Introduction

The signal recognition particle (SRP) plays a critical role in the targeting of proteins to the ER membrane, the plasma membrane in bacteria and archaea and the thylakoid membrane in chloroplasts. SRP first binds the N-terminal signal sequence of the nascent chain as it appears on the surface of translating ribosomes. The resulting ribosome–nascent chain–SRP complex is targeted to the membrane through interaction with the SRP receptor (Keenan et al. 2001). In a series of events where GTP hydrolysis is essential, the SRP is released and translocation of the protein is initiated.

Components of the SRP have been identified in all three domains of life (Rosenblad et al. 2003). The eukaryotic particle is typically composed of an RNA molecule (~300 nucleotides) and six different protein subunits. One of the proteins, SRP54, is composed of an NG domain with GTPase activity and an M domain that contains a signal sequence binding site as well as a binding site for the SRP RNA. Most eubacteria, including cyanobacteria, have a simplified SRP with only a small RNA (75–115 nucleotides) and a protein homologous to the eukaryotic SRP54, in bacteria denoted Ffh. The RNA component corresponds to the helix 8 region of eukaryotic SRP. The helix 8 region and the SRP54 protein form the most strongly conserved parts of the SRP.

Chloroplasts have a SRP54 homologue (cpSRP54), which is distinct from the cytoplasmic protein and which is encoded by the nuclear genome. While SRP typically operates co-translationally, the chloroplast protein was unexpectedly found to be involved in post-translational targeting of nuclear-encoded light-harvesting complex proteins (LHCPs) to the thylakoid membrane (Li et al. 1995) and it forms a complex with cpSRP43, a protein found only in chloroplasts (Schuenemann et al. 1998). The cpSRP43 protein binds to regions in the M domain of cpSRP54 that correspond to RNA binding sites in the cytoplasmic SRP54 (Groves et al. 2001). No RNA could be immunoprecipitated with cpSRP54, and RNase treatment of stroma had no effect on LHCP biogenesis (Schuenemann et al. 1998). Furthermore, purified cpSRP54 and cpSRP43 assemble into an active transit complex. This is in contrast to cytoplasmic SRPs that are known to act predominantly in co-translational membrane targeting and where the RNA component is required.

A co-translational mechanism for targeting chloroplast-encoded proteins to the thylakoid membrane in higher plants was discovered by cross-linking of stromal cpSRP54 to the ribosome-bound nascent chain of the thylakoid membrane protein D1. No interaction of cpSRP43 with D1 was detected (Nilsson et al. 1999).

So far SRP has been studied in the chloroplasts of higher plants but not in lower plants and algae. As indicated above, post-translational cpSRP is not likely to contain an RNA, but it cannot be excluded that an RNA is required in a co-translational mechanism. Furthermore, it cannot be ruled out that post-translational targeting in lower plants and algae chloroplasts require an RNA component. SRP RNA candidates in the chloroplasts of Porphyra and Odontella were found previously (Packer and Howe 1998). However, the general significance of these findings is not clear. So far, no other chloroplast SRP RNA genes have been reported and no such genes are annotated in any of the chloroplast genomes that have been sequenced. We have now more carefully examined a large number of chloroplast genomes and report on several novel SRP RNA genes in red algae and Chlorophyta.

Magnus Alm Rosenblad and Tore Samuelsson
Department of Medical Biochemistry, Goteborg University, Box 440, SE-405 30 Goteborg, Sweden

Identification of Chloroplast Signal Recognition Particle RNA Genes

The signal recognition particle (SRP) is a ribonucleoprotein complex responsible for targeting proteins to the ER membrane in eukaryotes, the plasma membrane in bacteria and the thylakoid membrane in chloroplasts. In higher plants two different SRP-dependent mechanisms have been identified: one post-translational for proteins imported to the chloroplast and one co-translational for proteins encoded by the plastid genome. The post-translational chloroplast SRP (cpSRP) consists of the protein subunits cpSRP54 and cpSRP43. An RNA component has not been identified and does not seem to be required for the post-translational cpSRP. The co-translational mechanism is known to involve cpSRP54, but an RNA component has not yet been identified. Several chloroplast genomes have been sequenced recently, making a phylogenetically broad computational search for cpSRP RNA possible. We have analysed chloroplast genomes from 27 organisms. In higher plant chloroplasts, no SRP RNA genes were identified. However, eight plastids from red algae and Chlorophyta were found to contain an SRP RNA gene. These results suggest that SRP RNA forms a complex in these plastids with cpSRP54, reminiscent of the eubacterial SRP.

