a northern blot analysis of *RPOD* mRNA levels in LD-entrained cells shifted to LL. In this experiment, we are testing for the

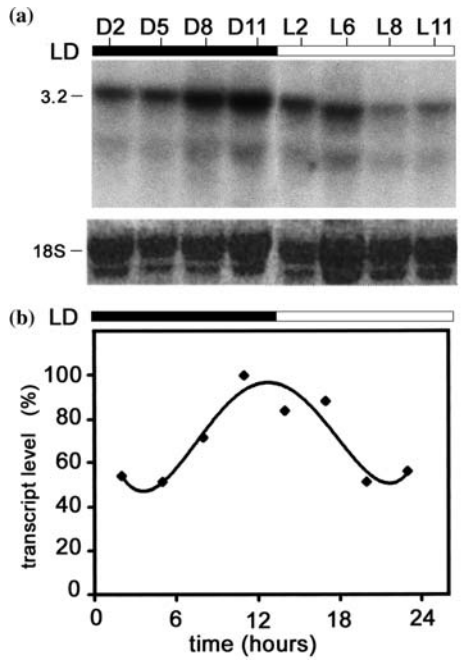

Figure 5. Northern blot analysis of *RPOD* mRNA levels during a light-dark (LD) cycle. (a) The upper panel shows an autoradiograph of a northern blot hybridized with the *RPOD*-specific probe. Total RNA (15 μg/lane), extracted the indicated times (top of figure) from cells growing synchronously under a 12:12 LD cycle, were separated on a denaturing agarose gel, and the blot was exposed to X-ray film for ~24 h. The 3.2-kb RNA corresponding to the *RPOD* transcript is indicated. The lower panel shows the methylene blue-stained 18S rRNA that was used as a loading control. (b) Graphical presentation of the relative levels of *RPOD* mRNA. The hybridization signals of the 3.2-kb RNA in panel (a) were quantified and corrected for unequal loading as described in Material and Methods. The maximum level was set to 100% for normalization. Time (in h) on the *X*-axis is from the beginning of the sampling period. Closed box, dark period; open box, light period.

second expected peak in LL. The figure shows that *RPOD* mRNA increased ~2.5-fold, peaking during the middle of the subjective dark period. These data show that the circadian rhythm of *RPOD* mRNA occurs in LL as well as DD.

Discussion

We have identified and characterized a novel nuclear gene of *C. reinhardtii* (*RPOD*) that encodes a putative sigma factor for the chloroplast RNAP. Importantly, our analysis of genomic Southern blots and low-stringency BLAST searches of the most recent *C. reinhardtii* genome assemblage (JGI version 2.0) and EST databases suggests that there may be only one sigma factor gene encoded within this model alga's genome – a situation that makes *C. reinhardtii* different from other photosynthetic organisms studied to date, which possess a family of sigma-70 factors. It should be noted, however, that the nuclear genome sequence is not 100% complete (it is estimated to be ~90% complete), so there could be other sigma factor genes yet to be found in this organism – although the fact that the EST databases also lack any other sigma-70 type genes would suggest that the expression of such genes, if they exist, is likely to be low. On the other hand, the expression of *RPOD* was detected by northern blot hybridization of total RNA, and there is an expanding collection of ESTs associated with the gene. Moreover, the fact that *RPOD* mRNA levels in LD-cycling cells peak just prior to, or coordinately with, the daily peak in chloroplast transcription

