*Published online 2 August 2017* *Nucleic Acids Research, 2017, Vol. 45, No. 18* ***10783–10799***
*doi: 10.1093/nar/gkx668*

# Small RNA profiling in *Chlamydomonas*: insights into chloroplast RNA metabolism

**Marina Cavaiuolo, Richard Kuras, Francis-André Wollman, Yves Choquet and Olivier Vallon***

Unité Mixte de Recherche 7141, CNRS/UPMC, Institut de Biologie Physico-Chimique, F-75005 Paris, France

Received April 01, 2017; Revised July 18, 2017; Editorial Decision July 19, 2017; Accepted July 28, 2017

## ABSTRACT

**In *Chlamydomonas reinhardtii,* regulation of chloroplast gene expression is mainly post-transcriptional. It requires nucleus-encoded *trans*-acting protein factors for maturation/stabilization (M factors) or translation (T factors) of specific target mRNAs. We used long- and small-RNA sequencing to generate a detailed map of the transcriptome. Clusters of sRNAs marked the 5′ end of all mature mRNAs. Their absence in M-factor mutants reflects the protection of transcript 5′ end by the cognate factor. Enzymatic removal of 5′-triphosphates allowed identifying those cosRNA that mark a transcription start site. We detected another class of sRNAs derived from low abundance transcripts, antisense to mRNAs. The formation of antisense sRNAs required the presence of the complementary mRNA and was stimulated when translation was inhibited by chloramphenicol or lincomycin. We propose that they derive from degradation of double-stranded RNAs generated by pairing of antisense and sense transcripts, a process normally hindered by the traveling of the ribosomes. In addition, chloramphenicol treatment, by freezing ribosomes on the mRNA, caused the accumulation of 32–34 nt ribosome-protected fragments. Using this '*in vivo* ribosome footprinting', we identified the function and molecular target of two candidate *trans*-acting factors.**

## INTRODUCTION

The chloroplast originates from an ancient photosynthetic cyanobacterium, engulfed by a eukaryotic host cell through endosymbiosis (1). During evolution, the endosymbiont was converted to a modern plastid with most genes of the ancestor either lost or transferred to the nucleus (1,2). In the model green alga *Chlamydomonas reinhardtii*, the 205 kilobase (kb) circular chloroplast chromosome, present in ∼80 copies per cell, harbors 109 genes (3). Most of these genes encode subunits of the photosynthetic apparatus or are involved in the expression of the plastid genome.

At variance with the cyanobacterial progenitor, the steady-state level of Cp transcripts is determined by post-transcriptional regulation of mRNA accumulation rather than by transcriptional control (4). Cp genes can be transcribed as monocistronic or polycistronic transcripts, but the latter are usually processed into monocistronic mRNAs through intercistronic cleavage by endo-ribonucleases and further trimming by exo-ribonucleases. The position of the 5′ end is determined by the binding of gene-specific protein factors (5,6), reviewed in (7,8). Transcription seems to terminate stochastically (9) and the 3′ ends of mature transcripts are generated by processing. They coincide either with stem-loop structures or with the binding site of an RNA-binding protein, both of which are able to stop the progression of 3′ → 5′ exonucleases. For example, by virtue of its tight binding, the maize protein PPR10 controls the formation of the 5′ and 3′ ends of the *atpH/rpl33* and *atpI/psaJ* transcripts, by blocking the progress of 5′ → 3′ and 3′ → 5′ exoribonucleases, respectively (6). A large number of these nucleus-encoded 'Organelle Trans-Acting Factors' (OTAFs) control the maturation/stability (M factors) and the translation (T factors) of Cp mRNAs, in a gene-specific manner. Most OTAFs belong to helical repeat protein families: the PPR, TPR and OPR (Penta-, Tetra- and Octo-tricoPeptide Repeat) proteins carry tandem repeats of a degenerated motif of respectively 35, 34 and 38 amino-acids, reviewed in (10). PPR repeats fold in two antiparallel α-helices, within which amino acids at specific positions interact with one specific nucleotide in the target (8,11). In contrast to land plants, *Chlamydomonas* contains only 14 PPR proteins (12) but >120 OPR proteins. Most OPRs are predicted to be targeted to organelles. While several have been identified as M or T factors (13–22), many still await a functional characterization.

Our current view of plastid transcripts in *Chlamydomonas* is mostly based on dedicated studies by RNA blot and 5′ or 3′ end mapping assays performed on a few genes. A better understanding of Cp RNA metabolism requires the characterisation of the Cp transcriptome on a genome-wide

*To whom correspondence should be addressed. Tel: +33 1 5841 5058; Fax: +33 1 5841 5022; Email: ovallon@ibpc.fr

scale. Here, using high throughput sequencing of small and long RNAs, we present a refined Cp transcriptomic map based on identification of sRNA mapping to primary or secondary (processed) 5′ ends of mRNAs. We show that sequencing of small RNAs (sRNA-Seq) provides a high accuracy in the determination of transcript 5′ ends. By analyzing long and small RNAs under transcriptional and translational inhibition, we could monitor changes in the stability of sense and antisense transcripts and propose specific pathways for their degradation.

## MATERIALS AND METHODS

### Strains and growth conditions

We used 137c-derived WT strains t222+ (CC-5101), CC-4533 and *atpB*-complemented *CC-373* (23) and mutant strains XS1 (*cw15 arg7 mt-*) (24), *mbb1*–222A (25), *mcd1* (26), *tca1* (27,28), *mca1* (28,29), pG-*petA* and *mca1* pG-*petA* (29), *tda1* (14), *mdb1* (30), *mde1* (Drapier D, Ozawa SI and Choquet Y, unpublished results), *PsaATr* (31) and insertion mutants (32) in *PPR1, PPR3, PPR6, OPR105, OPR56, OPR41, OPR24* and *OPR49* (resp. strains LMJ.RY0402.095219, .049122, .127874, .150140, .212388, .085518, .248644 and .253910). Strains were grown in Tris-acetate phosphate (TAP) medium (33) under low light (5–10 $\mu E\ m^{-2}\ s^{-1}$) or in minimum medium under medium light (20 $\mu E\ m^{-2}\ s^{-1}$) with the addition of 5% bubbled $CO_2$(g). Rifampicin was used at 350 $\mu g\ ml^{-1}$, lincomycin at 500 $\mu g\ ml^{-1}$ and chloramphenicol at 250 $\mu g\ ml^{-1}$.

### RNA extraction, Illumina sequencing and data analysis

Total RNA was extracted from 200 ml cultures (2–3 $\times$ $10^6$ cells $ml^{-1}$) according to (34) omitting the use of the aurintricarboxylic acid during extraction. For directional Whole Transcriptome Shotgun Sequencing (WTSS), RNA samples were treated with DNase-I (NEB), then with the Ribo-Zero Plant Kit to remove rRNAs. Libraries were prepared with the Illumina TruSeq Stranded Total RNA Sample Preparation and sequenced (HiSeq2000) at IGA Technology Services (Italy). Reads were mapped to the nuclear (Joint Genome Institute v5.5, chloroplast ('cv11', unpublished) and mitochondrial (CRU03843)) genomes using BWA aln (35), (samse algorithm, two mismatches allowed). For sRNA-Seq, RNA samples were eventually treated with RNA 5′ Polyphosphatase (RPP, Epicentre) to convert triphosphorylated small RNAs to the mono-phosphorylated form, then phenol-chloroform extracted. RPP- and mock-treated samples were sent to Fasteris Life Sciences SA (Switzerland) for sizing on acrylamide gel (<∼50-nt), multiplex library preparation (Illumina Small RNA Sample Preparation Kit) and sequencing (HiSeq2000). sRNAs-Seq reads (11–44 nt) were mapped either with BWA aln (perfect match) or with Bowtie2 (36) to allow soft-clipping. For the mapping of WTSS and sRNA-Seq data, the inverted repeat A (IRa) of the Cp genome was removed. Reads mapping to both the Cp and the nuclear genome were filtered out using SAMtools (37). Reads mapping at multiple locations were attributed randomly by the software. Mapping statistics of all sequencing data are shown in Supplementary Table S1. Alignments were displayed with the Integrative Genomics Viewer (IGV) (38). BEDtools (39) was used to compute coverage and read counts, normalized as reads per million (RPM) or reads Per Kilobase of transcript per Million mapped reads (RPKM) (40). Differential expression analysis was performed with the EdgeR package (41). The three-nt periodicity was determined using the RiboGalaxy tools (42). Raw datasets were deposited in the Short Read Archive (SRA) database as part of BioProject PRJNA379963. Finally, 313 bi-directional WTSS datasets of *C. reinhardtii* were collected from the SRA database. For each dataset, a coverage ratio CDS/non-CDS $\geq$ 20 was set as threshold to eliminate those with excessive rRNA or DNA contamination, resulting in 90 libraries (Supplementary Table S2).

### Annotation of the chloroplast genome

For the identification of transcript ends, we combined WTSS data and sRNA-Seq reads from WT t222+, *atpB*-complemented *CC-373*, *PsaATr*, *mcd1, mbb1* and *mde1*. A 5′ end was assigned where a cluster of organellar RNA (cosRNA) of at least 15 reads, with a sharp 5′ end, was found in correspondence to decreasing or null values of WTSS coverage. The 3′ end of the cosRNA was defined based on the size of the most represented sRNA-Seq read. To complement visual examination, we used the sRNAminer software (59) ($\geq$15 reads; 3′ heterogeneity up to 75%). A Transcription Start Site (TSS) was called when the ratio between RPP- and mock-treated libraries was $\geq$3 (except for the previously mapped TSSs of *WendyA, petA* and *rbcL*). The MotivFinder tool of the IGV program (version 2.3.34) was used to search for the Pribnow box motif 'TATAATAT' (up to four mismatches allowed, except in the first two positions) and of the TTGaca sequences, ∼10 or 35 nt upstream of the TSS, respectively. The position of 3′ ends was assigned (i) from literature, (ii) from circular RT-PCR (cRT-PCR) results (iii) from a strong predicted secondary structure or (iv) at the approximate position where WTSS coverage fell close to 0 (always <4 RPM). Consecutive genes were clustered in a polycistronic unit if WTSS coverage was continuous in between (except for the previously documented polycistronic clusters, *petA-petD* and *rpl36-rpl23*). Repeat regions between cistrons were considered transcribed if coverage by ambiguous reads was continuous on the expected strand.

### Other methods

RNA blots were carried out as described in (34), using PCR-generated DNA probes labeled with digoxigenin (Sigma). For reverse transcription, the first-strand cDNA synthesis kit (Invitrogen) was used. Quantitative PCR (qPCR) was performed using the SsoAdvanced™ universal SYBR® Green supermix (Biorad) according to the manufacturer's instructions. Reactions were run in duplicate in two independent assays. Expression levels relative to the Cp 16S rRNA gene were calculated using the delta-delta Cq method based on PCR efficiency (43). 5′RACE was performed using the GeneRacer Kit (Invitrogen) according to the manufacturer's instructions with and without tobacco acid pyrophosphatase treatment. cRT-PCR was performed

as described in (6). Primers are listed in Supplementary Table S3.

## RESULTS

To determine the boundaries of Cp transcripts on a genome-wide scale, we mapped Illumina WTSS and sRNA-Seq datasets to a newly assembled chloroplast genome ('cv11', kindly provided by S. Gallaher and S. Merchant, UCLA). Our genome browser at http://chlamy-organelles.ibpc.fr/ allows browsing the main mapping results, as well as right-clicking to download the sequence and annotation tracks. We collected bi-directional WTSS datasets from the SRA database and generated directional WTSS from WT strains grown in either mixotrophic or phototrophic conditions (Supplementary Table S1). 47% of directional sequencing reads mapped to the chloroplast vs. 0.6% in bi-directional libraries, due to the use of polyA-RNA. Using a cutoff of ≥1 read per million (RPM), ∼77% of the genome was covered by directional WTSS, with ∼6% transcribed from both strands, indicating the occurrence of 'antisense' transcription. Due to the presence of repeats (3), ∼0.4% of the reads mapped ambiguously, covering 3–4% of the genome on each strand.

### sRNA-Seq reveals footprints of M factors at the 5′ end of most transcripts

The 5′ end of 23 protein-coding genes has been previously described experimentally (20,24,29,34,44–48,50–57). WTSS coverage decreased progressively towards these 5′ ends and rarely reached them (Figure 1A). In contrast, we observed clusters of organellar sRNAs (cosRNA) at or very near the expected 5′ positions (Figure 1B), showing the same characteristics as the 'footprints' that have been shown to mark the binding sites of RNA-binding proteins in other organelles (58–60). This includes a sharp 5′-edge and a more heterogeneous 3′ end (59,61,62). Most of these cosRNA were detected by the software sRNAminer (59). The most abundant cosRNA started exactly at the mature 5′ end of the most abundant transcript, *psbA* (47). Because of an excellent correlation with known mRNA 5′ ends (Table 1; details in Supplementary Table S4), we assumed that the 5′ end of stable transcripts in *Chlamydomonas* will usually be marked by a cosRNA representing the footprint of an M factor. In total, we found 5′ end cosRNAs for 52 of the 75 protein-coding genes and for *tscA* which contains part of the first intron of the *trans*-spliced gene *psaA* (63,75). Taking into account published RNA blots for genes expressed as downstream CDS within of an uncleaved polycistronic transcript (e.g. *psbT, ycf3, ycf4, cemA*), only 11 protein-coding genes, all lowly expressed, failed to show the expected cosRNA at their 5′ end. Other cosRNAs were observed within transcripts but were considered irrelevant to gene annotation.

