The octatricopeptide repeat (OPR) protein Raa8 is required for chloroplast *trans*-splicing

Running title: An OPR protein as a chloroplast *trans*-splicing factor

Christina Marx, Christiane Wünsch and Ulrich Kück[#]

Lehrstuhl für Allgemeine und Molekulare Botanik, Ruhr-University Bochum, D-44780 Bochum, Germany

Author for correspondence (e-mail: ulrich.kueck@ruhr-uni-bochum.de; phone: +49 234 3226212; fax: +49 234 3214184)

Abstract

The mRNA maturation of the tripartite chloroplast *psaA* gene from the green alga *Chlamydomonas reinhardtii* depends on various nucleus-encoded factors that participate in *trans*-splicing of two group II introns. Recently, a multiprotein complex was identified that is involved in processing the *psaA* precursor mRNA. Using coupled tandem affinity purification (TAP) and mass spectrometry analyses with *trans*-splicing factor Raa4 as a bait protein, we recently identified a multisubunit ribonucleoprotein (RNP) complex comprising previously characterized *trans*-splicing factors Raa1, Raa3, Raa4, and Rat2, plus novel components. Raa1 and Rat2 share a common structural motif, an octatricopeptide repeat (OPR), that presumably functions as an RNA interaction module. Two of the novel RNP complex components also exhibit a predicted OPR motif, and were therefore considered as potential *trans*-splicing factors. Here, we selected BAC clones encoding these OPR proteins and conducted functional complementation assays using previously generated *trans*-splicing mutants. Our assay revealed that the *trans*-splicing defect of mutant F19 was restored by a new factor we named *RAA8;* molecular characterization of complemented strains verified that Raa8 participates in splicing of the first *psaA* group II intron. Three out of six OPR motifs are located in the C-terminal end of Raa8, which was shown to be essential for restoring *psaA* mRNA *trans*-splicing. Our results support the important role played by OPR proteins in chloroplast RNA metabolism, and also demonstrate that combining TAP and mass spectrometry with functional complementation studies represents a vigorous tool for identifying *trans*-splicing factors.

Introduction

Interaction modules are crucial for the molecular and cellular function of proteins, and a wide repertoire of diverse domains and motifs mediates a variety of regulatory mechanisms. Proteins that exhibit a repeating structural unit with α-helical architecture are classified as members of the α-solenoid superfamily (1). This large family is widely distributed in eukaryotes and includes, among others, the transcription activator-like (TAL) effectors, the tetratricopeptide repeat (TPR) proteins, and the pentratricopeptide repeat (PPR) proteins (2-4). The α-helical structure of diverse members mediates binding to either proteins or nucleic acids.

A well-characterized example for RNA binding proteins is the PPR protein family (5). The degenerated PPR motif is described as two anti-parallel, α-helical tracts of 35 amino acids, whose repeats are often arranged in tandem arrays. Recently, two approaches provided insights into the predictable RNA binding code of this motif (6, 7). Structural analyses indicate that RNA binding is mediated by helical tracts forming a superhelical groove with exposed residues. Despite a low sequence conservation of the PPR motif being restricted to several amino acids, PPR proteins specifically recognize single-stranded RNA sequences through a combinatorial amino acid code (2, 6, 7). Models of the modular recognition mechanism revealed a one nucleotide - one repeat binding of single-stranded RNA, where sequence specificity within each repeat is dictated by exposed key residues of the groove (6, 7).

The PPR protein family includes more than 450 members in land plants and 154 members in green algae (8, 9). These nucleus-encoded RNA binding proteins are involved in a range of post-transcriptional processes of organelle-derived transcripts.

Eukaryotic Cell

Presumably structurally related to the PPR motif, the recently detected octatricopeptide repeat (OPR) proteins harbour degenerated 38 amino acid repeats, thus representing new members of the α-solenoid superfamily (10). The OPR consensus sequence includes the amino acids PPPEW, with the first proline (P) and the tryptophan (W) being the most conserved (11).

OPR proteins were initially identified in the unicellular green alga *Chlamydomonas reinhardtii* with at least 43 nucleus-encoded members (10, 12). Similar to PPR proteins, OPR proteins are primarily predicted to be organelle-localized and proteins characterized so far promote processing, splicing or translation of specific chloroplast transcripts (10-15). Genetic analysis of photosynthesis-deficient mutants revealed that at least six OPR proteins function in transcript stability or translation in *C. reinhardtii*. Tbc2 is part of a 400 kDa complex that is involved in translation of the *psbC* mRNA and contains nine copies of the OPR motif located in the C-terminal region of the protein (11). A dual function was suggested for Tda1, which exhibits eight C-terminal repeats of the OPR motif and mediates trapping of a subset of untranslated *atpA* mRNAs into non-polysomic ribonucleoprotein (RNP) complexes (10). It is also essential for translating the *atpA* transcript. Tab1, a chloroplast translation factor of *psaB* mRNA, is the first OPR protein described to bind RNA *in vitro* (12). Recently, two OPR proteins, Mcg1 and Mbi1, were identified that specifically stabilize chloroplast *petG* and *psbI* mRNA, respectively (16). Stability and translation of chloroplast *psaA* mRNA are functions that are also mediated by OPR protein Taa1 (**t**ranslation of *ps**aA*** mRNA), which regulates *psaA* translation in response to iron limitation (17).