Keywords: Chloroplast — Evolution — RNA — Signal recognition particle (SRP).
Results

For the search for SRP RNAs in chloroplasts we considered 27 complete chloroplast genomes, including those of 16 land plants (Embryophyta), one Coleochaetales, one Mesostigmatales, three green algae, three Rhodophyta and three secondary endosymbionts (chloroplasts from *Euglena gracilis*, *Odontella sinensis* and *Guillardia theta*). A schematic phylogenetic tree including these organisms is shown in Fig. 1. A complete list of the organisms analysed is found in Materials and Methods.

Identification and analysis of chloroplast SRP RNA genes

To predict SRP RNA genes we combined a heuristic pattern-based search for conserved features of the helix 8 region with a COVE (Eddy and Durbin 1994) model of SRP RNA (Regalia et al. 2002). Many of the chloroplast RNAs could only be identified if we allowed one mismatch in the consensus helix 8 pattern for the pattern-based search. COVE is an implementation of the algorithms described by Durbin et al. (1998) that make use of probabilistic models to describe the sequence and secondary structure consensus of an RNA family. The model used in the searches was the non-Alu Prokaryote covariance model as described previously (Regalia et al. 2002).

We found a potential SRP RNA gene in eight of the plastids examined. Secondary structure models of these RNAs are shown in Fig. 2. No SRP RNA genes were identified in the higher plant chloroplasts, including the non-photosynthetic *Epifagus virginiana*. Of the eight identified RNAs only two (from *Porphyra* and *Odontella*) have been reported previously (Packer and Howe 1998). A multiple sequence alignment including previously known bacterial sequences showed, as expected, that the sequences of the chloroplast RNAs are most closely related to those of cyanobacteria. The *Porphyra* and *Odontella* sequences reported previously were found using a BLAST (Altschul et al. 1997) search using cyanobacterial SRP RNA as query sequence (Packer and Howe 1998). However, it is important to note that because SRP RNA sequences are poorly conserved most of the RNAs found in this work could not be found using BLAST or any other method based on sequence similarity.

The identified SRP RNA candidates aligned well to the COVE model, albeit with lower scores than those typically found in an analysis of bacterial SRP RNAs. The highest scores (~25) are similar to the lowest scores for bacterial RNAs (~30). MFOLD was able to fold the RNA in a manner that is consistent with the secondary structure predicted by COVE. There was no overlap between the predicted SRP RNA genes and other genes according to the genome annotation, except for *Cyanidioschyzon merolae*, which had an 8 nucleotide overlap with an upstream gene (see below). The predictions are also consistent with the observation that all the predicted RNAs have a TATA box element immediately upstream of the 5' end (see below). Finally, there is further support of many of the SRP RNA gene predictions from analysis of synteny as described further below.
Genomic localization of SRP RNA genes

In order to further evaluate the SRP RNA gene predictions we also examined genes that flank the RNA gene. The results of this analysis are shown in Fig. 4.

Fig. 4 Synteny of chloroplast SRP RNA genes. Chloroplast gene organization in regions where SRP RNAs were found. SRP RNA genes are indicated as ‘RNA’. Above is the red algal group, below the green group. The psbX gene of C. caldariorum is not in the annotation of the genome, but has been added here. (See the web supplement for an alignment of the psbX proteins.) Organisms listed are Porphyra purpurea (Pp), C. caldariorum (Cc), M. velata (Mv), G. theta (Gt), O. sinensis (Os), Mesostigma viride (Mv), Nephroselmis olivacea (No) and Chlorella vulgaris (Cv).

Analysis of promoter regions

In order to analyse promoter regions, the predicted chloroplast RNA sequences, including 40 nucleotides upstream of the 5' end, were aligned (Fig. 3). We relied on rules for TATA box location based on an analysis of cyanobacterial promoters as described previously (Vogel et al. 2003). The distance between the transcription start site and the TATA box was 5–9 nucleotides, the majority being 6–7 nucleotides. For comparison, the upstream region of known SRP RNA genes in cyanobacteria are shown (Fig. 3). No consensus elements corresponding to a –35 box were found, consistent with observations of Vogel et al. (2003).