In contrast to the mono-phosphorylated 5′ end generated by post-transcriptional processing (PTP), the 5′-triphosphorylated sRNAs corresponding to a transcription start site (TSS) can be integrated into the sequencing library only after removal of the 5′-pyrophosphate by 5′RNA polyphosphatase (RPP). Comparison of RPP- and mock-treated samples (Figure 1C) allowed us to identify 23 cosRNAs as marking a TSS in protein-coding genes and *tscA* (Supplementary Table S4). In all cases, the TSS was found 8–10 nt downstream of a conserved Pribnow box motif 'TATAATAT' (64). In total, excluding CDS, the motif was found 67 times. Usually, but not always, a TTGaca sequence was found upstream at a distance compatible with its marking a '-35' motif. For 13 protein-coding genes, a small RPP-dependent cosRNA was found upstream of an abundant RPP-independent 5′-PTP (see *petB* and *psbF* in Figure 1C, and Supplementary Table S4). For these genes, we considered that this usually minor upstream peak marks the TSS. In the case of *petB, psbF* and *psbK*, cRT-PCR identified the TSS only using RPP-treated RNA, validating these assignments (Supplementary Table S4). When we mapped sRNA-Seq reads allowing soft-clipping at the 3′ end, we found that ∼12% of the reads in 5′-end cosRNAs showed addition of one or two nucleotides (mostly A) at the 3′ end. Such 3′ tails are the hallmark of degradation by PNPase, the major 3′→5′ exonuclease of the Cp (65).

The TPR protein MBB1 protects the 5′ ends of the *psbB* and *psbH* transcripts, and RNA blot has shown the absence of the cognate footprints in the mutant (50). We analyzed sRNAs in *mbb1* and in three other M factor mutants, *mde1, mcd1* and *mca1* (22,25,29,66), which respectively lack the *atpE, petD* and *petA* mRNAs. In all mutants, the cognate footprint was missing, except for asingle read in *mcd1* (Figure 2) and *mca1*. sRNAs originating from other regions of the transcript were less severely affected, as expected if they represent degradation intermediates of a 5′-destabilized transcript. Supplementary Table S6 lists known M-factor and the cosRNAs tentatively assigned as their footprint.

In land plants footprints of RNA-binding proteins may coincide with Cp transcript 3′ ends (6), reviewed in (8,67). In *Chlamydomonas* no M factor has been found so far that targets a 3′ end. In contrast, stem-loops or secondary structures were shown to define the 3′ end of *atpB* (68), *rbcL* (69,70) and *psaB* (70,71). We found strong stem loops downstream of 33 genes (Supplementary Table S5), which overlapped with 22 unique cosRNAs, some in proximity to known 3′ ends (e.g. *psbI, psbA*). We therefore combined secondary structure prediction and cosRNAs to map the 3′ end of stable transcripts (Supplementary Figure S1 and Table S5).

### sRNAs mapping to tRNAs and rRNAs

A significant fraction of the Cp sRNAs (∼70%) mapped to tRNA genes, with a peak length of 32 nt (Supplementary Figure S2). They often started exactly at the mature 5′ end of the tRNA and most appear to result from cleavage in the anticodon loop, as reported earlier (72). A fraction of them carried A-tails at the 3′ end, suggesting participation of PNPase in tRNA degradation. Because mature tRNAs carry an added CCA sequence at their 3′ end, the corresponding 3′-cleavage products could only be identified after *in silico* selection and trimming of reads ending in CCA.

Interestingly, cosRNAs with characteristics of a TSS were found upstream of the *rrnS* gene and of 23 of the 29 tRNA genes (Supplementary Table S4, Figure S3). Because the

**Table 1.** cosRNA defining 5′ ends of protein-coding genes and *tscA*

| Gene | Strand | Type | 5′ end | Major small RNA in cosRNA |
|---|---|---|---|---|
| *wendyA** | - | TSS | 997 | AAATGTATTTAAAATTTTTCAACAAT |
| *petA** | + | TSS | 2640 | GAGAAGAAAAAAAATAAAAT |
| *petD** | + | 5PTP | 6038 | TTTAGCATGTAAACATTAGAAATA |
| *chlB* | + | TSS | 7616 | AATTATCAGGCAAAAACT |
| | | 5PTP | 8507 | CAATAGGCGAGACAACTGT |
| *psbK** | + | 5PTP | 11778 | TTTTATTTTAGAAAGAAAAAACGAGCTTTAAGGTGAGCTTA |
| *tufA** | + | TSS | 12637 | AACAGAACTACTGTAGTTTT |
| | | 5PTP | 12669 | TAAACCTGAAAAATTGGATTATATAGC |
| *rpl20* | - | 5PTP | 16556 | AACCAATCGTCTGTTGCAGT |
| *petB** | - | 5PTP | 20924 | GAAAGCCTAATGGTCATGTCAC |
| | | TSS | 20977 | ATAACTTTAATTTAAACT |
| *chlL* | + | TSS | 20988 | ATATATAAAATAAAAAAAACGTTAGTAATTC |
| *rpl36* | + | 5PTP | 22431 | TATAATTTAGGAAATT |
| *rpl16* | + | 5PTP | 27653 | TTTTAAAGTTGCTTGTTTTATA |
| *rpl14* | + | 5PTP | 28940 | TAGAATGACTAAAAGGAGT |
| *rps8* | + | 5PTP | 31812 | ATGGACTGCTATAATATAAGAAT |
| *psaA-1** | + | TSS | 33024 | ATATGATGTAAAAAAAACTATTTGTCT |
| *rps4* | + | 5PTP | 33980 | GTTAATTCATTAAAGCCGTTTATTTAAA |
| *psbA.1** | - | TSS | 56159 | ACCATGCTTTTAATAGAAG |
| | | 5PTP | 56106 | TTTACGGAGAAATTAAAAC |
| *psb30/ycf12* | - | 5PTP | 60455 | TGTTACTTTTTGATTTTGTATATA |
| *atpE** | - | 5PTP | 61465 | TTAAAGTATAGTTCAGAATATT |
| *rps7* | - | 5PTP | 62896 | AAAATTGCTTATTTGGTATG |
| *rps14* | - | 5PTP | 63641 | AGATAAATCGTGTCAGTTTTTGAATTGATAGC |
| *psbM* | - | 5PTP | 65306 | ACCTTTAGATCTCTGCATAGAGTATTTCCT |
| *psbZ** | - | 5PTP | 67304 | TTTTTCCTTTTTAGGTTCTATTTACAAAAGGATG |
| | | TSS | 68215 | ATAACATTAAAATTTTTGAACT |
| *psaA-3* | - | TSS | 72653 | ATAACACATTTATTTAAAAACAGCAAAAACTTGC |
| *wendyB* | - | TSS | 76235 | AAATGTATTTAAAATTTTTCAAAAATTTTTAAA |
| *psbH** | - | 5PTP | 77778 | TTTACAGAAAGTAAATAAAATAGCGCT |
| *psbN* | + | 5PTP | 78263 | AAAGAGAATAATTTTATTATTAAATG |
| *psbB** | - | 5PTP | 82553 | AATAATTAAGTAAAAAAATC |
| | | TSS | 82659 | AATTTAATTTAAAATCTTAAAAAAT |
| *rpoA* | - | 5PTP | 87322 | ATATAGCTAAAATGGACT |
| *rps2* | - | 5PTP | 91164 | AAGGGAAGTCTACTAACTC |
| *rps9* | - | 5PTP | 95104 | ATAAGATATATATAGGAG |
| *psbE** | - | 5PTP | 95610 | AAAGCAGACAAATTGTTGAAAAAGC |
| | | TSS | 95626 | ATAATACATTGATTATAAAGCAGACAAATTG |
| *rpoB2* | - | 5PTP | 98757 | TATTAGGAAATTACAAATTATATTACA |
| *rpoB1* | - | 5PTP | 102356 | ATACCTTTCTTTAAAACTAACCTAACAATTAGG |
| *psbF** | + | TSS | 102839 | ATTATATTTATTTTAAACTAATATT |
| | | 5PTP | 102877 | AACGAGTTAGCTTAATACAAAA |
| *petG** | + | 5PTP | 104047 | TCTTGAAGTGTGATGACTC |
| *rps3* | + | 5PTP | 104740 | TTATTAACGTATGGGAACCTTTTACT |
| *rpoC2* | + | 5PTP | 108567 | TTTTTCTGTTTTTTGTTTGTTAT |
| *psaB** | + | TSS | 119548 | ATATGTAATTAATCTGAAAATAGATTACT |
| | | 5PTP | 120085 | ACAGGATTATGGCGTAGTC |
| *rbcL** | - | TSS | 124868 | AAATGTATTTAAAATTTTTCAACAAT |
| *atpA** | + | TSS | 125209 | ACTATATAAATACATTTACC |
| | | 5PTP | 125243 | TTTACCTTTTTTTTAATTTGCATGATTTTAATGC |
| *psbI** | + | 5PTP | 127439 | ATTACTTTGTATATATAAACCAAAGTA |
| *atpH** | + | TSS | 129770 | GGTTGTTATCGATTTTATTGA |
| *tscA** | + | TSS | 136017 | AAGTGAAAAAATTAAAAATAAATAATG |
| *chlN* | + | 5PTP | 136635 | GTAAGTTTGAATACATTTAGT |
| *psbA.2** | + | TSS | 139566 | ACCATGCTTTTAATAGAAG |
| | | 5PTP | 139619 | TTTACGGAGAAATTAAAAC |
| *atpB** | - | TSS | 162779 | ATATATATAGTTAAATGAAAAAAC |
| | | 5PTP | 162753 | AAAAATAAGCGTTAGTGAATAA |
| *ycf1/orf1995* | - | TSS | 170076 | AAGTTTAAAAGTTATAGAATTTT |
| *rps12* | - | 5PTP | 171921 | ACATGATGTGGAATCATTT |
| *atpI** | - | 5PTP | 173682 | CTTTTGCATCAATCCATAGGATTGTATATACCA |
| *psbJ** | - | 5PTP | 175058 | AACGGCTCTTATTTTTTTAATAAGT |
| *psbD** | - | 5PTP | 177235 | AATTTAACGTAACGATGAGTTGTT |
| | | TSS | 177263 | ACACAATGATTAAAAT |
| *ycf2/orf2971* | + | TSS | 177492 | AGGAAAAAATTTAAAATTTAAAATGTTAGT |
| *psbC** | + | 5PTP | 188039 | TTTAAGTGTTACAAAGAAATTGAA |
| *psaC** | + | 5PTP | 191376 | GTCGATTCTCAATCTTCTTTTTG |

The asterisks indicate mRNAs whose 5′ end had been previously determined.

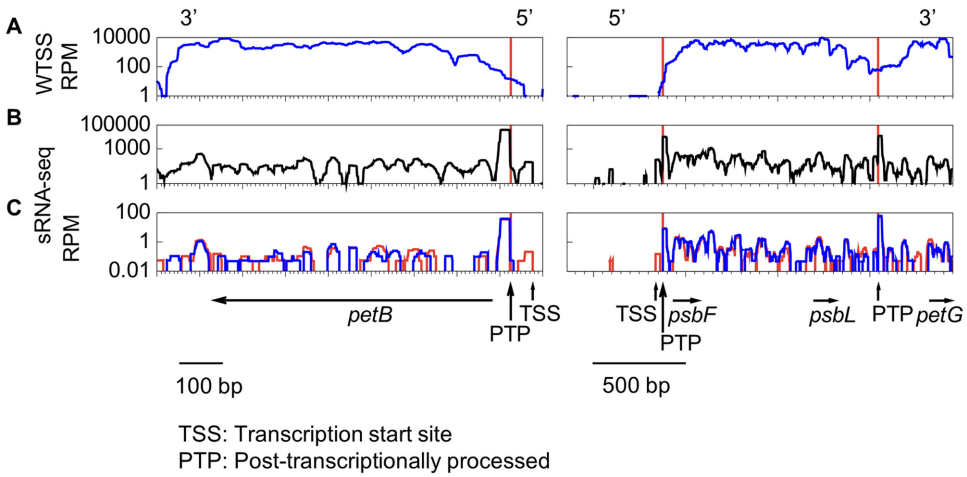


**Figure 1.** Transcriptional profile over *petB* and the *psbF-psbL-petG* polycistronic unit. Coverage (log scale) of (**A**) pooled bi-directional and directional WTSS; (**B**) pooled sRNA-Seq; (**C**) mock- (blue) versus RPP-treated (red) WT sRNA-Seq libraries. Red vertical lines indicate the position of the mature 5′ ends. Vertical arrows point to cosRNAs marking a TSS or PTP. In A and C, coverage is normalized as RPM.

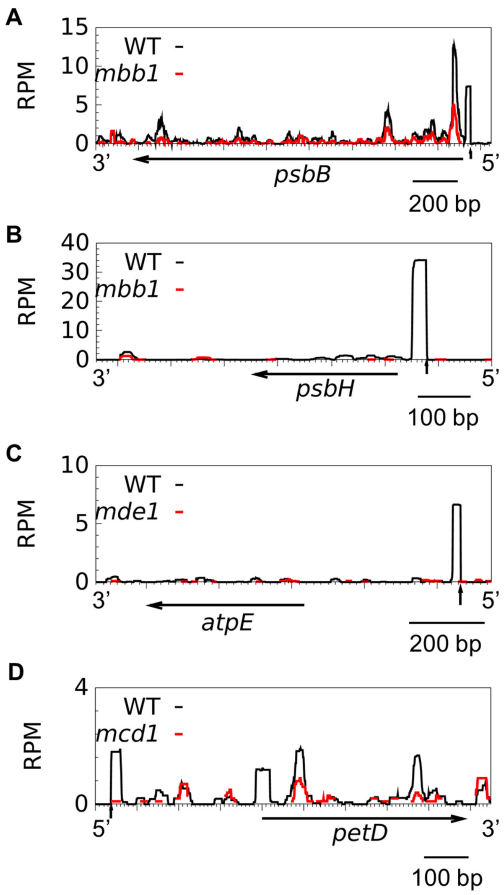


**Figure 2.** 5′-cosRNA are footprints of M factors. Distribution of sRNAs in WT (black) and mutant strains (red) over their target genes. Vertical arrows indicate the position of the mature 5′end in the WT. Coverage is expressed in RPM and averaged over two biological replicates.

cosRNAs lie downstream of appropriately spaced '-10' and '-35' sequences, we considered them as resulting from transcription initiation. Except for *trnK*, they were close to the mature tRNA 5′ end (usually 10–20 nt). Combining the sRNA-Seq and WTSS data, we also identified a 3′ extension for 15 of these putative tRNA precursors, Finally, one of the most abundant cosRNA lies 79 nt upstream of *rrn7*. It is RPP-independent, suggesting that the transcript originating at the *trnI* promoter is cleaved between *trnA* and *rrn7* and that the cleavage product is stabilized by the binding of an M factor before its final maturation. A cosRNA is found at a similar position in *Arabidopsis* (73).