Other OPR containing proteins are required for the maturation of chloroplast *psaA* mRNA. The *psaA* gene encodes the P_{700} chlorophyll α-apoprotein of photosystem I and

Eukaryotic Cell

84 comprises three exons that are scattered around the chloroplast genome (18). Each
85 exon is flanked by group II intron sequences, and thus two *trans*-splicing reactions are
86 necessary to generate mature *psaA* mRNA (19). Remarkably, the first group II intron is
87 tripartite comprising the two *psaA* primary transcripts of exon 1 and 2 as well as an
88 additional small chloroplast RNA (*tscA* RNA) (20). Processing of this *tscA* RNA is a
89 prerequisite for correct assembly and consecutive *trans*-splicing of the first group II
90 intron.
91 So far, two OPR proteins have been described that are required for *psaA trans*-splicing.
92 Rat2 (**R**NA maturation of *psa**A** **t**scA* RNA) and Raa1 (**R**NA maturation of *ps**aA*** RNA)
93 exhibit two and four OPR repeats, respectively (10, 14, 15). Mutants with a defect in
94 genes such as *RAT2* or *RAA1* are not only deficient in *trans*-splicing, but also have a
95 defect in photosynthesis since they lack of the P_{700} chlorophyll α-apoprotein. Genetic
96 analysis of photosystem I mutants revealed that at least 14 nucleus-encoded factors are
97 required for *psaA* mRNA maturation (18, 21, 22).
98 Recently, we were able to demonstrate that both Raa1 and Rat2 are part of a
99 chloroplast RNP complex involved in the first *trans*-splicing reaction (13). This complex
100 was identified by an extensive proteomic analysis and contains four *trans*-splicing
101 factors together with 19 as yet uncharacterized subunits. Of these 19 novel proteins, we
102 identified two further OPR proteins that also might represent *trans*-splicing factors
103 required for group II intron splicing (10, 12, 13).
104 Previously, *trans*-splicing factors were identified by analysis of mutants generated by
105 conventional or insertional mutagenesis (22). For example, we have identified mutant
106 genes by ligation-mediated suppression (LMS) PCR and genome walking in combination
107 with functional complementation analysis using BAC clones (23). As an alternative

approach to discover novel *trans*-splicing factors, we present here a complementation analysis of four as yet uncharacterized *trans*-splicing mutants (21, 24). The mutants were transformed with BAC clones, carrying genes for the above mentioned OPR proteins, which were so far detected solely by TAP followed by mass spectrometry. Evidence is provided that a 269 kDa protein carrying six copies of the OPR motif is required for *trans*-splicing of the first *psaA* group II intron. Importantly, our results extend the list of OPR proteins from *C. reinhardtii* involved in splicing and processing of chloroplast RNAs.

Materials and Methods

Strains and growth conditions

C. reinhardtii wild type 137C and previously described *trans*-splicing mutants F19, FI5, M10, and F31 (21, 24) were grown in liquid TAP (tris acetate phosphate) medium or on TAP agar plates (25). For photoautotrophic growth, test strains were spread onto HS (high salt minimal) or TAP agar plates (25). To generate autolysin, we used high efficiency mating strains CC-621 and CC-622 (25).

Transformation

Nuclear transformation was achieved by agitation with glass beads (26). To remove cell walls, cells were treated for at least 1 h with autolysin before transformation. Each transformation experiment was carried out at least three times with 0.5 to 10 µg BAC DNA and spread on single TAP agar plates. After 12 h to 24 h incubation under low light, plates were selected under high light conditions for 2 to 3 weeks.

Construction of plasmids

Recombinant plasmids and oligonucleotides used for PCR and RT-PCR experiments, or to generate transgenic algal strains are listed in Table S1 and Table S2. BAC clones of BAC library CRCCBa were obtained from CUGI (Clemson University, USA, https://www.genome.clemson.edu/cgi-bin/orders?page=productGroup&service=bacrc&productGroup=162). After plasmid isolation, BAC DNA was verified by restriction enzyme hydrolysis and partial sequencing.