As we have noted from previous SRP RNA gene predictions (Regalia et al. 2002) the exact location of the 5’ and 3’ ends cannot be predicted reliably. For the predicted sequences of Odontella we tentatively added 2 nucleotides at the 5’ end in order for the sequence to be consistent with the predicted TATA box promoter sequences. In addition, we removed from the initial COVE prediction of the C. merolae sequence 14 nucleotides at the 5’ end in order to make use of an alternative U-rich sequence which could be part of a transcription termination signal. The shorter version of the sequence resulted in a secondary structure which is more consistent with previously known SRP RNAs and also removed the overlap with the upstream apcD gene. In addition, the resulting gene sequence is a better alternative in the promoter analysis (Fig. 3). For most of the sequences shown in Fig. 2 we preferred a 5’ G, as this is the consensus of bacterial SRP RNAs. For a comparison of the sequences shown in Fig. 2 with the initial COVE predictions, the reader is referred to the web supplement to this paper (http://bio.lundberg.gu.se/srp04chl/).

Amino acid substitutions in cpSRP54 correlate with the absence of RNA

Higher plant cpSRP54 is unusual among SRP54 proteins in that it is not able to bind a bacterial SRP RNA effectively (Schuenemann et al. 1999). Groves et al. (2001) noted that cpSRP54 from Arabidopsis thaliana is unique among the SRP54 homologues in that a conserved glycine residue in the RNA binding domain is replaced by an aspartate, and that model-ling of cpSRP54 based on the crystal structure of the SRP54 M-domain–RNA complex (Batye et al. 2000) suggests that an aspartate in this position would introduce a negative charge in the immediate vicinity of the negatively charged phosphate backbone. This would reduce the affinity of cpSRP for SRP RNA. To further examine the phylogenetic distribution of this type of mutation as well as other changes in the RNA binding
Comparison between chloroplast and bacterial SRP RNAs

We noted differences in the predicted RNAs when compared with consensus features of bacterial RNAs. These are in (i) the tetraloop, (ii) the stem adjacent to the tetraloop and (iii) the symmetrical loop as described further below (Fig. 2).

The helix 8 ‘GNRA’ tetraloop in Mesostigma has a U in the first position. This nucleotide, as well as A and C, is compatible with the 3D structure from E. coli Ffh-SRP RNA (Batye et al. 2001) according to Leontis-Westhof isostericity matrices (Leontis et al. 2002). Wood et al. found that the mutations G→A and G→C in the first position could be tolerated without any growth effects, except for the most restrictive conditions (Wood et al. 1992). Another case where the tetraloop has a U in the first position is in Lactococcus lactis and Staphylococcus (Regalia et al. 2002), where the tetraloop is ‘UGAC’. The second nucleotide is U in O. sinensis instead of the bacterial consensus purine, a change that has been observed only in Trypanosoma brucei and Theileria annulata (Rosenblad et al. 2004). The fungi Aspergillus nidulans, Yarrowia lipolytica and Candida albicans have a C as second nucleotide.

In the closely related Nephroselmis and Chlorella the third base pair in the stem of the tetraloop stem–loop structure is A: instead of A:U, leaving only two canonical base pairs in this stem. Only Mycoplasma pulmonis SRP RNA is known to have a non-canonical base pair (G:A) in this stem.

Four of the sequences have an A:A (A49:A60) pair first in the symmetrical loop, instead of the usual G:A base pair. In Cyanidium caldarium this pair is G:U, found so far only in the fungi A. nidulans (Rosenblad et al. 2004) and Schizosaccharomyces pombe. A mutation to G:U (A60U) in E. coli was tested by Wood et al. (Wood et al. 1992) resulting in only a small effect on growth. In the C. caldarium sequence, the first base pair on the 3’ side of the symmetrical loop is C:G (consensus is pyrimidine:purine), whereas it is more commonly U:G or, as in some bacterial SRP RNAs, C:A. Only M. pulmonis SRP RNA has a non-consensus U:U pair.

Discussion

We have here reported a number of chloroplast SRP RNA genes that have not been identified previously. Based on primary sequence only a few of these have significant similarity to cyanobacterial SRP RNAs except for a shorter sequence that includes the tetraloop and the symmetrical loop of the helix 8 motif. This might be one reason they have escaped detection so far.

Although we cannot at this stage rule out that the genes identified in this work are pseudogenes and non-functional, there are a number of reasons why we believe they are bona fide SRP RNA genes. First, they all seem to have a typical upstream bacterial promoter sequence as well as a transcription termination signal. Furthermore, although their sequences in large parts are not conserved, they have preserved the primary and secondary structural elements characteristic of SRP RNA.

Evolution and distribution of chloroplast SRP RNA

We have shown that a number of red algae and Chlorophyta chloroplasts seem to encode an SRP RNA whereas this RNA is absent in the higher plants. This might reflect an evolution where the SRP RNA eventually was lost from the chloroplast genome, either because the RNA was not essential for SRP function, or because the gene was transferred to the nuclear genome.