### Co-transcription is widespread in the *Chlamydomonas* chloroplast

A few monocistronic genes such as *petB* (Figure 1A) showed distinct boundaries with null WTSS coverage at the 5′ and 3′ ends. But most of the times, the region between coding sequences (CDS) located on the same strand showed uninterrupted coverage (e.g. *psbF/psbL/petG* in Figure 1A, others in Supplementary Figure S4), indicating that they are co-transcribed in a polycistronic precursor. Based on this analysis and on the location of transcription start sites and promoter sequences in intergenic regions, we found many hitherto overlooked cases of co-transcription (Supplementary Table S4). In total we grouped 84 of the 109 genes into 22 polycistronic units. For 20 genes, RNA blot-based evidence for co-transcription is lacking but in some cases (especially for tRNAs) this is likely due to the efficient processing of the precursors. As an example of co-transcription, the tetracistronic *psbJ/atpI/psaJ/rps12* mRNA (74) is probably co-transcribed with the upstream bi-cistronic *psbD/psaA-2* from the *psbD* promoter (Supplementary Figure S4A) as observed in *trans*-splicing mutants (76). Similarly, we found evidence for a fusion of the *psbZ-psbM* (49) and *rps7-atpE* (77) clusters, also including *ycf12*. We extended the *rps9-ycf4-ycf3-rps18* cluster (78) to include *psbE* upstream and *rps2* downstream. The *rpl36-rpl23-rpl2-rps19* cluster (79) was fused with *chlL* upstream (Supplementary Figure S4B) and the *rpl16-rpl14-rpl5-rps8* genes downstream. Some clusters started with a tRNA gene, others contained an internal

tRNA gene marked by a TSS. The *atpH* promoter within the *atpA* cluster (34) was also marked by a TSS.

### Transcripts with stalled ribosomes yield ribosome-protected fragments

We wondered whether the stability of Cp transcripts, and hence the production of sRNAs, was affected by their association with ribosomes, as observed in prokaryotes. We analyzed RNA prepared from mixotrophically-growing cells treated for 10 min with lincomycin (Linc) or for 15 min or 3 h with chloramphenicol (CAP). Linc inhibits translation shortly after initiation by blocking the peptide exit channel, but allows previously engaged ribosomes to continue translation until they reach the stop codon (80). In contrast, CAP stops elongation by occupying the position of the amino acid attached to the tRNA in the A-site, thus forcing the ribosome to stall on the mRNA (81). Table 2 shows the general effects of these inhibitors on different types of Cp regions: UTRs, CDS, introns and intercistronic/intergenic regions. Since *psbA* is the most abundant transcript, whose 5′UTR alone generated ∼90% of all Cp sRNAs mapping to 5′UTRs, it was excluded from our general description and analyzed separately in Supplementary Tables S7, S8 and Figure S5.

As judged from directional WTSS, Linc treatment had no general effect on mRNA levels (Table 2). Differential expression analysis identified only one gene, *rpl23*, as significantly upregulated (Supplementary Table S9), which is interesting considering that Rpl23 lines the polypeptide exit site (82). Linc also had practically no effect on sRNA coverage, except for an increase over *rpl23* and the downstream *rpl2*. CAP treatment, in contrast, led to a marked increase in sRNA coverage over most CDS and many UTRs. This effect was already apparent after 15 min and was exacerbated after 3 h, at which time 61 CDS showed a significant increase (>2-fold) in sRNA coverage (Supplementary Table S9). Interestingly, the newly-accumulating sRNAs were mostly in the range of 32–34 nt (Figure 3). Based on the size of *in vitro*-generated ribosome footprints characterized in other studies, e.g. ∼30–35 nt in Cp of maize (83), 27–35 in *Chlamydomonas* Cp (84), we tentatively assigned this population to *in vivo*-generated ribosome-protected mRNA fragments (RPFs). Looking for a relationship between the frame and the position of the RPFs (85), we observed that the difference in RPF sizes was mostly due to a variable extent of trimming at the 5′ and 3′ ends (Figure 4A). For example, in the T0 and CAP15min samples, the 33 nt RPFs started mostly at frame 0 and the 32 nt RPFs at frame 1, indicating variability in the trimming on the 5′ side of the ribosome. In the CAP3h samples, the 32 nt and 33 nt RPFs started at frame 0, suggesting variable trimming at the 3′ side.

After 3h of CAP treatment, the proportion of RPFs over CDS among the total sRNAs increased from 8% in the control to 58% (Supplementary Table S10). For individual genes, it ranged from extremely low (the purely intronic *tscA,* the probably untranslated *WendyB*, the intronic *orf5* in *psbA*) to 89% (*psbK*, where it was already 54% in the control). The non-conserved *orf58* (86) showed no RPFs which we take as an indication that it is probably not translated

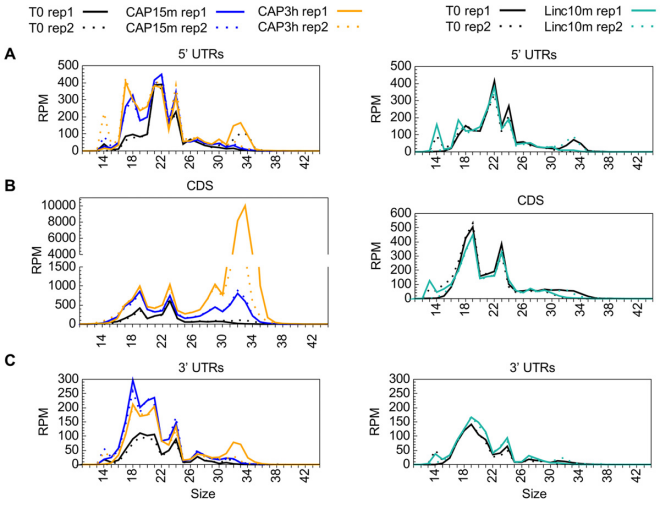


**Figure 3.** Size distribution of sRNAs in control, CAP- and Linc-treated samples, over 5′UTRs (**A**), CDS (**B**) and 3′UTRs (**C**). For each experiment, the two replicates are shown. *psbA* was excluded from this representation.

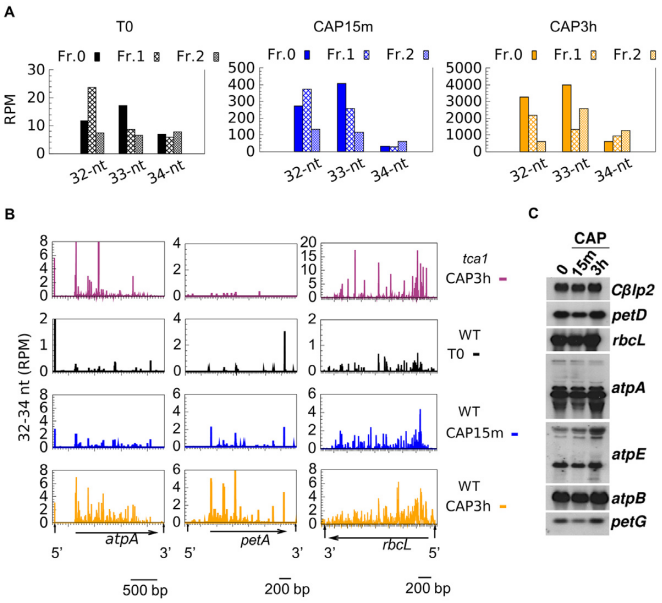


**Figure 4.** Characteristics of RPFs (32–34 nt) in control and CAP-treated samples. (**A**) Position of the 5′ end of 32–34-nt reads respective to the reading frame (all CDS). (**B**) Profiles of 5′-end positions of 32–34 nt sRNAs sequences (average of two replicates) in CDS (horizontals arrow) and untranslated regions (marked by vertical arrows) of *atpA*, *petA* and *rbcL*. (**C**) Accumulation of chloroplast transcripts upon CAP treatments, with the nuclear *Cβlp2* gene as a loading control.

into a protein. The distribution of RPF 5′ ends (Figure 4B, others in Supplementary Figure S6) was not homogeneous, with the strongest peaks usually observed around the start codon and in the 5′ part of the CDS, consistent with the notion that ribosomes travel more slowly during the initial phases of translation (85,87). In the *tca1* mutant that is unable to translate the *petA* gene (27,28) CAP treatment led to an increase in 32–34 nt sRNAs over all CDS except *petA* (Figure 4B), confirming that their production indeed requires translation.

**Table 2.** The effects of translational inhibition on the accumulation of Cp small RNA and RNA

| Cp region | | WTSS (RPMx10∧3) | | sRNA-Seq (RPM × $10^3$) | | | | |
|---|---|---|---|---|---|---|---|---|
| Length (kb) | Type | T0 | Linc 10 min | T0 | Linc 10 min | T0 | CAP 15 min | CAP 3 h |
| 21 | *5′UTRs* | 43±8 | 67±17 | 1.9±0.1 | 1.9±0 | 1.9±0.1 | 2.8±0.05 | 3.3±0.09 |
| 88 | *CDS* | 542±67 | 683±70 | 2±0 | 1.9±0 | 1.6±0.1 | 6.4±0.1 | 23±12 |
| 13 | *3′UTRs* | 32±8 | 51±10 | 0.8±0 | 1±0 | 0.7±0.05 | 1.7±0.02 | 1.5±0 |
| 0.1 | *Introns* | 0.2±0 | 0.3±0.1 | 0.006±0 | 0.0±0 | 0.0±0 | 0.0±0.0 | 0.006±0 |
| 40 | *Intercistronic* | 42±6 | 69±13 | 5.2±1 | 7±1 | 5.5±1.2 | 6.3±0.03 | 8±0.3 |
| 2 | *mature tRNA* | 28±6 | 42±8 | 49±13 | 57±15 | 62±16 | 81.8±0.4 | 130±20 |
| 5 | *mature rRNA* | 7±0.7 | 5.6±0.3 | 28±4 | 40±1.8 | 25±5 | 24±0.7 | 34±0.9 |

sRNA-Seq and WTSS coverage, reported as the sum of the reads normalized to the total mapped to the nuclear and mitochondrial genomes for each Cp region and averaged between two biological replicates (RPM±SD). Introns are for *psaA*. The inverted repeat A, *psbA* and intercistronic regions between convergent gene units are excluded.

The *in vivo* generation of RPFs by CAP is likely due to endonucleolytic cleavage in the region between the stalled ribosomes, followed by 5′→3′ and 3′→5′ exonucleolytic trimming that determines the size of the RPFs. However, RNA blot for three highly (*rbcL*, *atpB* and *atpA*) and three moderately (*petD*, *petG* and *atpE)* abundant transcripts showed no significant change in transcript levels after 15min or 3h of CAP treatment, compared to a control nuclear mRNA (Figure 4C), indicating that ongoing transcription compensates for transcript cleavage between the stalled ribosomes.

### The expression level of Cp protein coding genes is generally not affected by growth conditions

When we analyzed the expression level of Cp genes in cells grown in mixo- or photo-trophic conditions by directional WTSS, we found that the relative abundance of Cp transcripts was highly correlated between the two growth conditions, with a Pearson's coefficient ($R^2$) of 0.97 (Figure 5). Only 7 genes were identified as differentially expressed, at a False Discover Rate (FDR) ≤ 0.05 (Supplementary Table S11): the *WendyA* transposon, *orf528*, *tufA*, *rps11*, *rps12* and two homing endonucleases encoded by the *psbA* and *rrnL* introns (88,89). We averaged the RPKMs between the two growth conditions and arbitrarily classified genes into low (RPKM < 1000), moderate (1000 ≤ RPKM < 10000) and high (RPKM ≥ 10 000) expression categories (Supplementary Table S12). The most highly expressed genes were those encoding the major subunits of photosynthetic proteins. The only photosynthetic gene present in the low category was *psbI*. Differentially expressed genes were all in the 'moderate' or 'low' category.

Interestingly, analysis of nuclear gene expression in the same samples yielded a completely different picture. Using the same criteria, 37% of nuclear genes showed differential expression between mixotrophic and phototrophic conditions (Supplementary Table S13). In particular, many helical repeat protein genes among which most known M factors (RAT1, RAT2, MRL1, NAC2, MCD1, MBB1, MCA1, RAA4, MCG1, PPR1, MAC1) were less expressed in phototrophic conditions.

### Inhibition of transcription reduces mRNA and sRNA levels differentially in mixotrophic and phototrophic conditions

To assess the stability of transcripts, we treated the cells with rifampicin (Rif), a specific inhibitor of transcription initiation in bacteria and Cp (90). After 6 h in Rif, WTSS showed a large decrease in coverage over all Cp regions and genes (Table 3, Supplementary Table S8 for *psbA*). Over two thirds of individual regions showed a >2-fold decrease that could be considered significant at a FDR ≤ 0.05 (Supplementary Table S9). For most genes, the effect was stronger in phototrophic than in mixotrophic conditions (Figure 6A and B), as shown previously for several photosynthetic genes (4). Overall, photosynthesis genes were less affected than genes involved in translation, transcription or other functions (Figure 6C), correlating with their higher abundance (Figure 6A). Genes from a same polycistronic unit often displayed a very different sensitivity to Rif-treatment (e.g. in the *atpA*, *psbB* and *psbD* clusters).