Derivatives of BAC 31O13 were generated to identify regions sufficient for rescue of mutant F19 by partial restriction of 31O13 with *Sgs*I (31O13-1, 31O13-2), or *Bln*I (31O13-3, 31O13-4). For subcloning of *RAA8*, 31O13 was first digested with *Sfu*I and *Cla*I, resulting in plasmid 31O13-5. In a second step, 31O13-5 was partially digested with *Mlu*I (31O13-6). 31O13-6 only comprises *RAA8* and Cre10.g440050. To generate a plasmid that contains solely *RAA8*, 31O13-6 was partial digested with *Pvu*I resulting in 31O13-7.

Molecular genetic techniques

Standard molecular techniques were used as reported elsewhere (13, 23). *C. reinhardtii* total RNA and RNA blot experiments were prepared as described previously (15, 18, 23). After transfer of RNA samples to nylon membranes, hybridization was performed with a radioactively labelled probe. For isolation of genomic DNA, algal cells were treated as described elsewhere (27).

cDNA synthesis was conducted with the SuperScript™ III first-strand synthesis system for RT-PCR (Invitrogen) and purified using Amicon™ columns (Millipore). RT-PCR and PCR were performed using Taq DNA polymerase (Eppendorf), Phusion® high-fidelity DNA polymerase (Thermo Scientific), GoTaq® DNA polymerase (Promega), and FastStart Taq DNA polymerase (Roche).

For sequencing of the *RAA8* gene locus, PCR fragments were directly sequenced or cloned into pDrive (Qiagen). For verification of point mutations, two independent PCR reactions were conducted with different primer combinations followed by sequencing.

For PsaA immunoblot analysis, total protein extracts were loaded on 12% SDS-polyacrylamide gels with 6 M urea. After gel electrophoresis and blotting onto a PVDF

membrane (Roche), the membrane was blocked with 5% non-fat dry milk in washing buffer. After incubation with rabbit polyclonal PsaA antibody (Agrisera) at 4°C overnight, the membrane was decorated with peroxidase-linked anti-rabbit IgG for 1 h. Signals were detected using western ECL substrate (Biorad).

Bioinformatic analyses

Gene models and corresponding protein data were obtained from Phytozome v10.1 (http://phytozome.jgi.doe.gov/pz/portal). Protein motifs and domains were predicted by MEME suite (28) and Motif Scan (29). Alignments of protein sequences were visualized using ClustalX (30). RNA binding regions were predicted by BindN (31). Secondary structure was analyzed using Phyre^2 (32). Chloroplast target peptides were predicted by WoLF PSORT (33), TargetP (34), ChloroP (35), and PredAlgo (36).

Results

Functional complementation analysis of class C mutants using BAC DNA

Recently, we identified a chloroplast RNP complex containing *trans*-splicing factors, intron RNA, and as yet uncharacterized protein subunits (13). This complex was shown to be involved in processing *psaA* exon 1 and 2 precursors. Two of these *trans*-splicing factors, Rat2 and Raa1, were described as OPR proteins, and two uncharacterized proteins were considered as further OPR proteins of the splicing complex (10, 12). These are Cre17.g698750 and Cre10.g440000, and besides OPR repeats both have no conserved domains. To determine whether both OPR proteins are required for *trans*-splicing of *psaA* mRNA, BAC clones encoding these uncharacterized OPR components were selected for a complementation approach. BAC clones 13K19 and 31O13 were obtained from the CUGI BAC library and encode Cre17.g698750 and Cre10.g440000, respectively.

The BAC clones were used for complementation studies of photosynthesis mutants. These mutants were previously generated by random mutagenesis, and further characterized to detect defects in *trans*-splicing of the *psaA* mRNA (21, 24). Depending on the splicing defect, mutants can be divided in three distinct classes, namely A, B, and C. Mutants unable to splice the second group II intron are designated as class A, and class B mutants are blocked in both the first and second *trans*-splicing reactions (21). Class C mutants show a defect in the first *trans*-splicing reaction, and they are either unable to fuse exon 1 and 2, or alternatively have a defect in *tscA* RNA 3' end processing. Here, we chose a number of class C mutants from a type culture collection

(ChlamyStation, Paris, France), namely F19, M10, F31, and FI5 (21, 24), for a complementation approach.

The four class C mutants were separately transformed with BAC clones 13K19 and 31O13, and transformants were selected under high light conditions. Growth under these growth conditions indicated restoration of photosystem I activity, and thus complementation of the *psaA trans*-splicing defect. Transformation of M10, FI5, and F31 with each of the two BACs revealed no transformants, while several transformants were obtained from transformation of F19 with BAC 31O13.

Photoautotrophic growth tests were performed on four selected transformants derived from complementation of F19 with BAC 31O13 (T404-2/6, T401-1/1, T401-2/1, and T401-2/2) (Fig. 1a). TAP agar plates contained acetate as an alternative carbon source, whereas HS medium was used as a minimal medium. Therefore, photosystem I deficient strains are unable to grow under high light conditions both on TAP plates and on HS medium without acetate, but are able to grow under low light conditions on TAP agar plates (Fig. 1a, Fig. S1). Wild type 137C and mutant F19 served as controls for photoautotrophic growth tests.