No import of RNA into chloroplasts has yet been demonstrated. However, import of different kinds of RNA into mitochondria is well known (tRNA, RNase P RNA, 5S rRNA) (Entelis et al. 2001). Also, the plastid genome of the non-
photosynthetic parasitic plant *E. virginiana* contains only 17 of the 30 tRNA genes normally found in angiosperm plastid DNA. It has been suggested that cytosolic tRNAs are imported into plastids to make up a complete set of tRNAs required for translation (Lohan and Wolfe 1998). Therefore, the possibility that a nuclear-encoded cpSRP RNA is imported into the chloroplast of higher plants cannot be ruled out. However, we have searched the complete nuclear genomes of *Arabidopsis* and *Chlamydomonas* and have not been able to find a convincing bacterial type of SRP RNA. Using our search method, where pattern matching is combined with covariance models, we find only RNAs that are of the eukaryotic type that includes a structure with the helices 5, 6 and 8 (data not shown). The available data therefore indicate that higher plants and *Chlamydomonas* do not have a chloroplast SRP RNA.

Experiments to show that cpSRP is involved in post-translational targeting were carried out in higher plants. We do not know at this stage if the RNAs in red algae and Chlorophyta are at all involved in post-translational targeting. A SRP43 homologue has been identified in *C. merolae*, suggesting that SRP43 and SRP RNA are not mutually exclusive. We have failed to identify a SRP43 homologue in available genome sequences of *T. pseudonana*. Only when more nuclear genomes of the red and green algal group become available will it be possible to know how the phylogenetic distribution of SRP43 compares with that of cpSRP RNA.

### Comparison with RNase P and tmRNA

Interestingly, both tmRNA and the RNA component of RNase P (P RNA) have a phylogenetic distribution reminiscent of cpSRP RNA, reflecting an evolution where the RNA was lost in higher plant chloroplasts. Hence, tmRNA is found in all species (Gueneau de Novoa and Williams 2004) where we identified an SRP RNA, except *Chlorella*. Furthermore, the P RNA gene is found in the chloroplasts of *C. merolae* (Ohta et al. 2003), *C. caldarium* (Seif et al. 2003), *Nephroselmis* (Turmel et al. 1999) and *Porphyra* (Reith and Munholland 1995) where we found an SRP RNA. P RNA, as well as tmRNA, are also found in *Cyanophora*.

No gene encoding a homologue to the bacterial RNase P RNA has been found in higher plants and algae and there is evidence that spinach chloroplast RNase P does not contain an

### Table 1  Plastid genomes analysed

<table>
<thead>
<tr>
<th>Group</th>
<th>Organism of plastid genome</th>
<th>GenBank accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhodophyta</td>
<td><em>P. purpurea</em></td>
<td>NC_000925.1</td>
</tr>
<tr>
<td>C. merolae</td>
<td>NC_004799.1</td>
<td></td>
</tr>
<tr>
<td>C. caldarium</td>
<td>NC_001840.1</td>
<td></td>
</tr>
<tr>
<td>Secondary (red algal) endosymbionts</td>
<td><em>O. sinensis</em></td>
<td>NC_001713.1</td>
</tr>
<tr>
<td>G. theta</td>
<td>NC_000926.1</td>
<td></td>
</tr>
<tr>
<td>Chlorophyta (green algae)</td>
<td><em>N. olivacea</em></td>
<td>NC_000927.1</td>
</tr>
<tr>
<td>C. vulgaris</td>
<td>NC_001865.1</td>
<td></td>
</tr>
<tr>
<td><em>Chlamydomonas reinhardtii</em></td>
<td>BK000554.2</td>
<td></td>
</tr>
<tr>
<td>Secondary (green algal) endosymbionts</td>
<td><em>E. gracilis</em></td>
<td>NC_001603.2</td>
</tr>
<tr>
<td>Mesostigmatales</td>
<td><em>M. viride</em></td>
<td>NC_002186.1</td>
</tr>
<tr>
<td>Colecocahetales</td>
<td>Chaetosphaeridium globosum</td>
<td>NC_004115.1</td>
</tr>
<tr>
<td>Land plants (Embryophyta)</td>
<td>Physcomitrella patens subsp. <em>patens</em></td>
<td>NC_005087.1</td>
</tr>
<tr>
<td>Amborella trichopoda</td>
<td>NC_005086.1</td>
<td></td>
</tr>
<tr>
<td>Calycanthus floridus var. glaucus</td>
<td>NC_004993.1</td>
<td></td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>NC_001320.1</td>
<td></td>
</tr>
<tr>
<td>Pinus koraiensis</td>
<td>NC_004677.1</td>
<td></td>
</tr>
<tr>
<td>A. thaliana</td>
<td>NC_000932.1</td>
<td></td>
</tr>
<tr>
<td>Anthoceros formsae</td>
<td>NC_004543.1</td>
<td></td>
</tr>
<tr>
<td>Adiantum capillus-veneris</td>
<td>NC_004766.1</td>
<td></td>
</tr>
<tr>
<td>Psilotum nudum</td>
<td>NC_003386.1</td>
<td></td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>NC_002762.1</td>
<td></td>
</tr>
<tr>
<td>Lotus japonicus</td>
<td>NC_002694.1</td>
<td></td>
</tr>
<tr>
<td>Nicotiana tabacum</td>
<td>NC_001879.1</td>
<td></td>
</tr>
<tr>
<td>Pinus thumbergii</td>
<td>NC_001631.1</td>
<td></td>
</tr>
<tr>
<td><em>E. virginiana</em> (non-photosynthetic)</td>
<td>NC_001568.1</td>
<td></td>
</tr>
<tr>
<td>Marchantia polymorpha</td>
<td>NC_001319.1</td>
<td></td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>NC_003119.6</td>
<td></td>
</tr>
</tbody>
</table>
essential RNA component (Gegenheimer 1996). Recently it was shown that the P RNAs of *Porphyra* and *Nephroselmis* do not bind to the pre-tRNA substrate in vitro, probably because they are unable to fold into a functional tertiary structure (de la Cruz and Vioque 2003). The P RNA may have lost the substrate binding capability during evolution and the protein may now instead be responsible for this function. In the light of these ideas one may speculate that the SRP of chloroplasts has been subject to a similar evolution.