At the sRNA level, we observed a decrease in coverage for all type of Cp regions (Table 3), but in contrast to the WTSS data, the Rif-induced decrease of sRNA was less pronounced in phototrophic than in mixotrophic cells, es-

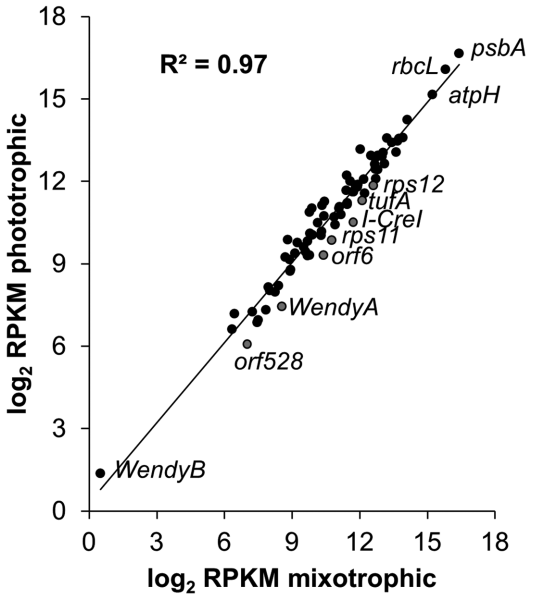


**Figure 5.** Gene expression in mixotrophic and phototrophic conditions ($log_2$ transformed RPKM values). RPKM values were computed on CDS, on the three exons of *psaA* and on the *WendyB* and *tscA* genes. The value for *psbA* is the average of the RPKM computed independently for the five exons. Differentially expressed genes are in grey.

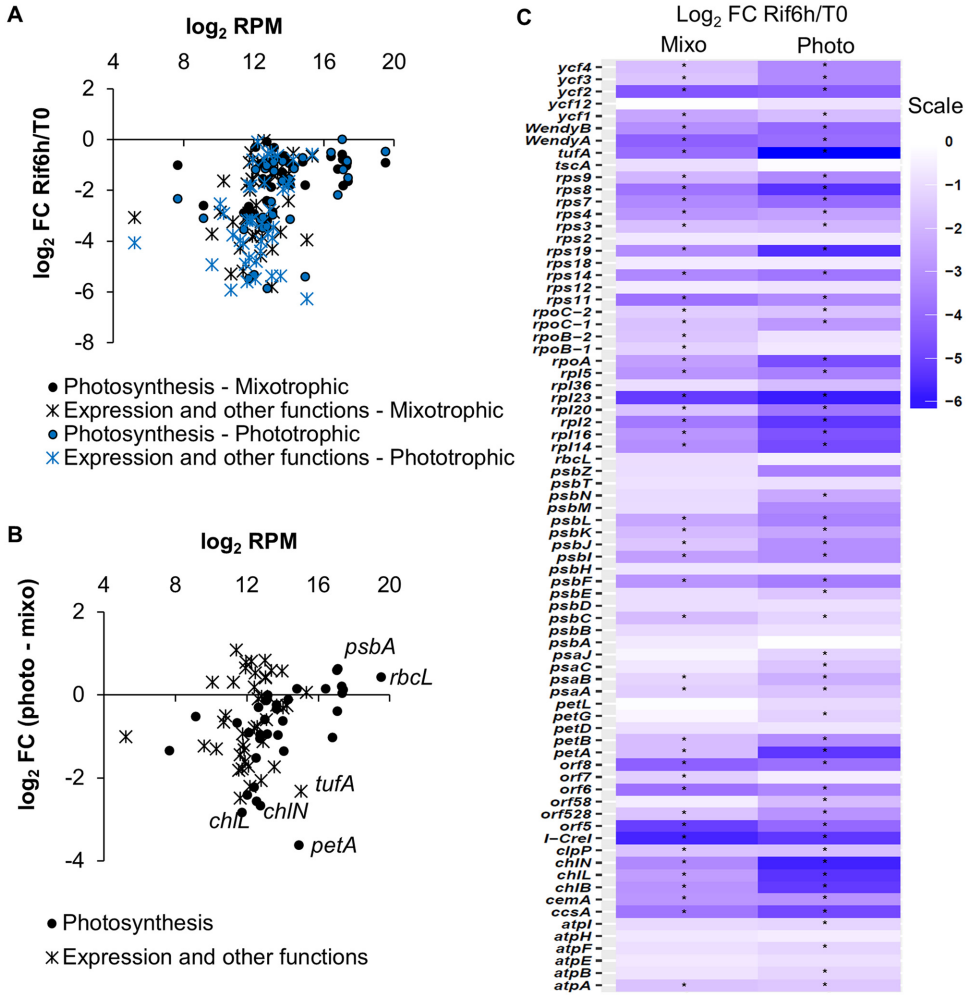


**Figure 6.** Cp transcripts display different stability between mixotrophic and phototrophic growth. (**A**) MA plot reporting the $log_2$FC (Rif-treated over control) against $log_2$RPM of the RNA levels, distinguishing genes involved in photosynthesis (circle) or other functions (cross) in mixotrophic (black) and phototrophic (blue) condition. (**B**) Difference between $log_2$FC in the two conditions. (**C**) Heatmap of the $log_2$FC. Genes with significant changes at FDR $\leq 0.05$ are marked with an asterisk.

**Table 3.** The effects of transcriptional inhibition on the accumulation of Cp small RNA and RNA

| Cp region | | WTSS (RPM × $10^3$) | | | | | | sRNA-Seq (RPM × $10^3$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mixotrophic | | | Phototrophic | | | Mixotrophic | | | Phototrophic | | |
| Length (~kb) | Type | T0 | Rif 6 h | $log_2$FC | T0 | Rif 6 h | $log_2$FC | T0 | Rif 6 h | $log_2$FC | T0 | Rif 6 h | $log_2$FC |
| 21 | *5′ UTRs* | 43±8 | 17±5 | -1.2 | 46±9 | 7.7±0.8 | -2.6 | 1.9±0.1 | 0.5±0 | -1.8 | 0.9±0 | 0.4±0 | -1.3 |
| 88 | *CDS* | 542±67 | 226±48 | -1.2 | 573±61 | 169±16.5 | -1.8 | 2±0 | 0.4±0 | -2.3 | 1.4±0.1 | 1.0±0.1 | -0.5 |
| 13 | *3′ UTRs* | 32±8 | 13±4 | -1.2 | 34±6 | 8.0±0.2 | -2.1 | 0.8±0 | 0.1±0 | -2.3 | 0.5±0 | 0.1±0 | -1.9 |
| 0.1 | *Introns* | 0.2±0 | 0.05±0 | -2.2 | 0.27±0 | 0.03±0 | -3.0 | 0.0±0 | 0.0±0 | -2.9 | 0.0±0 | 0.0±0 | -1.9 |
| 40 | *Intercistronic* | 42±6 | 12±3 | -1.7 | 47±8 | 7.2±0.5 | -2.7 | 5.2±1 | 1.8±0 | -1.4 | 4±1 | 2.3±0.3 | -1.0 |
| 2 | *mature tRNA* | 28±6 | 11±3 | -1.4 | 28±5 | 7.0±0.1 | -2.0 | 49±13 | 31±0.5 | -0.6 | 31±3 | 38±4 | 0.3 |
| 5 | *mature rRNA* | 7±0.7 | 1.5±0.3 | -2.2 | 6±0.7 | 1.5±0.6 | -2.2 | 28±4 | 1.9±0.2 | -1.1 | 28±0.5 | 18±1 | -0.6 |

sRNA-Seq and WTSS coverage, reported as the sum of the reads normalized to the total mapped to the nuclear and mitochondrial genomes for each Cp region and averaged between two biological replicates (RPM±SD). Introns are for *psaA*. The inverted repeat A, *psbA* and intercistronic regions between convergent gene units are excluded. The $log_2$ Fold-Change (FC) between Rif- and control is indicated.

pecially over CDS (Figure 7; Supplementary Table S9, Supplementary Table S8 for *psbA*). This was in part explained by the appearance over most of the CDS of a population of 32–34 nt sRNAs reminiscent of the CAP-induced RPFs (Figure 7, Supplementary Figure S5, Supplementary Table S10). The Rif-induced 32–34 nt were distributed throughout the CDS of most genes (Figure 8A and B) and showed a three-nt periodicity (Figure 8C) as expected from RPFs. They may be caused by ribosomes stalled at the 3′ end of a truncated CDS generated by partial degradation.

### Antisense sRNAs accumulate when translation is inhibited

In bacteria antisense RNAs (asRNAs) are implicated in fine-tuning of gene expression (91): asRNAs transcribed from the complementary strand of a gene can base-pair with the corresponding mRNA, modifying its stability and/or

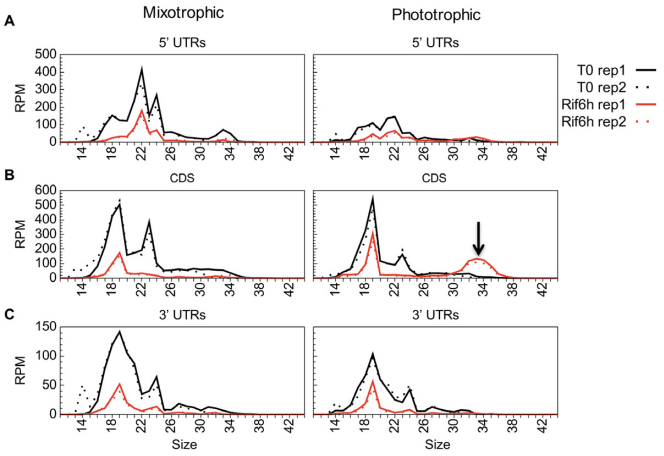


**Figure 7.** Size distribution of small RNAs in Rif-treated samples. Abundance over 5′ UTRs (**A**), CDS (**B**) and 3′ UTRs (**C**). For each experiment, the two replicates are shown. *psbA* was excluded from this representation.

translational efficiency (92). Such asRNAs were identified throughout the Cp genome of land plants (93,94) but only a few have a proposed function. The chloroplast-encoded AS5, whose over-expression leads to decreased 5S rRNA stability, has been proposed to prevent the accumulation of misprocessed 5S rRNA (95). An asRNA to *psbT* was shown to base-pair with *psbT* mRNA causing its translational inactivation by blocking the access of the ribosome (96) and allowing the processing of the *psbT-psbH* intergenic region (97). Finally, asRNAs that over-accumulated in the *Arabidopsis* RNase J knock-down line form duplexes with mRNAs and prevent their translation (98).

Searching for a possible regulatory role of antisense transcripts or their degradation products in *Chlamydomonas*, we quantitatively profiled both the long and small antisense RNAs (Supplementary Table S14). For protein-coding genes, WTSS coverage on the antisense strand was 100–1000 times lower than on the sense strand and decreased strongly upon Rif treatment. By contrast, antisense sRNA (as-sRNA) coverage was much higher (Figure 9A) and partially resistant to Rif (Figure 10A and Supplementary Figure S7). In control conditions as-sRNAs were in similar amounts or even more abundant than sense sRNAs (s-sRNAs) over many regions (Supplementary Table S14). Strand-specific RT-PCR demonstrated the existence of long antisense RNAs (lg-asRNAs), from which as-sRNAs likely derived, at all the tested loci (*atpA*, *atpB*, *atpI* and *petA,* Figure 9C and D). In the case of *petA*, we identified by 5′-RACE a major 5′ end for an antisense transcript that corresponded precisely to the highest peak of as-sRNAs in Figure 9C. This 5′ end could be amplified without RPP treatment and therefore results from the processing of a longer transcript, in agreement with RT-PCR results (Figure 9D) and with the identification of an antisense promoter in the *petA-petD* intergenic region (99). For *atpB*, qPCR showed that the lg-asRNA accumulated to ∼0.004% of the sense RNA, i.e. even less than predicted from WTSS data. In the *mdb1* mutant that lacks the *atpB* sense transcript (30), the *atpB* sense signal decreased 30-fold compared to WT, as expected, but we observed a 63-fold over-accumulation of the lg-asRNA (Figure 9E). In contrast, sRNA-Seq of the mutant showed a massive reduction not only of the sense but also of the as-sRNAs on *atpB*. These results suggest that degradation of the lg-asRNA and the resulting production of as-sRNAs depend on the presence of the sense transcript to which it can base-pair. Indeed, other M factor mutants showed decreased amounts of as-sRNAs mapping to the cognate target mRNA (Figure 9B and *mca1* in Figure 10A). Conversely, increasing accumulation of the *petA* mRNA by introducing a poly-G tract in its 5′-UTR (29) led to an increase in *petA* as-sRNAs, even in an *mca1* background (Figure 10A).

Base-pairing between sense and antisense transcripts and hence production of as-sRNAs should be favored when ribosomes are prevented from translating the mRNA. Indeed, inhibition of translation by either Linc or CAP led to a marked increase in the abundance of as-sRNAs (Figure 10B; Supplementary Table S15), with no change in size distribution (Supplementary Figure S7). This increase was strongest over CDS (55 showed an increase ≥3-fold after 3h in CAP) but was also observed over non-translated regions. Similarly, the *tda1* mutant that is unable to translate the *atpA* mRNA (14) showed a specific increase in the accumulation of as-sRNAs over *atpA* (Figure 10C), comparable to that obtained in WT after 3 h in CAP.

### Using sRNA-Seq to identify the target of PPR and OPR proteins

Based on the results above, we tried to develop a protocol for rapidly identifying the target and mode of action of an OTAF of unknown function. After CAP-treatment, an M factor mutant is expected to show decreased or null sRNA coverage over the target gene (especially over the footprint), while a pure T factor mutant would simply show absence of the 32–34 nt RPFs over the CDS, sRNAs of other sizes being unaffected. We selected from the CliP mutant collection (32) 8 mutants carrying insertions in a gene encoding a helical repeat protein. For each mutant, the absence of the WT copy of the OTAF gene was confirmed by PCR.