The four selected transformants were able to grow under high light conditions on HS and TAP agar plates, thus demonstrating a restored photosystem I activity. For Northern analysis, RNA was isolated from 137C, F19, and the four complementation strains. A *psaA* exon 1 probe detected a 2,800 nt signal representing the mature *psaA* transcript in the wild type 137C and all transformants, whereas the signal of about 400 nt corresponds to the *psaA* exon 1 primary transcript in mutant F19 (Fig. 1b). Furthermore, a *tscA* probe detected the 450 nt processed *tscA* transcript in all analyzed strains. Rehybridization with an *RBCS1* probe served as a loading control. The Northern

analysis confirmed that mutant F19 has a defect in *trans*-splicing of the first group II intron, but not in 3' end processing of *tscA* RNA. We previously showed that *trans*-splicing mutants with a defect in *tscA* processing generated precursor RNA molecules of 1,700 nt and 2,800 nt that are clearly distinguished from the processed 450 nt *tscA* RNA (15, 37). All analyzed complementation strains exhibited restored photosystem I activity and regained the ability to generate mature *psaA* mRNA.

To further analyze the transformants, we performed immunodetection with α-PsaA antibody. As expected, since mutant F19 was unable to generate mature *psaA* mRNA transcript, no PsaA protein was detected. In contrast, all tested transformants showed a signal for PsaA (Fig. 1c). Comparison with the wild type signal revealed slightly reduced PsaA signal intensities in all transformants, presumably as a consequence of ectopic integration and thus suboptimal gene expression in the complemented strains.

These analyses revealed that recombinant BAC clone 31O13 carrying Cre10.g440000 and twelve further genes restores the *trans*-splicing defect of mutant F19. Since Cre10.g440000 is most likely responsible for rescuing the *trans*-splicing mutant F19, we called the gene *RAA8* (RNA maturation of *psaA* mRNA).

To rule out that another gene apart from *RAA8* is necessary for complementation, we generated seven derivatives of BAC 31O13 (Fig. 2). After transformation of recipient F19 with all plasmids, we only obtained photosynthesis-competent strains, when the full length *RAA8* gene was present on the recombinant BAC plasmid. We failed to obtain transformants when 31O13-1 lacking *RAA8* and 31O13-4 carrying a truncated version of *RAA8* (exon 1 to exon 5, coding for 1,973 aa residues) were used for transformation.

As further proof for *psaA* pre-mRNA *trans*-splicing, RT-PCR was performed with cDNA to detect the mature *psaA* mRNA in photosynthetic-competent transformants. This

included strains complemented with 31O13-7, which carries a deletion upstream of the ATG start codon of *RAA8*. As a control, cDNA from wild type and F19 was used, demonstrating again that F19 is a class C mutant, unable to perform exon 1 - exon 2 *trans*-splicing (Fig. S2). In summary, the lack of *RAA8* in F19 is responsible for the *psaA* pre-mRNA splicing defect.

Photosynthesis mutant F19 carries an in-frame stop codon in *RAA8*

Since the BAC derivative 31O13-7 was sufficient to complement mutant F19, the F19 mutation must affect *RAA8*, the only gene encoded by 31O13-7, with a predicted gene locus comprising nine exons (Fig. 3a).

Previously, F19 was generated by 5-fluorouracyl mutagenesis, which is usually associated with DNA point mutations (38). To identify mutations responsible for the mutant phenotype, we performed PCR analysis and sequencing of the F19 DNA. By sequencing the entire *RAA8* locus (exon 1 - exon 9, 10,597 bp), we detected a point mutation at base pair position 5,537 within exon 3. PCR analysis of wild type 137C DNA and 31O13-7 followed by sequence alignment revealed that the point mutation is unique for mutant F19 (Fig. 3b). The transversion of guanine to thymine leads to an in-frame stop codon instead of glutamic acid (E1722). Whereas *RAA8* encodes a protein of 2,745 residues, the stop codon in F19 leads to premature translation termination resulting in a truncated protein of 1,720 residues. We conclude that the N-terminal part of Raa8 alone is not sufficient for complementation.

Raa8 has a molecular mass of 269 kDa and exhibits six putative OPR repeats

Raa8 has a predicted size of 269 kDa and is the largest *trans*-splicing factor described so far (Fig. 4a). Peptide counts obtained from mass spectrometry analysis (13) revealed a coverage of about 37.7% of the total Raa8 sequence. Using TargetP, ChloroP, PredAlgo, and WoLF PSORT, a chloroplast localization can be predicted for Raa8 (13, 33-36). In addition to the putative chloroplast signal peptide, Motif Scan identified several stretches of low complexity regions (29). These regions exhibit little diversity regarding their amino acid composition and are common in many eukaryotic proteins (39-41). Raa8 carries an alanine-rich region that covers nearly the entire protein, in addition to serine-, glycine-, glutamine- and proline-rich stretches (Fig. 4a). The protein exhibits a high content of alanine (24.4%), glycine (11.8%), and leucine (12.0%). Moreover, BindN predicts seven putative RNA binding residues (31). Analysis of Raa8, together with several OPR proteins such as Tab1, Tda1, Tbc2, and *trans*-splicing factors, by the program MEME identified at least six OPR repeats (28).