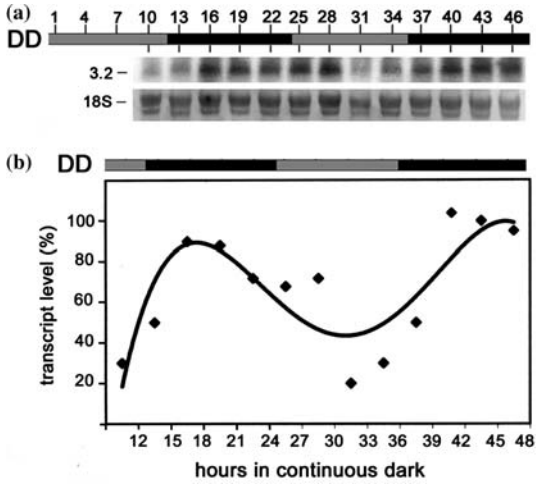

Figure 6. Northern blot analysis of *RPOD* mRNA in continuous darkness (DD). (a) LD-entrained cells were shifted into DD, total RNA was isolated at the indicated times (h 10 – 46, top of figure), and 10 μg/lane was hybridized with the *RPOD*-specific probe. The position of the 3.2-kb *RPOD* transcript in the autoradiograph is indicated. The lower panel shows the methylene blue-stained pattern of 18S rRNA that was used as a loading control. (b) Graphical representation of relative *RPOD* mRNA levels. Quantification of the 3.2-kb RNA in (a) was adjusted for unequal loading, and the maximum level was set to 100%. Continuous time on the X-axis is measured from the beginning of the light period in which sampling began. Gray boxes, subjective light periods; black boxes, subjective dark periods.

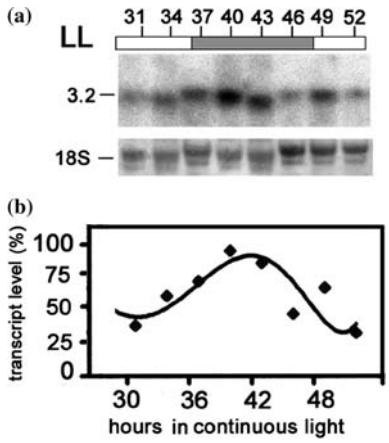

Figure 7. Northern blot analysis of *RPOD* mRNA in continuous light (LL). (a) LD-entrained cells were shifted to LL, RNA was isolated at the indicated time points (h of continuous light, top of figure) and 10 μg/lane was separated. The blot was hybridized with a *RPOD*-specific cDNA, and exposed to X-ray film for ~24 h. The position of the 3.2-kb *RPOD* transcript is indicated. The lower panel shows the methylene blue-stained pattern of 18S rRNA that was used as a loading control. (b) Graphical presentation of the quantification of the northern blot shown in (a). The *RPOD* mRNA levels were corrected for unequal loading, and the maximum level was set to 100%. Continuous time is also shown, and is measured from the beginning of the light period in which sampling began. Open boxes, subjective light periods; gray box, subjective dark period.

(Hwang et al. 1996) suggests that *RPOD* plays an important role in this process.

The mature *C. reinhardtii RPOD* protein is predicted to be 78 kDa, and may therefore correspond to the 82-kDa polypeptide detected by Troxler et al. (1994) using antibodies to an *Anabaena* (PCC 7120) sigma factor; the 82-kDa protein also correlated with RNAP activity in chromatographic fractions. The relationship of RPOD to the 51-kDa protein (with sigma factor activity) purified by Surzycki and Shellenbarger (1976) is not clear. If the 51-kDa protein represented a different sigma factor (than RPOD) one would expect a protein of such abundance to be represented by ESTs, but no other sigma-70 sequences could be found. A more likely explanation, in our opinion, is that the 51-kDa protein was a proteolytic fragment of RPOD containing the conserved *C*-terminal sigma-70 regions (Figure 3). It is intriguing that PEST sequences, believed to cause protein instability (Rechsteiner and Rogers 1996), are found near the middle of RPOD, and, moreover, cleavage of the protein near these sequences would generate an ~50-kDa polypeptide.

Our finding that *C. reinhardtii* appears to possess only a single sigma factor gene, together with the apparent absence of a chloroplast NEP (Smith and Purton 2002), suggests that the nuclear regulation of chloroplast transcription could be much simpler in this green alga than in higher plants. This may be due, at least in part, to the fact that the chloroplast in *C. reinhardtii* does not undergo the elaborate process of plastid differentiation and maturation seen in angiosperms. Rather, a fully functional mature chloroplast is formed in each *C. reinhardtii* cell, even when grown under heterotrophic conditions in complete darkness (Harris 1989). This notion that a simplified transcription apparatus is found in organisms with a 'constitutive plastid' is supported by our finding that the apicomplexan parasite, *Plasmodium falciparum*, which possesses a single type of non-pigmented plastid, also appears to have only one sigma factor gene and no plastid NEP gene in its nuclear genome (our unpublished results).