Among the five mutants in OPR genes that we analyzed, only *opr56* showed a strong phenotype, being non-phototrophic with fluorescence induction kinetics typical of PSII mutants. Accordingly, we observed a near disappearance of the sRNA coverage on the gene encoding the PSII core subunit *psbC*, both on the sense and antisense strands (Figure 11A). The *psbC* 5′-PTP cosRNA disappeared completely, as did the mRNA itself (Figure 11C), confirming the assignment of OPR56 as an M factor for *psbC*. The gene was renamed *MBC1* in accordance with the nomenclature of OTAFs in *Chlamydomonas*. The other OPR mutants that we analyzed, located in the *OPR24*, *OPR41*, *OPR49* and *OPR105* genes, displayed no or only mild growth defects and showed only minor changes in sRNA coverage. Their targets thus remain unknown.

PPR1 is the ortholog of land plants HCF152 (12) which in maize controls the splicing and stability of the *petB* mRNA (100). The *Chlamydomonas ppr1* mutant was non-phototrophic and had a fluorescence induction phenotype typical of cytochrome $b_6f$ mutants. sRNA-Seq showed a 25/1.7-fold decrease in sRNA coverage over the *petB* CDS

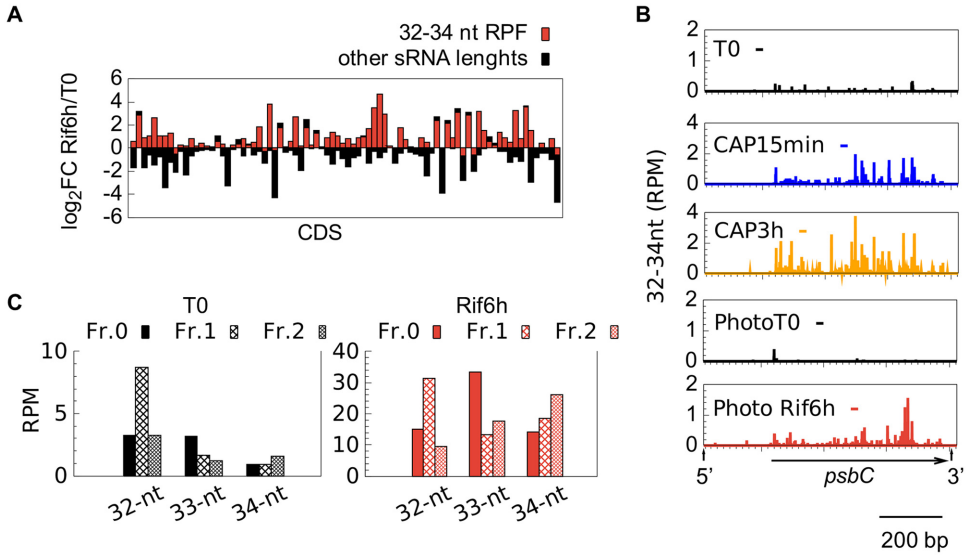


**Figure 8.** Characteristics of RPFs (32–34 nt) after Rif-treatment in phototrophically grown cells. (**A**) $\log_2$FC (Rif-treated over control) of RPF (red) and non-RPF sRNAs (black) for each chloroplast CDS, following order in the genome from *petA* to *WendyA*. (**B**) Profiles of the 5′-end positions of 32–34-nt sequences on *psbC* after CAP or Rif treatment compared to the controls. (**C**) Position of the 5′ end of 32–34-nt reads respective to the reading frame (all CDS).

for RPFs and non-RPFs respectively (Supplementary Table S16), indicating that *petB* is the evolutionarily-conserved target of PPR1 in plants and algae. The stronger effect on RPFs than on non-RPF sRNAs suggests a role in translation, but the mutant showed an overall decrease over all *petB* regions, suggesting a general destabilization of the mRNA. In particular, the *petB* 5′-PTP footprint decreased 71-fold (Figure 11B). In accordance with these data, the *petB* mRNA was severely reduced but still detectable by RNA blot (Figure 11C). cRT-PCR indicated the presence in *ppr1* of precursor transcripts starting at the TSS, while no transcript could be detected carrying the mature 5′ end. We conclude that PPR1 is necessary for translation of the *petB* mRNA and in addition contributes to its stabilization. We therefore propose to rename it *TCB1*. In contrast, mutants in PPR3 and PPR6, two PPR-SmR-cyclins of unknown function (12), showed no growth phenotype and no significant change in sRNA-Seq (Supplementary Table S16) or RNA blots (not shown), including over the candidate targets suggested by the PPR code, *rps4* and *psbF*.

## DISCUSSION

### A description of the *Chlamydomonas* Cp transcriptome

In this work, we used a combination of WTSS and sRNA-Seq to characterize the Cp transcripts of *Chlamydomonas*, in particular to delineate their 5′ ends where stabilizing M factors are expected to bind. By contrast with the more demanding 'Directional mRNA-Seq' Illumina protocol used by (93), which, in *Arabidopsis*, allowed to capture some true 5′ ends (101), the TruSeq methods were not appropriate to define 5′ or 3′ ends. However, a previous data mining study (50) had revealed the presence of sRNA footprints in the Cp of *Chlamydomonas*, similar to those found in higher plants (6,59,60). We therefore used sRNA-Seq to produce a more comprehensive description of 5′ ends. In total, we found 65 cosRNAs marking the 5′ end of protein-coding genes, of which 14 had been previously identified by (50). A cosRNA was found for all 23 genes whose 5′ end had been mapped previously and we confirmed 4 newly identified 5′ ends by cRT-PCR. We conclude that all stable Cp mRNAs in *Chlamydomonas* show a 5′ cosRNA, the likely footprint of an M factor. Because some footprints can be of low abundance (e.g. *petA*, *rbcL*), additional 5′ ends may remain to be discovered.

At all loci tested, the footprint was absent or strongly reduced in the cognate M factor mutant, while a sRNA signal was still detectable over the rest of the transcript. Similar results have been presented before using RNA blots (50,102,103), RNase protection assay (60) or sRNA-Seq (103,104). When the footprint was not completely abolished, as for *mcd1*, *mca1* or *mcg1* (20) the reason could be that 5′ ends unspecifically bind a protein factor, or that 3′→5′ exonucleases can drop-off prematurely, leaving behind an unprotected 5′-end sRNA. Another possibility is that the mutated gene functions in conjunction with other factors that can provide a low level of protection. Indeed, *MCD1* has been proposed to cooperate with the unknown *MCD4* gene product for *petD* stabilization (105). Cooperative binding of several proteins could also explain the shape of some cosRNAs such as that of *psbA*, with its two major 3′ ends (Supplementary Figure S5). In a 5′-PTP, the 3′ end is always less sharp than the 5′ end. The fact that it often carries a short A-rich tail suggests repeated attempts of PNPase to degrade the sRNA and indirectly implies that the protein remains bound to the footprint after the rest of the mRNA has been degraded. It will be important in the future to determine whether the sRNAs generated as M factor footprints can compete with the mRNA for the binding of the protein, as has been suggested for PPR10 (59).

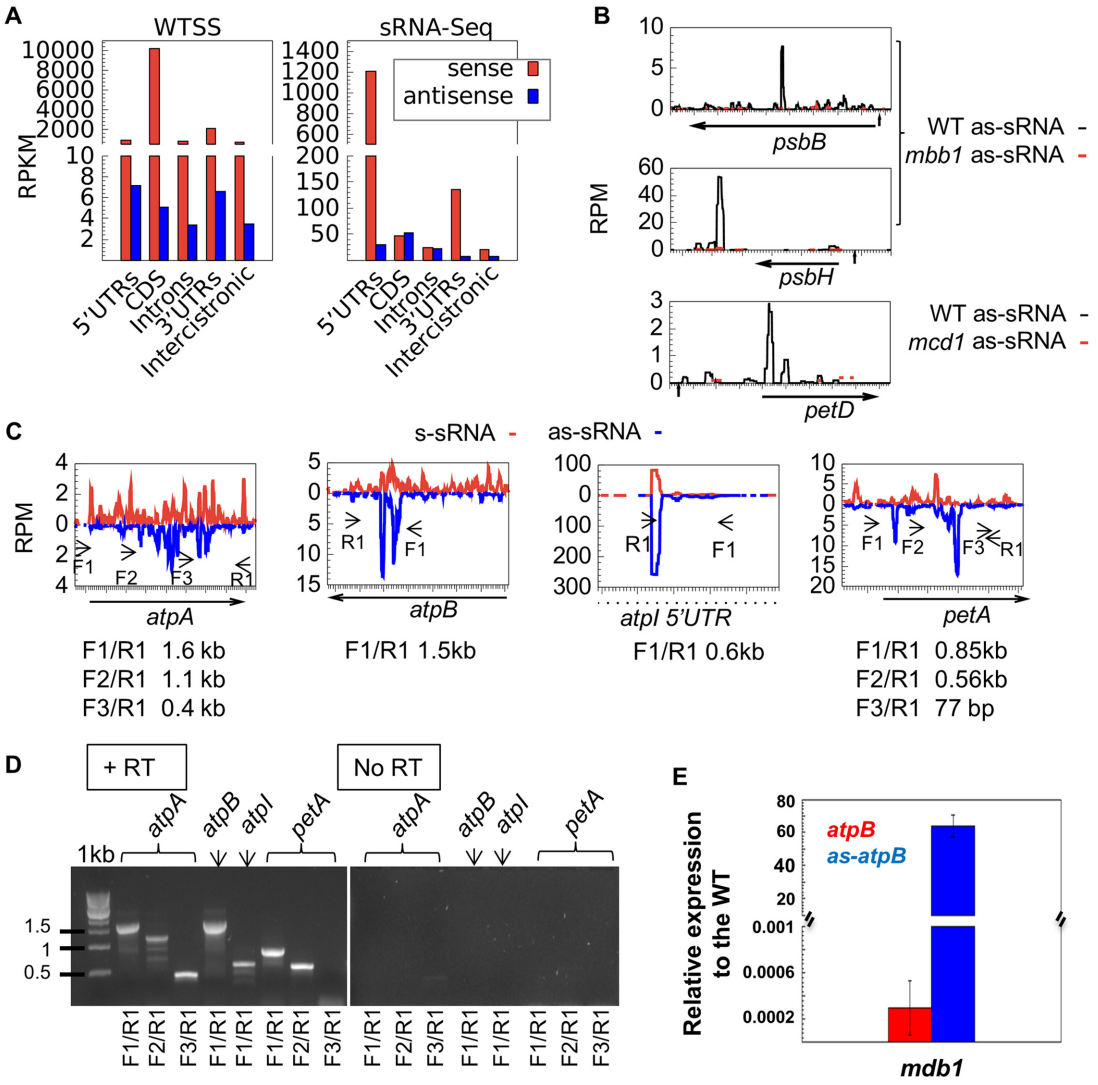


**Figure 9.** Antisense transcription in the Cp genome. (**A**) Comparison of sense (red) and antisense (blue) coverage over each chloroplast region as averaged RPKMs from four directional WTSS datasets and 4 sRNA-Seq datasets derived from the same RNA samples. (**B**) Coverage in RPM of antisense sRNAs in WT (black) and mutant strains (red). The orientation of the horizontal arrow indicate transcription direction of the mRNA. (**C**) Profile of s-sRNAs (red) and as-sRNAs (blue) at the *atpA*, *atpB*, *atpI* and *petA* loci. F1,F2 and F3 primers were used for reverse transcription, then combined with primers R1 for PCR. (**D**) Strand-specific RT-PCR in the presence (left panel) and absence of the reverse transcriptase (RT) (right panel). (**E**) Expression level of *atpB* mRNA (red) and antisense-*atpB* (in blue) in the *mdb1* mutant relative to the WT by qPCR. The values are the average of two independent qPCR assays ± SD.

In land plants chloroplasts, RNA-binding proteins can stabilize transcripts also against 3′→5′ exonucleases and thus define the 3′ end (6,59) and in this case the Cp footprints show a sharp 3′ end (59). In *Chlamydomonas*, all available studies link formation of the mature 3′ end to a secondary structure (68–71,106). We found no evidence for 3′-sharp cosRNAs at the 3′ end of transcripts, but sRNAs were often found in conjunction with a predicted stem-loop downstream of a CDS. These were used to annotate the 3′ ends, in addition to those identified by cRT-PCR or collected from the literature (34,46,47,56-57,69–71,107,108).

Comparison of RPP- and mock-treated samples revealed that 45 of our 89 5′-cosRNAs had a triphosphorylated 5′ end and were actually marking a TSS. Our analysis was sensitive enough to identify an unstable primary 5′ end for 13 protein-coding genes, i.e. a low level TSS upstream of a strong PTP signal. For several genes, primer extension experiments have already shown that the precursor and mature transcripts start at these respective positions (47,48,51,54,109).