Sequence conservation of the OPR motif is restricted to several amino acids, including the mostly conserved first proline and the tryptophan of the degenerated PPPEW pattern (Fig. 4b) (10, 11). Since OPR proteins belong to the α-solenoid superfamily with α-helical architecture, we performed secondary structure prediction of partial Raa8 sequences using Phyre^2 (32). Raa8 contains several paired α-helices with a high confidence value. Some of these α-helices match the six OPR repeats and all are located in the C-terminal half of Raa8 (Fig. S3).

Our functional complementation studies revealed that 31O13-4 carrying a truncated version of *RAA8* (exon 1 - exon 5) was not able to rescue mutant F19. *In silico* analysis

revealed that 31O13-4 encodes a 192 kDa instead of a 269 kDa protein. This indicates that the genomic region of exon 1 to exon 5 is not sufficient for functional complementation. Importantly, this truncated gene version does not encode the last three C-terminal OPR repeats. We therefore conclude that the complete set of OPR repeats is required for the *trans*-splicing function of Raa8 and thus for complementation.

Discussion

Here, we demonstrate that functional complementation studies as a follow up to a previous TAP and mass spectrometry approach enabled us to identify a gene defect in a previously generated *trans*-splicing class C mutant. Detection of the mutation in F19 revealed an uncharacterized OPR protein as a *trans*-splicing factor, which we named Raa8. This protein is part of a RNP complex that was previously identified by TAP and mass spectrometry, and is involved in the first *trans*-splicing reaction of *psaA* mRNA. Hence, at least five *trans*-splicing factors, Rat2, Raa1, Raa3, Raa4, and the newly identified Raa8, are part of this complex.

Functional complementation studies using photosynthesis mutants to complement a combined TAP and mass spectrometry approach represents a vigorous tool to identify *trans*-splicing factors

Homologous recombination is a prerequisite to generate mutant strains by insertion or deletion of distinct genes. In *C. reinhardtii*, the frequency of homologous recombination is exceptionally low and most often transforming DNA integrates randomly by the non-homologous end joining repair pathway (42). Up to now, a rapid method to generate targeted mutations with a moderate success rate is still missing.

C. reinhardtii mutants are commonly used as sources for forward genetic approaches to identify genes controlling specific processes. Usually, the screening of mutants relies on a distinct mutant phenotype that allows uncovering specific gene functions based on complementation analysis. Commonly, insertional mutagenesis by random integration of

a selection marker was used to generate mutants with a desired phenotype (42, 43). More recently, an alternative approach used whole genome sequencing to identify gene mutations, but is hampered by the mutational profile of strains analyzed (44). Despite single nucleotide polymorphisms (indels) that lead to a high number of changes to the reference genome, additional chromosomal rearrangements or suppressed recombination impede the identification of mutations (44, 45).

Here, we use an alternative approach to identify genes involved in chloroplast *trans*-splicing by using data from extensive TAP and mass spectrometry analysis of an RNP complex (13). This RNP complex is involved in the first *trans*-splicing reaction and contains four characterized *trans*-splicing factors and additional as yet uncharacterized protein subunits. We conducted functional complementation studies using previously generated *trans*-splicing mutants. More than 30 years ago, these *trans*-splicing mutants of *C. reinhardtii* were generated by chemical treatment or UV radiation (21, 24, 46). Since the long-term propagation of strain collections leads to accumulation of spontaneous mutations, identification of mutations by whole genome sequencing is time consuming and laborious, requiring numerous back-crossings and segregation analyses. Here, we took advantage of previously generated class C mutants and used BAC DNA encoding either of the two uncharacterized OPR proteins, Cre10.g440000 (Raa8) and Cre17.g698750, for transformation. Both proteins were already known to be part of the *trans*-splicing complex (13). Functional complementation studies and further molecular characterization of complemented strains enabled us to identify the novel *trans*-splicing factor Raa8. We univocally demonstrate that Raa8 is involved in *trans*-splicing of the first *psaA* group II intron, but is not crucial for *tscA* 3' end processing. Thus, this is the first described OPR protein required exclusively for splicing group II intron RNA. With the

discovery of Raa8, we extend the rather short list of OPR proteins from *C. reinhardtii* that are involved in RNA splicing and/or processing.