RPOD has two predicted protein motifs that have not been reported previously for sigma factors: a LZ and PEST sequences. LZs usually occur in certain eukaryotic transcription factors, most notably bZIP proteins, where they mediate dimerization via a coiled-coil structure (Bornberg-Bauer et al. 1998; Deppmann et al. 2004); it should be emphasized that the LZ region in RPOD is also predicted to be a coiled-coil. To our knowledge, sigma factors have not been reported to exist as dimers, although RPOD could be the first. Alternatively, the LZ domain could interact with coiled-coil domains of other proteins, possibly to regulate RPOD function.

PEST sequences have been found mostly in regulatory proteins, where they often confer reduced stability (Rechsteiner and Rogers 1996); there are two adjacent PEST sequences predicted for the non-conserved portion of *C. reinhardtii* RPOD. Although PEST sequences have not been reported for sigma factors, we found that ~40% of them contain at least one good PEST sequence; the PEST sequences are found in conserved as well as non-conserved regions of the proteins (our unpublished results). Interestingly, one PEST-containing sigma factor that has been well studied, the *E. coli rpoS* gene product, is an unstable protein whose stability is regulated by the *sprE*

gene, a response regulator (Pratt and Silhavy 1996).

These results also have implications for the control of chloroplast transcription by the circadian clock. Previously, it was shown that total chloroplast transcription, as well as the transcription of several major chloroplast-encoded genes, is under circadian control, and that this control is mediated by the nucleus (Hwang et al. 1996; Kawazoe et al. 2000). It was suggested that a nuclear-encoded sigma factor could be responsible for the circadian program of chloroplast transcription, with the most obvious mechanism being simply the circadian expression of the sigma factor. If this were the case, one would expect a daily peak of sigma factor mRNA preceding and/or coincident with the peak of chloroplast gene transcription. Indeed, this is what is observed for *RPOD* mRNA, which typically peaked in the mid to late dark period, right before or during (in the case of a late dark period peak) the peak in chloroplast transcription around the dark-to-light transition. Moreover, having only a single sigma factor gene facilitates global control of chloroplast transcription by the circadian clock. It is important to note, however, that the amplitude of the *RPOD* mRNA fluctuation during the 24-h LD cycle (2.5-fold) is not as great as the transcription amplitudes of some of the genes analyzed previously, except for *psbA*, which exhibited a relatively low (2–3-fold) amplitude (Hwang et al. 1996). Thus, there are likely to be additional mechanisms involving RPOD, such as differential protein stability, or completely distinct mechanisms (e.g. Salvador et al. 1998) that overlay the circadian expression of *RPOD* to help achieve robust clock control of chloroplast transcription. Further analysis of *RPOD* may reveal such mechanisms.

Acknowledgements

We thank the Kazusa Research Institute for cDNA clones, and Ryo Kawazoe for subcloning the *RPOD* cDNA from AV392956 into pBluescript. This research was supported by grants from the US Dept. of Energy (DE-FG03-02ER15352), the Robert A. Welch Foundation (F-1164), and the Biotechnology and Biological Sciences Research Council, UK.

References

Allison LA (2000) The role of sigma factors in plastid transcription. Biochimie 82: 537–548

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Asamizu E, Nakamura Y, Sato S, Fukuzawa H and Tabata S (1999) A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags. DNA Res 6: 369–373

Baginsky S, Tiller K and Link G (1997) Transcription factor phosphorylation by a protein kinase associated with chloroplast RNA polymerase from mustard (*Sinapis alba*). Plant Mol Biol 34: 181–189

Beardslee TA, Roy-Chowdhurry S, Jaiswal P, Buhot L, Lerbs-Mache S, Stern DB and Allison LA (2002) A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. Plant J 31: 199–209

Bornberg-Bauer E, Rivals E and Vingron M (1998) Computational approaches to identify leucine zippers. Nucleic Acids Res 26: 2740–2746