A Pribnow box 'TATAATAT' was identified starting 11–13 nt upstream of all TSS for protein-coding genes, but the Gilbert box (–35; TTGaca) was less clear and even completely missing in some genes. Interestingly, the promoters upstream of the tRNA genes tended to show a weaker match to the –10 TATAATAT consensus (3/23 perfect matches versus 24/29 for protein genes), but a stronger match to the –35 TTGACA consensus (with 7/23 perfect matches versus only 2/29 for protein genes). This was noted before for *rrnS* (110). However, there was no obvious cor-

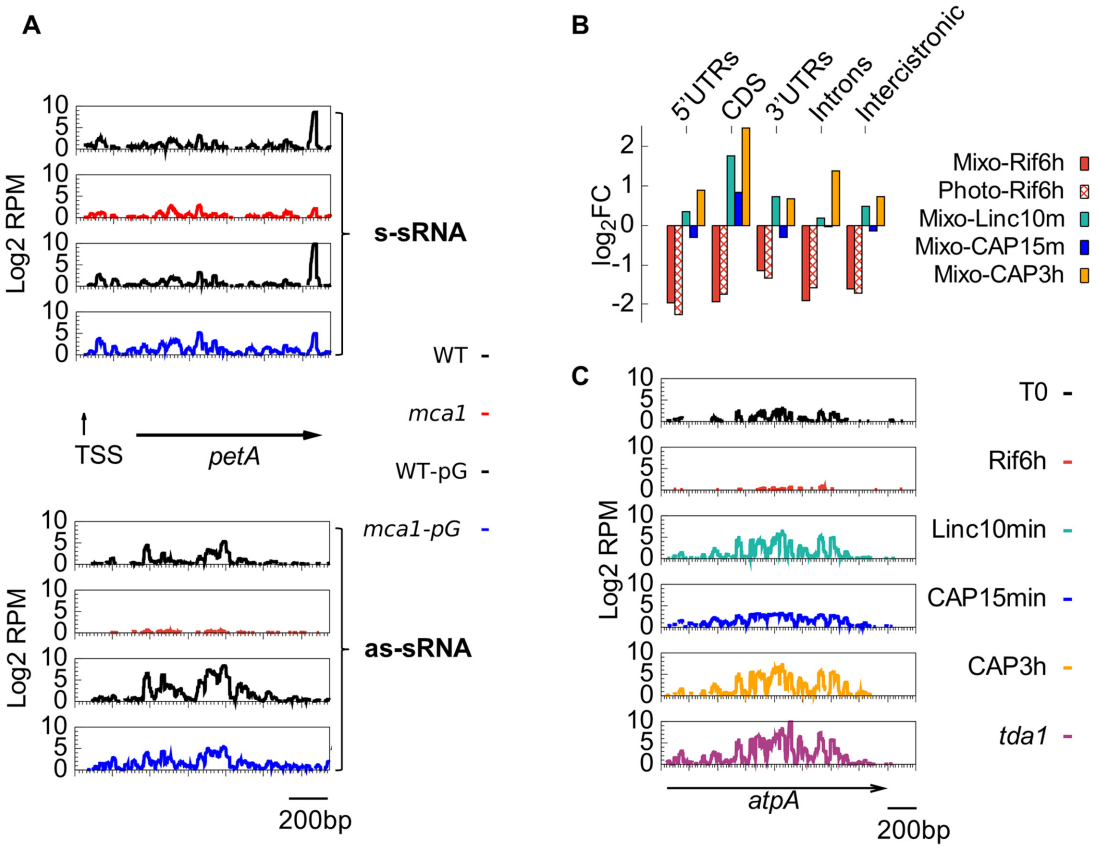


**Figure 10.** The effects of transcription and translation inhibition on the production of as-sRNAs. (**A**) Coverage of sense (upper panels) and antisense (lower panels) sRNA over the *petA* gene in (top to bottom) WT, *mca1*, *WT-pG* and *mca1-pG*). (**B**) $\log_2$FC of the averaged RPMs of drug-treated samples over the control per Cp region. (**C**) Coverage of antisense sRNAs along the *atpA* gene, following Rif, Linc or CAP treatment and in mutant *tda1*.

relation between adhesion to the consensus and transcript accumulation.

With at least 70% of the genes found in polycistronic units, co-transcription is much more prevalent in the Cp of *Chlamydomonas* than previously thought (111). Increasing complexity, some genes, although co-transcribed, have their own promoter which may lead to the production of a TSS cosRNA (e.g. *atpH*), but not always (e.g. *petD*). Many promoters remain to be identified, including those that drive formation of antisense transcripts.

### Transcript stability

For some *Chlamydomonas* Cp genes, mRNA accumulation has been shown to be determined by the amount of a dedicated M factor (19,29,102). Since expression level of several OTAFs vary depending on environmental conditions (19,20,28,102), we expected mRNA levels for the OTAFs and their targets to change coordinately between phototrophic and mixotrophic conditions. To our surprise, in spite of the fact that Cp transcripts decay more rapidly in phototrophic conditions, we found that very few Cp genes showed differential mRNA accumulation between the two conditions. Because transcription is not the limiting factor for Cp mRNA accumulation (29), we speculate that the lesser stability of transcripts in phototrophic conditions is compensated for by a higher efficiency of the initial stabilization step, i.e. the binding of the M factor. This suggests that the level of M factor proteins remains globally unchanged, despite possible variations in their mRNA levels, and that those released by Cp mRNA degradation can shed the footprint and rebind a newly-synthesized transcript. Whatever the mechanisms, the system thus appears to buffer changes in Cp transcript production and stability in the different growth conditions so that transcript levels remain stable.

### Effect of translation on the production of sRNAs

Beside M factor footprints, our study revealed another example of protection against RNases: the Ribosome-Protected Fragments (RPFs) that over-accumulate during CAP treatment. Their absence upon Linc treatment, their prevalence over CDS regions, their size similar to that generated by *in vitro* ribosome footprinting (83,84) strongly suggest that they represent degradation end-products of mRNA carrying stalled ribosomes. Indeed, a *tca1* mutant unable to translate the *petA* transcript also fails to accumulate RPFs specifically over the *petA* CDS. Thus sRNA-Seq of CAP-treated cells can be considered a form of '*in vivo* ribosome footprinting' which, although not as quantitative or resolutive as its *in vitro* counterpart, is sensitive enough to identify translated regions: thus *orf528* and *WendyA*, even if lacking orthologs in closely-related species, are likely trans-

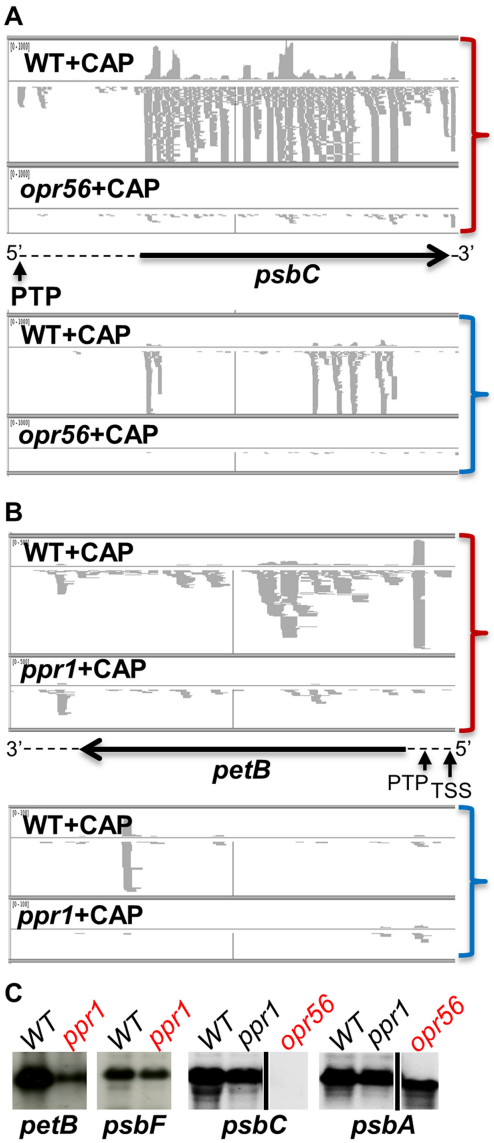


**Figure 11.** Snapshot of the IGV browser showing the alignment of sense (top, red parentheses) and antisense (bottom, blue parentheses) sRNA from CAP-treated WT and mutant strains. (**A**) *opr56* over *psbC*; (**B**) *ppr1* over *petB*. For each panel, the upper track displays the coverage, the lower track the reads. Arrows and dashed lines represent the CDS and UTRs and orientation on the genome; vertical arrows point to the 5′ ends of the transcripts. (**C**) Accumulation of *petB* mRNA in WT and *ppr1* (with *psbF* as a loading control) and *psbC* mRNA in WT, *ppr1* and *opr56* (with *psbA* as loading control).

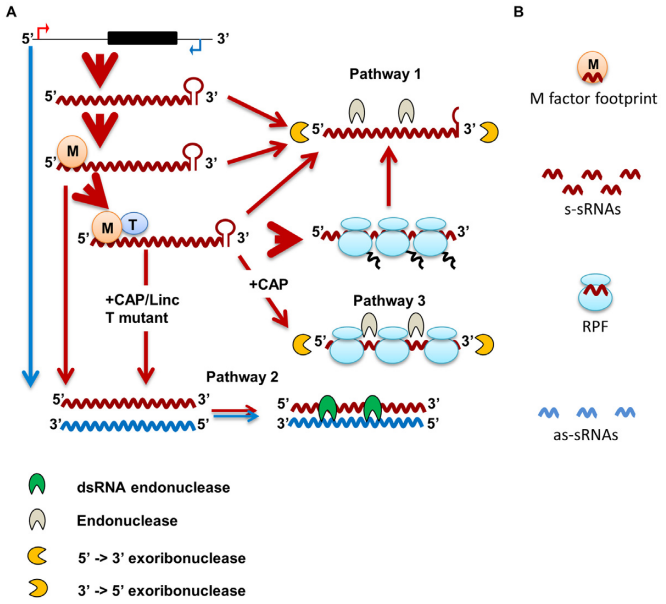


**Figure 12.** Model for Cp mRNA degradation and generation of small RNAs. (**A**) Transcription occurs on both strands of a gene locus and generates abundant mRNA (red) and rare antisense transcripts (blue). M factors stabilize the mRNA and T factors activate its translation. After translation, the mRNA can be delivered to degradation Pathway 1, starting with endoribonucleolytic cleavage followed by exoribonucleolytic degradation. Transcripts in excess, namely mRNAs not bound by an M factor or not activated for translation, are mostly directed toward Pathway 1, but they can also base-pair with antisense RNA: the generated dsRNA is substrate for a dsRNA endonuclease (Pathway 2). A block in translation (T factor mutant, Linc or CAP) will exacerbate Pathway 2. In addition, a block in translation with CAP induces the degradation of the mRNA engaged with ribosomes through endoribonucleolytic cleavage between the stalled ribosomes. (**B**) By-products of RNA degradation comprise mononucleotides (not shown), M factor footprints, ribosome-protected fragments and s-sRNA (derived from Pathways 1 and 3) as well as as-sRNAs derived from Pathway 2.

lated, while the putative *orf58* (86) is not translated, at least in mixotrophic conditions.

The steady state level of Cp mRNAs is not affected when they are not translated (Linc or CAP treatment). Similarly, mutations in T factors generally do not affect mRNA accumulation (14,30,55,112–114), except when they interact with the M factor (29). This contrasts with the situation observed in the prokaryotic ancestors where untranslated transcripts are degraded, but is consistent with the existence in the Cp of large pools of untranslated mRNAs (4). Translation can even decrease Cp transcript stability (57,115). There are many ways in which a traveling ribosome may affect the stability of the mRNA, for example by displacing RNA-bound proteins, disrupting stabilizing secondary structures or base-paired antisense RNA (see below). sRNA-Seq informs us on the final products of transcript degradation and hence on its mechanisms, but unfortunately not on its rate.

## A putative role of antisense RNA in regulation of Cp gene expression

Our results uncover the existence of antisense transcripts in the Cp of *Chlamydomonas* and provide insights into the possible consequences of their base-paring with mRNAs. Antisense transcripts can be generated when transcription units converge (five sites in the genome) or by the firing of 'antisense promoters' anywhere. Such a promoter has been described before for *petA* (99) and our results suggest that there are many more, for example those responsible for the long asRNA transcripts revealed by strand specific RT-PCR at the *atpB*, *atpI* and *atpA* loci.

Based on WTSS, long antisense transcripts are expressed to low levels and are rather unstable upon Rif. But their degradation is obviously more prone to yield sRNAs, be-

cause as-sRNAs accumulate to comparable levels than sense sRNAs. Moreover, the generation of as-sRNAs over a target gene appears to be dependent on the presence of the sense mRNA, since in M factor mutants both sense and antisense sRNAs are decreased over the target gene. Conversely, the artificial over-accumulation of a sense mRNA (i.e. in the *petA-pG* strains) induces an overall increase of as-sRNAs. These observations suggest that degradation of lg-asRNAs to yield as-sRNAs requires their pairing with the mRNA. Accordingly, in the *mdb1* mutant, the lack of *atpB* mRNA correlates with an increased accumulation of the long antisense transcript and a decrease of as-sRNAs. Ribosomes traveling on the mRNA would limit the base-pairing with antisense transcripts and, indeed, as-sRNA coverage increases when translation is abolished (Linc or CAP treatment, *tda1* mutation). Because mRNAs are very abundant, changes in the translation status only marginally affects their sRNA yield (except for RPFs), while it dramatically affects the fate of antisense transcripts. Cleavage of dsRNAs, followed by exonucleolytic degradation, would explain the variable size of as-sRNAs and also why lg-asRNA transcripts are usually under-represented in WTSS datasets. Candidate chloroplast-targeted enzymes for double-stranded RNA cleavage include the stem-loop endoribonuclease CSP41 (Cre10.g440050), a 'mini-III' RNAse (Cre11.g482841) orthologous to that described in vascular plants (73) or a distant homolog of the RNase M5 (Cre12.g497101), which in bacteria is involved in processing of the 5S rRNA.

These results raise the question of whether antisense transcription and processing of dsRNA substrates have a biological function. We propose that degradation of dsRNAs could participate in the removal of transcripts in excess that could not be activated for translation due to limiting amounts of T factors, thus contributing to set mRNA steady state levels. Consequently, varying the levels of lg-asRNAs could impact expression of the complementary mRNA, as previously shown in land plants (96–98). In this respect, antisense RNA could acquire regulatory functions.

By taking into account our results from the effects of the treatments on the sRNAs, we propose a model for Cp mRNA degradation that is articulated in three pathways (Figure 12). In steady state conditions, the major pathway for mRNA degradation (Pathway 1) initiates with an internal endonucleolytic cleavage of the mRNA followed by 5′→3′ and 3′→5′ exoribonucleolytic trimming, that generates nucleoside-monophosphates and a small proportion of sRNAs. Transcripts in excess, those that are not stabilized due to a limiting amount of M factor, are degraded through Pathway 1, but they can also base-pair with low abundance complementary as-RNAs forming dsRNAs degraded by Pathway 2. Blocking translation initiation will thus indirectly stimulate Pathway 2. In addition, when translation is inhibited with CAP, those transcripts engaged by the ribosomes are degraded to RPFs by endo- and exonucleolytic attack of the regions between the stalled ribosomes (Pathway 3).