Raa8 is a member of the OPR protein family that functions in chloroplast *psaA* mRNA processing

Mitochondria and chloroplasts originated from symbiotic incorporation of either α-proteobacteria or cyanobacteria (47, 48). Thus, both organelles show diverse similarities to bacteria such as their circular genome, replication mode, RNA processing steps, as well as their protein transport and translation machinery (49-52).

The endosymbiosis resulted in reconstruction of genetic information mediated by gene transfer from the ancestral organelle genomes to the nucleus (53). The persistent compartmentation of the genetic systems requires crosstalk between different retrograde and anterograde signalling pathways to control plastid and mitochondrial biogenesis and their distinct biochemical processes (53-55). Organelle genes are often constitutively expressed and regulated at their post-transcriptional level. Therefore, nucleus-encoded factors are required to control functionality of polycistronic transcripts by affecting their stability, editing, or maturation (56-58).

In algae and plants, chloroplast RNA metabolism is mostly controlled by PPR and OPR proteins, both members of the α-solenoid superfamily (8, 10, 12, 58, 59). Numerous PPR proteins mediate translation, RNA stability, cleavage, editing and splicing (8, 60). Whereas PPRs expanded in land plants, the sole OPR protein from *A. thaliana* (61, 62) contains in addition to four OPR repeats an RAP (**R**NA binding domain abundant in **Ap**icomplexans) domain that also is present in *C. reinhardtii* chloroplast proteins Taa1

and *trans*-splicing factor Raa3 (17, 63). Rap functions by binding to the 16S rRNA and via 5' end processing mediates maturation of 16S rRNA (62). Similar to *C. reinhardtii* OPR proteins such as Tda1 and Tbc2, the OPR repeats are mainly localized in the C-terminal half of Rap (10, 11, 62). Interestingly, the C-terminal part of Tda1 was shown to be sufficient to promote *atpA* mRNA translation. Moreover, complementation studies using truncated versions of both *TAB1* and *TDA1,* as mentioned in the introduction, suggest that the entire set of OPR repeats is required for functionality of the proteins (10, 12).

The striking role of OPR-containing protein regions in chloroplast RNA metabolism is also supported by the results of this work. Similar to other OPR proteins such as Tda1 or Tbc2, our analysis revealed that the six OPR domains of Raa8 are localized in the C-terminal half of the protein primary structure. Characterization of mutant F19 revealed that a stop codon in exon 3 of *RAA8* results in a truncated protein. Thus, the mutant encodes a 167 kDa protein lacking five out of six OPR repeats. Moreover, restoration of photosynthesis in F19 failed when a truncated 192 kDa protein (expressed by derivative 31O13-4) exhibiting only three out of six OPRs was used for complementation. Our results show that the C-terminal region of Raa8 is required for *trans*-splicing function, suggesting that the missing OPRs in the truncated versions provide essential functions for group II intron splicing.

In vivo, trans-splicing of *psaA* mRNA requires *trans*-splicing factors, which mostly share structural features such as low complexity regions and OPR motifs. Accurate splicing of the group II introns is a prerequisite for successful translation of PsaA. Therefore, OPR proteins may adjust the tripartite intron fragments of the first group II intron through nucleotides binding to OPR repeats. The common α-helical structure of PPR and OPR

repeats suggest that OPR motifs similar to the PPR motifs bind single nucleotides of the target RNA (6, 7).

Besides a direct function of OPR proteins such Raa8 in *trans*-splicing, other OPR proteins such as Raa1 and Rat2 are involved in 3' end processing of *tscA* RNA. Presumably, these proteins enhance proper processing of the precursor transcript through sequence-specific binding of the OPR motif. Conversely, proteins without OPR repeats such as Raa4 and Raa3 may act as chaperones. Previously, binding of Raa4 to RNA was demonstrated, thus suggesting that Raa4 promotes folding and stability of the intron RNA (23).

The result of our work expands the roles of OPR proteins in mRNA splicing. The characterization of chloroplast-localized OPR proteins supports their specific function in various roles of chloroplast RNA metabolism, and is presumably not restricted to algal organisms, but will most probably also be discovered in higher plant chloroplasts.

Acknowledgements:

This work was funded by the Deutsche Forschungsgemeinschaft (KU 517/13-1). We wish to thank Ingeborg Godehardt for excellent technical assistance, and the ChlamyStation (France) for providing algal strains. We are grateful to Dr. J. Jacobs for helpful comments on the manuscript.