Delwiche CF (1999) Tracing the thread of plastid diversity through the tapestry of life. Am Nat 154: S164-S177

Deppmann CD, Acharya A, Rishi V, Wobbes B, Smeekens S, Taparowsky EJ and Vinson C (2004) Dimerization specificity of all 67 B-ZIP motifs in *Arabidopsis thaliana*: a comparison to *Homo sapiens* B-ZIP motifs. Nucleic Acids Res 32: 3435–3445

Emanuelsson O, Nielsen H and von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Prot Sci 8: 978–984

Emanuelsson O, Nielsen H, Brunak S and von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300: 1005–1016

Fujiwara M, Nagashima A, Kanamaru K, Tanaka K and Takahashi H (2000) Three new nuclear genes, *sigD*, *sigE* and *sigF*, encoding putative plastid RNA polymerase σ factors in *Arabidopsis thaliana*. FEBS Lett 481: 47–52

Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA and Kay SA (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. Science 290: 2110–2113

Harris E (1989) The Chlamydomonas Sourcebook. Academic Press, San Diego, California

Herrin DL and Schmidt GW (1988) Rapid, reversible staining of northern blots prior to hybridization. Biotechniques 6: 196–200

Herrin DL, Michaels AS and Paul A-L (1986) Regulation of genes encoding the large subunit of ribulose-1,5-bisphosphate carboxylase and the photosystem II polypeptides D-1 and D-2 during the cell cycle of *Chlamydomonas reinhardtii*. J Cell Biol 103: 1837–1845

Hess WR and Borner T (1999) Organellar RNA polymerases of higher plants. Int Rev Cytol 190: 1–59

Hwang S and Herrin DL (1994) Control of *lhc* transcription by the circadian clock in *Chlamydomonas reinhardtii*. Plant Mol Biol 26: 557–569

Hwang S, Kawazoe R and Herrin DL (1996) Transcription of *tufA* and other chloroplast- encoded genes is regulated by a circadian clock in *Chlamydomonas*. Proc Natl Acad Sci USA 93: 996–1000

Isono K, Shimizu M, Yoshimoto K, Niwa Y, Satoh K, Yokota A and Kobayashi H (1997) Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of σ^{70} factors of bacterial RNA polymerases in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 94: 14948–14953

Kanamaru K, Fujiwara M, Seki M, Katagiri T, Nakamura M, Mochizuki N, Nagatani A, Shinozaki K, Tanaka K and Takahashi H (1999) Plastidic RNA polymerase sigma factors in *Arabidopsis*. Plant Cell Physiol 40: 832–842

Kanamaru K, Nagashima A, Fujiwara M, Shimada H, Shirano Y, Nakabayashi K, Shibata D, Tanaka K and Takahashi H (2001) An *Arabidopsis* sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. Plant Cell Physiol 42: 1034–1043

Kawazoe R, Hwang S and Herrin DL (2000) Requirement for cytoplasmic protein synthesis during circadian peaks of transcription of chloroplast-encoded genes. Plant Mol Biol 44: 699–709

Lahiri SD, Yao J, McCumbers C and Allison LA (1999) Tissue-specific and light-dependent expression within a family of nuclear-encoded sigma-like factors from *Zea mays*. Mol Cell Bio Res Comm 1: 14–20

Leu S, White D and Michaels A (1990) Cell cycle-dependent transcriptional and post-transcriptional regulation of chloroplast gene expression in *Chlamydomonas reinhardtii*. Biochim Biophys Acta 1049: 311–317

Morikawa K, Ito S, Tsunoyama Y, Nakahira Y, Shiina T and Toyoshima Y (1999) Circadian-regulated expression of a nuclear-encoded plastid sigma factor (*sigA*) in wheat seedlings. FEBS Lett 451: 275–278

Oikawa K, Fujiwara M, Nakazato E, Tanaka K and Takahasi H (2000) Characterization of two plastid σ factors, SigA1 and SigA2, that mainly function in matured chloroplasts of *Nicotiana tabacum*. Gene 261: 221–228