### Using sRNA-Seq to identify the target of candidate OTAFs

The genetic network linking Cp genes and their nuclear-encoded OTAFs has thus far been built mostly by forward genetics but reverse genetics is progressively taking over. We show here that sRNA-Seq of CAP-treated cells can allow the rapid identification of the molecular target and mode of action of a candidate OTAF. We demonstrate that the targets of OPR56 and PPR1 are respectively *psbC* and *petB*. While the near total disappearance of *psbC* s- and as-sRNAs in *opr56* clearly qualifies it is an M factor, PPR1 appears to act primarily in translation, while contributing to the stability of the mRNA. In the T factor mutant *taa1-F23*, an even larger decrease in *psaA* mRNA was also observed, and it was necessary to stabilize the mRNA by a poly-G track to prove that TAA1 is indeed a T factor (19).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Sorel Fitz-Gibbon, Sean Gallagher and Sabeeha Merchant (UCLA Molecular Biology Institute) for providing the sequence of the chloroplast genome 'cv11' used in this study. We thank Benoist Laurent for creating the chloroplast genome browser and Marc Dreyfus for critical reading of the manuscript.

## FUNDING

Centre National de la Recherche Scientifique; Université Pierre et Marie Curie, Paris 06; Agence Nationale de la Recherche [ChloroRNP ANR-13-BSV7-0001-001]; 'Initiative d'Excellence' [DYNAMO ANR-11-LABX-0011-01]. Funding for open access charge: Centre National de la Recherche Scientifique.
*Conflict of interest statement.* None declared.

## REFERENCES

1. Keeling,P.J. (2010) The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. Roy. Soc. Lond. B: Biol. Sci.*, **365**, 729–748.
2. Martin,W., Rujan,T., Richly,E., Hansen,A., Cornelsen,S., Lins,T., Leister,D., Stoebe,B., Hasegawa,M. and Penny,D. (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12246–12251.
3. Maul,J.E., Lilly,J.W., Cui,L., dePamphilis,C.W., Miller,W., Harris,E.H. and Stern,D.B. (2002) The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. *Plant Cell*, **14**, 2659–2679.
4. Eberhard,S., Drapier,D. and Wollman,F.A. (2002) Searching limiting steps in the expression of chloroplast-encoded proteins: relations between gene copy number, transcription, transcript abundance and translation rate in the chloroplast of Chlamydomonas reinhardtii. *Plant J.*, **31**, 149–160.
5. Stern,D.B., Goldschmidt-Clermont,M. and Hanson,M.R. (2010) Chloroplast RNA metabolism. *Annu. Rev. Plant Biol.*, **61**, 125–155.
6. Pfalz,J., Bayraktar,O.A., Prikryl,J. and Barkan,A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.*, **28**, 2042–2052.
7. Barkan,A. (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.*, **155**, 1520–1532.

8. Barkan,A. and Small,I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.*, **65**, 415–442.
9. Stern,D.B. and Gruissem,W. (1987) Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell*, **51**, 1145–1157.
10. Hammani,K., Bonnard,G., Bouchoucha,A., Gobert,A., Pinker,F., Salinas,T. and Giege,P. (2014) Helical repeats modular proteins are major players for organelle gene expression. *Biochimie*, **100**, 141–150.
11. Barkan,A., Rojas,M., Fujii,S., Yap,A., Chong,Y.S., Bond,C.S. and Small,I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.*, **8**, e1002910.
12. Tourasse,N.J., Choquet,Y. and Vallon,O. (2013) PPR proteins of green algae. *RNA Biol.*, **10**, 1526–1542.
13. Rahire,M., Laroche,F., Cerutti,L. and Rochaix,J.D. (2012) Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. *Plant J.*, **72**, 652–661.
14. Eberhard,S., Loiselay,C., Drapier,D., Bujaldon,S., Girard-Bascou,J., Kuras,R., Choquet,Y. and Wollman,F.A. (2011) Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of atpA transcripts in Chlamydomonas reinhardtii chloroplasts. *Plant J.*, **67**, 1055–1066.
15. Auchincloss,A.H., Zerges,W., Perron,K., Girard-Bascou,J. and Rochaix,J.D. (2002) Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast psbC mRNA in Chlamydomonas reinhardtii. *J. Cell Biol.*, **157**, 953–962.
16. Merendino,L., Perron,K., Rahire,M., Howald,I., Rochaix,J.D. and Goldschmidt-Clermont,M. (2006) A novel multifunctional factor involved in trans-splicing of chloroplast introns in Chlamydomonas. *Nucleic Acids Res.*, **34**, 262–274.
17. Balczun,C., Bunse,A., Hahn,D., Bennoun,P., Nickelsen,J. and Kück,U. (2005) Two adjacent nuclear genes are required for functional complementation of a chloroplast trans-splicing mutant from *Chlamydomonas reinhardtii*. *Plant J.*, **43**, 636–648.
18. Marx,C., Wunsch,C. and Kück,U. (2015) The octatricopeptide repeat protein Raa8 is required for chloroplast trans splicing. *Eukaryot. Cell*, **14**, 998–1005.
19. Lefebvre-Legendre,L., Choquet,Y., Kuras,R., Loubery,S., Douchi,D. and Goldschmidt-Clermont,M. (2015) A nucleus-encoded chloroplast protein regulated by iron availability governs expression of the photosystem I subunit PsaA in Chlamydomonas reinhardtii. *Plant Physiol.*, **167**, 1527–1540.
20. Wang,F., Johnson,X., Cavaiuolo,M., Bohne,A.V., Nickelsen,J. and Vallon,O. (2015) Two Chlamydomonas OPR proteins stabilize chloroplast mRNAs encoding small subunits of photosystem II and cytochrome b6 f. *Plant J.*, **82**, 861–873.
21. Reifschneider,O., Marx,C., Jacobs,J., Kollipara,L., Sickmann,A., Wolters,D. and Kück,U. (2016) A ribonucleoprotein supercomplex involved in trans-splicing of organelle group II introns. *J. Biol. Chem.*, **291**, 23330–23342.
22. Murakami,S., Kuehnle,K. and Stern,D.B. (2005) A spontaneous tRNA suppressor of a mutation in the Chlamydomonas reinhardtii nuclear MCD1 gene required for stability of the chloroplast petD mRNA. *Nucleic Acids Res.*, **33**, 3372–3380.
23. Shepherd,H.S., Boynton,J.E. and Gillham,N.W. (1979) Mutations in nine chloroplast loci of Chlamydomonas affecting different photosynthetic functions. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 1353–1357.
24. Johnson,X., Wostrikoff,K., Finazzi,G., Kuras,R., Schwarz,C., Bujaldon,S., Nickelsen,J., Stern,D.B., Wollman,F.A. and Vallon,O. (2010) MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in *Chlamydomonas* and *Arabidopsis*. *Plant Cell*, **22**, 234–248.
25. Monod,C., Goldschmidt-Clermont,M. and Rochaix,J.D. (1992) Accumulation of chloroplast psbB RNA requires a nuclear factor in Chlamydomonas reinhardtii. *Mol. Gen. Genet.*, **231**, 449–459.
26. Drager,R.G., Girard-Bascou,J., Choquet,Y., Kindle,K.L. and Stern,D.B. (1998) In vivo evidence for 5′→3′ exoribonuclease degradation of an unstable chloroplast mRNA. *Plant J.*, **13**, 85–96.
27. Wostrikoff,K., Choquet,Y., Wollman,F.A. and Girard-Bascou,J. (2001) TCA1, a single nuclear-encoded translational activator specific for *petA* mRNA in *Chlamydomonas reinhardtii* chloroplast. *Genetics*, **159**, 119–132.
28. Raynaud,C., Loiselay,C., Wostrikoff,K., Kuras,R., Girard-Bascou,J., Wollman,F.A. and Choquet,Y. (2007) Evidence for regulatory function of nucleus-encoded factors on mRNA stabilization and translation in the chloroplast. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 9093–9098.
29. Loiselay,C., Gumpel,N.J., Girard-Bascou,J., Watson,A.T., Purton,S., Wollman,F.A. and Choquet,Y. (2008) Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in Chlamydomonas reinhardtii chloroplasts. *Mol. Cell. Biol.*, **28**, 5529–5542.
30. Drapier,D., Girard-Bascou,J. and Wollman,F.A. (1992) Evidence for nuclear control of the expression of the atpA and atpB chloroplast genes in Chlamydomonas. *Plant Cell*, **4**, 283–295.
31. Wostrikoff,K., Girard-Bascou,J., Wollman,F.A. and Choquet,Y. (2004) Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of Chlamydomonas. *EMBO J.*, **23**, 2696–2705.
32. Li,X., Zhang,R., Patena,W., Gang,S.S., Blum,S.R., Ivanova,N., Yue,R., Robertson,J.M., Lefebvre,P.A., Fitz-Gibbon,S.T. *et al.* (2016) An indexed, mapped mutant library enables reverse genetics studies of biological processes in Chlamydomonas reinhardtii. *Plant Cell*, **28**, 367–387.
33. Harris,E.H. (1989) *The Chlamydomonas Source Book: A Comprehensive Guide to Biology and Laboratory Use*. Academic Press, San Diego.
34. Drapier,D., Suzuki,H., Levy,H., Rimbault,B., Kindle,K.L., Stern,D.B. and Wollman,F.A. (1998) The chloroplast atpA gene cluster in Chlamydomonas reinhardtii. Functional analysis of a polycistronic transcription unit. *Plant Physiol.*, **117**, 629–641.
35. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
36. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
37. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Subgroup,G.P.D.P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
38. Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative Genomics Viewer. *Nat Biotechnol*, **29**, 24–26.
39. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
40. Mortazavi,A., Williams,B.A., McCue,K., Schaeffer,L. and Wold,B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
41. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
42. Michel,A.M., Mullan,J.P., Velayudhan,V., O'Connor,P.B., Donohue,C.A. and Baranov,P.V. (2016) RiboGalaxy: a browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biol.*, **13**, 316–319.
43. Pfaffl,M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, e45.
44. Fan,W.H., Woelfle,M.A. and Mosig,G. (1995) Two copies of a DNA element, 'Wendy', in the chloroplast chromosome of Chlamydomonas reinhardtii between rearranged gene clusters. *Plant Mol. Biol.*, **29**, 63–80.
45. Sturm,N.R., Kuras,R., Büschlen,S., Sakamoto,W., Kindle,K.L., Stern,D.B. and Wollman,F.A. (1994) The petD gene is transcribed by functionally redundant promoters in Chlamydomonas reinhardtii chloroplasts. *Mol. Cell. Biol.*, **14**, 6171–6179.
46. Turmel,M., Choquet,Y., Goldschmidt-Clermont,M., Rochaix,J.D., Otis,C. and Lemieux,C. (1995) The trans-spliced intron 1 in the psaA gene of the Chlamydomonas chloroplast: a comparative analysis. *Curr. Genet.*, **27**, 270–279.
47. Erickson,J.M., Rahire,M. and Rochaix,J.D. (1984) Chlamydomonas reinhardtii gene for the 32 000 mol. wt. protein of photosystem II contains four large introns and is located entirely within the chloroplast inverted repeat. *EMBO J.*, **3**, 2753–2762.