**Conclusion**

We report here on plastid-encoded SRP RNA genes in eight different organisms. The predictions are supported by the findings that all of the genes have promoter elements compatible with previously reported cyanobacterial promoters and that a high degree of syney is observed. Only two of these SRP RNA genes have previously been reported (Packer and Howe 1998). Chloroplast SRP RNAs are found in the red algae and Chlorophyta branches but not in the higher plants. It seems likely that in these organisms the cpSRP RNA is part of a complex analogous to the eubacterial SRP, and involved in co-translational targeting.

The possibility that the cpSRP RNA gene has been transferred to the nuclear genome in the higher plants may not be formally ruled out. In such a case there would have to be a mechanism to import such RNAs to the chloroplast. More plausible, however, is the possibility that higher plants do not have a functional cpSRP RNA. In this scenario cpSRP54 might have evolved to manage without the RNA in the higher plants, whereas in red algae and in the Chlorophyta branch the SRP is more closely related in structure to the eubacterial SRP.

**Materials and Methods**

**Sources of genomic sequences**

The genomic sequence of *Chlamydomonas reinhardtii* (genome. jgi-psf.org/chlre1/chlre1.home.html and chlre2/chlre2.home.html) was incomplete and represents unpublished material. The genome sequence of *C. merolae* was obtained from the *C. merolae* Genome Project (http://merolae.biol.s.u-tokyo.ac.jp/download/) and genome sequences of *T. pseudonana* were downloaded at http://genome.jgi-psf.org/thaps1/thaps1.home.html. *Chlamydomonas* and *Thalassiosira* data were provided by the DoE Joint Genome Institute.

Twenty-seven complete chloroplast genomes were searched, including 16 land plants (*Embryophyta*), one Coleochaetales, one Mesostigmales, three green algae, three Rhodophyta and three secondary endosymbionts (chloroplasts from *E. gracilis*, *O. sinensis* and *G. theta*) as listed in Table 1.

**Identification of SRP RNA sequences and prediction of RNA secondary structure**

SRP RNA genes were predicted as described previously (Regalia et al. 2002) by applying a combination of pattern searches using *rnabob* (www.genetics.wustl.edu/eddy/software/rnabob) and COVE (Eddy and Durbin 1994). RNA secondary structure predictions were carried out by COVE and Mfold (Zuker 1989) and were complemented by analysis of compensatory base changes. Multiple alignments of RNA primary sequences were made with CLUSTALW (Thompson et al. 1994). Multiple alignments of secondary structures were made with covariance. All predicted RNA sequences and secondary structures are shown in a web supplement at http://bio.lundberg.gu.se/srp04chl/.

**Acknowledgment**

This work was supported by the EU grant QLK3-CT-200-00082 to T.S.

**References**


(Received July 1, 2004; Accepted August 28, 2004)