References

1. **Kobe B, Kajava AV.** 2000. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends Biochem Sci **25:**509-515.
2. **Deng D, Yan C, Wu J, Pan X, Yan N.** 2014. Revisiting the TALE repeat. Protein Cell **5:**297-306.
3. **Blatch GL, Lässle M.** 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays **21:**932-939.
4. **Small ID, Peeters N.** 2000. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci **25:**46-47.
5. **Manna S.** 2015. An overview of pentatricopeptide repeat proteins and their applications. Biochimie **113:**93-99.
6. **Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang HW, Wang J, Zhu JK, Shi Y, Yan N.** 2013. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature **504:**168-171.
7. **Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I.** 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet **8:**e1002910.
8. **Barkan A, Small I.** 2014. Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol **65:**415-442.
9. **Tourasse NJ, Choquet Y, Vallon O.** 2013. PPR proteins of green algae. RNA Biol **10:**1526-1542.
10. **Eberhard S, Loiselay C, Drapier D, Bujaldon S, Girard-Bascou J, Kuras R, Choquet Y, Wollman FA.** 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.

11. **Auchincloss AH, Zerges W, Perron K, Girard-Bascou J, Rochaix JD.** 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.

12. **Rahire M, Laroche F, Cerutti L, Rochaix JD.** 2012. Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. Plant J **72:**652-661.

13. **Jacobs J, Marx C, Kock V, Reifschneider O, Fränzel B, Krisp C, Wolters D, Kück U.** 2013. Identification of a chloroplast ribonucleoprotein complex containing *trans*-splicing factors, intron RNA, and novel components. Mol Cell Proteomics **12:**1912-1925.

14. **Merendino L, Perron K, Rahire M, Howald I, Rochaix JD, Goldschmidt-Clermont M.** 2006. A novel multifunctional factor involved in *trans*-splicing of chloroplast introns in *Chlamydomonas*. Nucleic Acids Res **34:**262-274.

15. **Balczun C, Bunse A, Hahn D, Bennoun P, Nickelsen J, 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.

16. **Wang F, Johnson X, Cavaiuolo M, Bohne AV, Nickelsen J, Vallon O.** 2015. Two *Chlamydomonas* OPR proteins stabilize chloroplast mRNAs encoding small subunits of photosystem II and cytochrome b_6f. Plant J **82:**861-873.

17. **Lefebvre-Legendre L, Choquet Y, Kuras R, Loubéry S, Douchi D, Goldschmidt-Clermont M.** 2015. A nucleus-encoded helical-repeat protein which is regulated by iron availability controls chloroplast *psaA* mRNA expression in *Chlamydomonas reinhardtii*. Plant Physiol **167:**1527-1540.

18. **Kück U, Choquet Y, Schneider M, Dron M, Bennoun P.** 1987. Structural and transcription analysis of two homologous genes for the P_{700} chlorophyll α-apoproteins in *Chlamydomonas reinhardtii*: evidence for *in vivo trans*-splicing. EMBO J **6:**2185-2195.

19. **Glanz S, Kück U.** 2009. *Trans*-splicing of organelle introns - a detour to continuous RNAs. Bioessays **31:**921-934.

20. **Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD.** 1991. A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. Cell **65:**135-143.

21. **Goldschmidt-Clermont M, Girard-Bascou J, Choquet Y, Rochaix JD.** 1990. *Trans*-splicing mutants of *Chlamydomonas reinhardtii*. Mol Gen Genet **223:**417-425.

22. **Jacobs J, Glanz S, Bunse-Grassmann A, Kruse O, Kück U.** 2010. RNA *trans*-splicing: identification of components of a putative chloroplast spliceosome. Eur J Cell Biol **89:**932-939.

23. **Glanz S, Jacobs J, Kock V, Mishra A, Kück U.** 2012. Raa4 is a *trans*-splicing factor that specifically binds chloroplast *tscA* intron RNA. Plant J **69:**421-431.

24. **Girard J, Chua NH, Bennoun P, Schmidt G, Delosme M.** 1980. Studies on mutants deficient in the photosystem I reaction centers in *Chlamydomonas reinhardtii*. Curr Genet **2:**215-221.

25. **Harris EH.** 1989. The *Chlamydomonas* sourcebook, A comprehensive guide to biology and laboratory use. Academic Press, San Diego, CA.

26. **Kindle KL.** 1990. High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A **87:**1228-1232.

27. **Newman SM, Boynton JE, Gillham NW, Randolph-Anderson BL, Johnson AM, Harris EH.** 1990. Transformation of chloroplast ribosomal RNA genes in *Chlamydomonas*: molecular and genetic characterization of integration events. Genetics **126:**875-888.

28. **Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS.** 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res **37:**202-208.

29. **Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Falquet L.** 2004. MyHits: a new interactive resource for protein annotation and domain identification. Nucleic Acids Res **32:**332-335.

30. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-2948.

31. **Wang L, Brown SJ.** 2006. BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic Acids Res **34:**243-248.

32. **Kelley LA, Sternberg MJ.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Prot **4:**363-371.

33. **Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K.** 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res **35:**585-587.

34. **Emanuelsson O, Brunak S, von Heijne G, Nielsen H.** 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. Nat Prot **2:**953-971.

35. **Emanuelsson O, Nielsen H, Von Heijne G.** 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci **8:**978-984.