48. Bruick,R.K. and Mayfield,S.P. (1998) Processing of the psbA 5′ untranslated region in Chlamydomonas reinhardtii depends upon factors mediating ribosome association. *J. Cell Biol.*, **143**, 1145–1153.
49. Swiatek,M., Kuras,R., Sokolenko,A., Higgs,D., Olive,J., Cinque,G., Muller,B., Eichacker,L.A., Stern,D.B., Bassi,R. *et al.* (2001) The chloroplast gene ycf9 encodes a photosystem II (PSII) core subunit, PsbZ, that participates in PSII supramolecular architecture. *Plant Cell*, **13**, 1347–1367.
50. Loizeau,K., Qu,Y., Depp,S., Fiechter,V., Ruwe,H., Lefebvre-Legendre,L., Schmitz-Linneweber,C. and Goldschmidt-Clermont,M. (2014) Small RNAs reveal two target sites of the RNA-maturation factor Mbb1 in the chloroplast of Chlamydomonas. *Nucleic Acids Res.*, **42**, 3286–3297.
51. Vaistij,F.E., Goldschmidt-Clermont,M., Wostrikoff,K. and Rochaix,J.D. (2000) Stability determinants in the chloroplast psbB/T/H mRNAs of Chlamydomonas reinhardtii. *Plant J.*, **21**, 469–482.
52. Stampacchia,O., Girard-Bascou,J., Zanasco,J.L., Zerges,W., Bennoun,P. and Rochaix,J.D. (1997) A nuclear-encoded function essential for translation of the chloroplast psaB mRNA in chlamydomonas. *Plant Cell*, **9**, 773–782.
53. Woessner,J.P., Gillham,N.W. and Boynton,J.E. (1986) The sequence of the chloroplast atpB gene and its flanking regions in Chlamydomonas reinhardtii. *Gene*, **44**, 17–28.
54. Nickelsen,J., van Dillewijn,J., Rahire,M. and Rochaix,J.D. (1994) Determinants for stability of the chloroplast psbD RNA are located within its short leader region in Chlamydomonas reinhardtii. *EMBO J.*, **13**, 3182–3191.
55. Rochaix,J.D., Kuchka,M., Mayfield,S., Schirmer-Rahire,M., Girard-Bascou,J. and Bennoun,P. (1989) Nuclear and chloroplast mutations affect the synthesis or stability of the chloroplast psbC gene product in Chlamydomonas reinhardtii. *EMBO J.*, **8**, 1013–1021.
56. Takahashi,Y., Goldschmidt-Clermont,M., Soen,S.Y., Franzen,L.G. and Rochaix,J.D. (1991) Directed chloroplast transformation in Chlamydomonas reinhardtii: insertional inactivation of the psaC gene encoding the iron sulfur protein destabilizes photosystem I. *EMBO J.*, **10**, 2033–2040.
57. Zicker,A.A., Kadakia,C.S. and Herrin,D.L. (2007) Distinct roles for the 5′ and 3′ untranslated regions in the degradation and accumulation of chloroplast tufA mRNA: identification of an early intermediate in the in vivo degradation pathway. *Plant Mol .Biol.*, **63**, 689–702.
58. Zhelyazkova,P., Hammani,K., Rojas,M., Voelker,R., Vargas-Suarez,M., Borner,T. and Barkan,A. (2012) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.*, **40**, 3092–3105.
59. Ruwe,H., Wang,G., Gusewski,S. and Schmitz-Linneweber,C. (2016) Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.*, **44**, 7406–7417.
60. Ruwe,H. and Schmitz-Linneweber,C. (2012) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.*, **40**, 3106–3116.
61. Prikryl,J., Rojas,M., Schuster,G. and Barkan,A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 415–420.
62. Germain,A., Kim,S.H., Gutierrez,R. and Stern,D.B. (2012) Ribonuclease II preserves chloroplast RNA homeostasis by increasing mRNA decay rates, and cooperates with polynucleotide phosphorylase in 3′ end maturation. *Plant J.*, **72**, 960–971.
63. Goldschmidt-Clermont,M., Choquet,Y., Girard-Bascou,J., Michel,F., Schirmer-Rahire,M. and Rochaix,J.D. (1991) A small chloroplast RNA may be required for trans-splicing in Chlamydomonas reinhardtii. *Cell*, **65**, 135–143.
64. Klinkert,B., Schwarz,C., Pohlmann,S., Pierre,Y., Girard-Bascou,J. and Nickelsen,J. (2005) Relationship between mRNA levels and protein accumulation in a chloroplast promoter-mutant of Chlamydomonas reinhardtii. *Mol. Genet. Genomics*, **274**, 637–643.
65. Zimmer,S.L., Schein,A., Zipor,G., Stern,D.B. and Schuster,G. (2009) Polyadenylation in Arabidopsis and Chlamydomonas organelles: the input of nucleotidyltransferases, poly(A) polymerases and polynucleotide phosphorylase. *Plant J.*, **59**, 88–99.
66. Drager,R.G., Higgs,D.C., Kindle,K.L. and Stern,D.B. (1999) 5′ to 3′ exoribonucleolytic activity is a normal component of chloroplast mRNA decay pathways. *Plant J.*, **19**, 521–531.
67. Germain,A., Hotto,A.M., Barkan,A. and Stern,D.B. (2013) RNA processing and decay in plastids. *Wiley interdiscipl. Rev. RNA*, **4**, 295–316.
68. Stern,D.B. and Kindle,K.L. (1993) 3′end maturation of the Chlamydomonas reinhardtii chloroplast atpB mRNA is a two-step process. *Mol. Cell Biol.*, **13**, 2277–2285.
69. Goldschmidt-Clermont,M., Rahire,M. and Rochaix,J.D. (2008) Redundant cis-acting determinants of 3′ processing and RNA stability in the chloroplast rbcL mRNA of Chlamydomonas. *Plant J.*, **53**, 566–577.
70. Blowers,A.D., Klein,U., Ellmore,G.S. and Bogorad,L. (1993) Functional in vivo analyses of the 3′ flanking sequences of the Chlamydomonas chloroplast rbcL and psaB genes. *Mol. Gen. Genet.*, **238**, 339–349.
71. Lee,H., Bingham,S.E. and Webber,A.N. (1996) Function of 3′ non-coding sequences and stop codon usage in expression of the chloroplast psaB gene in Chlamydomonas reinhardtii. *Plant Mol. Biol.*, **31**, 337–354.
72. Megel,C., Morelle,G., Lalande,S., Duchêne,A.M., Small,I. and Maréchal-Drouard,L. (2015) Surveillance and cleavage of eukaryotic tRNAs. *Int. J. Mol. Sci.*, **16**, 1873–1893.
73. Hotto,A.M., Castandet,B., Gilet,L., Higdon,A., Condon,C. and Stern,D.B. (2015) Arabidopsis chloroplast mini-ribonuclease III participates in rRNA maturation and intron recycling. *Plant Cell*, **27**, 724–740.
74. Liu,X.Q., Gillham,N.W. and Boynton,J.E. (1989) Chloroplast ribosomal protein gene rps12 of Chlamydomonas reinhardtii. Wild-type sequence, mutation to streptomycin resistance and dependence, and function in Escherichia coli. *J. Biol. Chem.*, **264**, 16100–16108.
75. Kück,U., Choquet,Y., Schneider,M., Dron,M. and Bennoun,P. (1987) Structural and transcription analysis of two homologous genes for the P700 chlorophyll a-apoproteins in Chlamydomonas reinhardii: evidence for in vivo trans-splicing. *EMBO J.*, **6**, 2185–2195.
76. Choquet,Y., Goldschmidt-Clermont,M., Girard-Bascou,J., Kück,U., Bennoun,P. and Rochaix,J.D. (1988) Mutant phenotypes support a trans-splicing mechanism for the expression of the tripartite psaA gene in the C. reinhardtii chloroplast. *Cell*, **52**, 903–913.
77. Robertson,D., Boynton,J.E. and Gillham,N.W. (1990) Cotranscription of the wild-type chloroplast atpE gene encoding the CF1/CF0 epsilon subunit with the 3′ half of the rps7 gene in Chlamydomonas reinhardtii and characterization of frameshift mutations in atpE. *Mol. Gen. Genet.*, **221**, 155–163.
78. Boudreau,E., Takahashi,Y., Lemieux,C., Turmel,M. and Rochaix,J.D. (1997) The chloroplast ycf3 and ycf4 open reading frames of Chlamydomonas reinhardtii are required for the accumulation of the photosystem I complex. *EMBO J.*, **16**, 6095–6104.
79. Cui,L., Leebens-Mack,J., Wang,L.S., Tang,J., Rymarquis,L., Stern,D.B. and dePamphilis,C.W. (2006) Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. *BMC Evol. Biol.*, **6**, 13.
80. Vazquez,D. (1979) Inhibitors of protein biosynthesis. *Mol. Biol. Biochem. Biophys.*, **30**, 1–312.
81. Dreyfus,M. (2009) Killer and protective ribosomes. *Progr. Mol. Biol. Transl. Sci.*, **85**, 423–466.
82. Bieri,P., Leibundgut,M., Saurer,M., Boehringer,D. and Ban,N. (2017) The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J.*, **36**, 475–486.
83. Chotewutmontri,P. and Barkan,A. (2016) Dynamics of chloroplast translation during chloroplast differentiation in maize. *PLoS Genet.*, **12**, e1006106.
84. Chung,B.Y., Hardcastle,T.J., Jones,J.D., Irigoyen,N., Firth,A.E., Baulcombe,D.C. and Brierley,I. (2015) The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA*, **21**, 1731–1745.

85. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
86. Takahashi,Y., Rahire,M., Breyton,C., Popot,J.L., Joliot,P. and Rochaix,J.D. (1996) The chloroplast *ycf7* (*petL*) open reading frame of *Chlamydomonas reinhardtii* encodes a small functionally important subunit of the cytochrome $b_6f$ complex. *EMBO J.*, **15**, 3498–3506.
87. Tuller,T., Veksler-Lublinsky,I., Gazit,N., Kupiec,M., Ruppin,E. and Ziv-Ukelson,M. (2011) Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.*, **12**, R110.
88. Durrenberger,F., Thompson,A.J., Herrin,D.L. and Rochaix,J.D. (1996) Double strand break-induced recombination in Chlamydomonas reinhardtii chloroplasts. *Nucleic.Acids.Res.*, **24**, 3323–3331.
89. Odom,O.W., Holloway,S.P., Deshpande,N.N., Lee,J. and Herrin,D.L. (2001) Mobile self-splicing group I introns from the psbA gene of Chlamydomonas reinhardtii: highly efficient homing of an exogenous intron containing its own promoter. *Mol. Cell. Biol.*, **21**, 3472–3481.
90. Pato,M.L. and Von Meyenburg,K. (1970) Residual RNA synthesis in *Escherichia coli* after inhibition of initiation of transcription by Rifampicin. *Cold Spring Harb. Symp. Quant. Biol.*, **35**, 497–504.
91. Wagner,E.G. and Simons,R.W. (1994) Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.*, **48**, 713–742.
92. Georg,J. and Hess,W.R. (2011) cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol. Mol. Biol. Rev.*, **75**, 286–300.
93. Hotto,A.M., Schmitz,R.J., Fei,Z., Ecker,J.R. and Stern,D.B. (2011) Unexpected diversity of chloroplast noncoding RNAs as revealed by deep sequencing of the Arabidopsis transcriptome. *G3*, **1**, 559–570.
94. Zhelyazkova,P., Sharma,C.M., Förstner,K.U., Liere,K., Vogel,J. and Börner,T. (2012) The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell*, **24**, 123–136.
95. Hotto,A.M., Huston,Z.E. and Stern,D.B. (2010) Overexpression of a natural chloroplast-encoded antisense RNA in tobacco destabilizes 5S rRNA and retards plant growth. *BMC Plant Biol.*, **10**, 213.
96. Zghidi-Abouzid,O., Merendino,L., Buhr,F., Malik Ghulam,M. and Lerbs-Mache,S. (2011) Characterization of plastid psbT sense and antisense RNAs. *Nucleic Acids Res.*, **39**, 5379–5387.
97. Chevalier,F., Ghulam,M.M., Rondet,D., Pfannschmidt,T., Merendino,L. and Lerbs-Mache,S. (2015) Characterization of the psbH precursor RNAs reveals a precise endoribonuclease cleavage site in the psbT/psbH intergenic region that is dependent on psbN gene expression. *Plant Mol. Biol.*, **88**, 357–367.
98. Sharwood,R.E., Halpert,M., Luro,S., Schuster,G. and Stern,D.B. (2011) Chloroplast RNase J compensates for inefficient transcription termination by removal of antisense RNA. *RNA*, **17**, 2165–2176.
99. Thompson,R.J. and Mosig,G. (1987) Stimulation of a Chlamydomonas chloroplast promoter by novobiocin in situ and in E. coli implies regulation by torsional stress in the chloroplast DNA. *Cell*, **48**, 281–287.
100. Meierhoff,K., Felder,S., Nakamura,T., Bechtold,N. and Schuster,G. (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell*, **15**, 1480–1495.
101. Castandet,B., Hotto,A.M., Fei,Z. and Stern,D.B. (2013) Strand-specific RNA sequencing uncovers chloroplast ribonuclease functions. *FEBS Lett.*, **587**, 3096–3101.
102. Douchi,D., Qu,Y., Longoni,P., Legendre-Lefebvre,L., Johnson,X., Schmitz-Linneweber,C. and Goldschmidt-Clermont,M. (2016) A nucleus-encoded chloroplast phosphoprotein governs expression of the photosystem I subunit PsaC in Chlamydomonas reinhardtii. *Plant Cell*, **28**, 1182–1199.
103. Wu,W., Liu,S., Ruwe,H., Zhang,D., Melonek,J., Zhu,Y., Hu,X., Gusewski,S., Yin,P., Small,I.D. *et al.* (2016) SOT1, a pentatricopeptide repeat protein with a small MutS-related domain, is required for correct processing of plastid 23S-4.5S rRNA precursors in Arabidopsis thaliana. *Plant J.*, **85**, 607–621.
104. Aryamanesh,N., Ruwe,H., Sanglard,L.V., Eshraghi,L., Bussell,J.D., Howell,K.A., Small,I. and des Francs-Small,C.C. (2017) The pentatricopeptide repeat protein EMB2654 is essential for trans-splicing of a chloroplast small ribosomal subunit transcript. *Plant Physiol.*, **173**, 1164–1176.
105. Rymarquis,L.A., Webster,B.R. and Stern,D.B. (2007) The nucleus-encoded factor MCD4 participates in degradation of nonfunctional 3′ UTR sequences generated by cleavage of pre-mRNA in Chlamydomonas chloroplasts. *Mol. Genet. Genomics*, **277**, 329–340.
106. Jiao,H.S., Hicks,A., Simpson,C. and Stern,D.B. (2004) Short dispersed repeats in the Chlamydomonas chloroplast genome are collocated with sites for mRNA 3′ end formation. *Curr. Genet.*, **45**, 311–322.
107. Komine,Y., Kwong,L., Anguera,M.C., Schuster,G. and Stern,D.B. (2000) Polyadenylation of three classes of chloroplast RNA in Chlamydomonas reinhadtii. *RNA*, **6**, 598–607.
108. Johnson,C.H. and Schmidt,G.W. (1993) The psbB gene cluster of the Chlamydomonas reinhardtii chloroplast: sequence and transcriptional analyses of psbN and psbH. *Plant Mol. Biol.*, **22**, 645–658.
109. Rimbault,B. (2001) Etude de l'expression des gènes chloroplastiques atpA et atpB dans l'algue verte C. reinhardtii. *Thèse de Doctorat de l'Université Paris XI*.
110. Klein,U., De Camp,J.D. and Bogorad,L. (1992) Two types of chloroplast gene promoters in Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci . U.S.A.*, **89**, 3453–3457.
111. Klein,U. (2009) In: Harris,EE, Stern,DB and Whitman,G (eds). *The Chlamydomonas Sourcebook*, 2nd edn. Academic press, NY, Vol. **2**, pp. 893–914.
112. Girard-Bascou,J., Pierre,Y. and Drapier,D. (1992) A nuclear mutation affects the synthesis of the chloroplast psbA gene production Chlamydomonas reinhardtii. *Curr. Genet.*, **22**, 47–52.
113. Schwarz,C., Elles,I., Kortmann,J., Piotrowski,M. and Nickelsen,J. (2007) Synthesis of the D2 protein of photosystem II in Chlamydomonas is controlled by a high molecular mass complex containing the RNA stabilization factor Nac2 and the translational activator RBP40. *Plant Cell*, **19**, 3627–3639.
114. Dauvillee,D., Stampacchia,O., Girard-Bascou,J. and Rochaix,J.D. (2003) Tab2 is a novel conserved RNA binding protein required for translation of the chloroplast psaB mRNA. *EMBO J.*, **22**, 6378–6388.
115. Klaff,P. and Gruissem,W. (1991) Changes in chloroplast mRNA stability during leaf development. *Plant cell*, **3**, 517–529.