Ogrzewalla K, Piotrowski M, Reinbothe S and Link G (2002) The plastid transcription kinase from mustard (*Sinapis alba* L.). A nuclear-encoded CK2-type chloroplast enzyme with redox-sensitive function. Eur J Biochem 269: 3329–3337

Paget MSB and Helmann JD (2003) The σ^{70} family of sigma factors. Genome Biol. 4(1): 203

Pratt LA and Silhavy TJ (1996) The response regulator *sprE* controls the stability of *rpoS*. Proc Natl Acad Sci USA 93: 2488–2492

Purton S and Rochaix J-D (1995) Characterization of the *ARG7* gene of *Chlamydomonas reinhardtii* and its application to nuclear transformation. Eur J Phycol 30: 141–148

Rechsteiner M and Rogers SW (1996) PEST sequences and regulation by proteolysis. Trends Biochem Sci 21: 267–271

Rivier C, Goldschmidt-Clermont M and Rochaix JD (2001) Identification of an RNA-protein complex in chloroplast group II intron *trans*-splicing in *Chlamydomonas reinhardtii*. EMBO J 20: 1765–1773

Rochaix J-D, Mayfield S, Goldschmidt-Clermont M and Erickson J (1988) Molecular biology of *Chlamydomonas*. In: Shaw JN (ed) Plant Molecular Biology: a Practical Approach, pp 253–257. IRL Press, Oxford

Rost B (1996) PredictPotein. Meth Enzymol 266: 525–539

Salvador ML, Klein U, and Bogorad L (1993) Light-regulated and endogenous fluctuations of chloroplast transcript levels in *Chlamydomonas*. Regulation by transcription and RNA degradation. Plant J 3: 213–219

Salvador ML, Klein U and Bogorad L (1998) Endogenous fluctuations of DNA topology in the chloroplast of *Chlamydomonas reinhardtii*. Mol Cell Biol 18: 7235–7242

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Biology: a Laboratory Manual. Cold Spring Harbor Laboratory Press, New York

Schein AI, Kissinger JC and Ungar LH (2001) Chloroplast transit peptide prediction: a peek inside the black box. Nucleic Acids Res 29: e82

Shrager J, Hauser C, Chang C-W, Harris EH, Davies J, McDermott J, Tamse R, Zhang Z and Grossman AR (2003) *Chlamydomonas reinhardtii* genome project. A guide to the generation and use of the cDNA information. Plant Physiol 131: 401–408

Silflow C (1998) Organization of the nuclear genome. In: Rochaix J-D, Goldschmidt-Clermont M and Merchant S (eds) Advances in Photosynthesis. The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas, Vol 7, pp 25–40. Kluwer Academic Publishers, Dordrecht, The Netherlands

Smith AC and Purton S (2002) The transcriptional apparatus of algal plastids. Eur J Phycol 37: 301–311

Surzycki, SJ and Shellenbarger DL (1976) Purification and characterization of a putative sigma factor from *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 73: 3961–3965

Tan S and Troxler RF (1999) Characterization of two chloroplast RNA polymerase sigma factors from *Zea mays*: Photoregulation and differential expression. Proc Natl Acad Sci USA 96: 5316–5321

Tanaka K, Shiina T and Takahashi H (1988) Multiple principal sigma factor homologs in eubacteria: Identification of the '*rpoD* Box'. Science 242: 1040–1042

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Thum KE, Kim M, Morishige DT, Eibl C, Koop H-U and Mullet JE (2001) Analysis of barley chloroplast *psbD* light-responsive promoter elements in transplastomic tobacco. Plant Mol Biol 47: 353–366

Tozawa Y, Tanaka K, Takahashi H and Wakasa K (1998) Nuclear encoding of a plastid σ factor in rice and its tissue- and light-dependent expression. Nucleic Acids Res 26: 415–419

Troxler RF, Zhang F, Hu J and Bogorad L (1994) Evidence that σ factors are components of chloroplast RNA polymerase. Plant Physiol 104: 753–759

Vaistij F, Boudreau E, Lemaire SD, Goldschmidt-Clermont M and Rochaix JD (2000) Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast *psbB*/*psbT*/*psbH* gene cluster in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 97: 14813–14818