36. **Tardif M, Atteia A, Specht M, Cogne G, Rolland N, Brugiere S, Hippler M, Ferro M, Bruley C, Peltier G, Vallon O, Cournac L.** 2012. PredAlgo: a new subcellular localization prediction tool dedicated to green algae. Mol Biol Evol **29:**3625-3639.

37. **Hahn D, Nickelsen J, Hackert A, Kück U.** 1998. A single nuclear locus is involved in both chloroplast RNA *trans*-splicing and 3′ end processing. Plant J **15:**575-581.

38. **Murakami S, Kuehnle K, Stern DB.** 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.

39. **Golding GB.** 1999. Simple sequence is abundant in eukaryotic proteins. Protein Sci **8:**1358-1361.

40. **Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D.** 1999. A census of protein repeats. J Mol Biol **293:**151-160.

41. **DePristo MA, Zilversmit MM, Hartl DL.** 2006. On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. Gene **378:**19-30.

42. **Jinkerson RE, Jonikas MC.** 2015. Molecular techniques to interrogate and edit the *Chlamydomonas* nuclear genome. Plant J **82:**393-412.

43. **Nickelsen J, Kück U.** 2000. The unicellular green alga *Chlamydomonas reinhardtii* as an experimental system to study chloroplast RNA metabolism. Naturwissenschaften **87:**97-107.

44. **Dutcher SK, Li L, Lin H, Meyer L, Giddings TH, Jr., Kwan AL, Lewis BL.** 2012. Whole-genome sequencing to identify mutants and polymorphisms in *Chlamydomonas reinhardtii*. G3 (Bethesda) **2:**15-22.

45. **Lin H, Miller ML, Granas DM, Dutcher SK.** 2013. Whole genome sequencing identifies a deletion in protein phosphatase 2A that affects its stability and localization in *Chlamydomonas reinhardtii*. PLoS Genet **9:**e1003841.

46. **Garnier J, Guyon D, Picaud A.** 1979. Characterization of new strains of nonphotosynthetic mutants of *Chlamydomonas reinhardtii*. I: Fluorescence, photochemical activities, chlorophyll-protein complexes. Plant Cell Physiol **20:**1013-1027.

47. **Gray MW, Burger G, Lang BF.** 2001. The origin and early evolution of mitochondria. Genome Biol **2:**Reviews 1018.1011–1015.

48. **Gould SB, Waller RF, McFadden GI.** 2008. Plastid evolution. Annu Rev Plant Biol **59:**491-517.

49. **Wakasugi T, Tsudzuki T, Sugiura M.** 2001. The genomics of land plant chloroplasts: Gene content and alteration of genomic information by RNA editing. Photosynth Res **70:**107-118.

50. **Small ID, Rackham O, Filipovska A.** 2013. Organelle transcriptomes: products of a deconstructed genome. Curr Opin Microbiol **16:**652-658.

51. **Wang P, Dalbey RE.** 2011. Inserting membrane proteins: the YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. Biochim Biophys Acta **1808:**866-875.

52. **Gray MW.** 2015. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. Proc Natl Acad Sci U S A doi:10.1073/pnas.1421379112.

53. **Timmis JN, Ayliffe MA, Huang CY, Martin W.** 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet **5:**123-135.

54. **Woodson JD, Chory J.** 2008. Coordination of gene expression between organellar and nuclear genomes. Nat Rev Genet **9:**383-395.

55. **Martin W, Koonin EV.** 2006. Introns and the origin of nucleus-cytosol compartmentalization. Nature **440:**41-45.

56. **Stern DB, Goldschmidt-Clermont M, Hanson MR.** 2010. Chloroplast RNA metabolism. Annu Rev Plant Biol **61:**125-155.

57. **Binder S, Brennicke A.** 2003. Gene expression in plant mitochondria: transcriptional and post-transcriptional control. Philos Trans R Soc Lond B Biol Sci **358:**181–189.

58. **de Longevialle AF, Small ID, Lurin C.** 2010. Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. Mol Plant **3:**691-705.

59. **Germain A, Hotto AM, Barkan A, Stern DB.** 2013. RNA processing and decay in plastids. Wiley Interdiscip Rev RNA **4:**295-316.

60. **Schmitz-Linneweber C, Small I.** 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci **13:**663-670.

61. **Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H.** 2007. A novel class of bacteria-induced small RNAs in *Arabidopsis*. Genes Dev **21:**3123-3134.

62. **Kleinknecht L, Wang F, Stübe R, Philippar K, Nickelsen J, Bohne AV.** 2014. RAP, the sole octotricopeptide repeat protein in *Arabidopsis*, is required for chloroplast 16S rRNA maturation. Plant Cell **26:**777-787.

63. **Lee I, Hong W.** 2004. RAP - a putative RNA-binding domain. Trends Biochem Sci **29:**567-570.

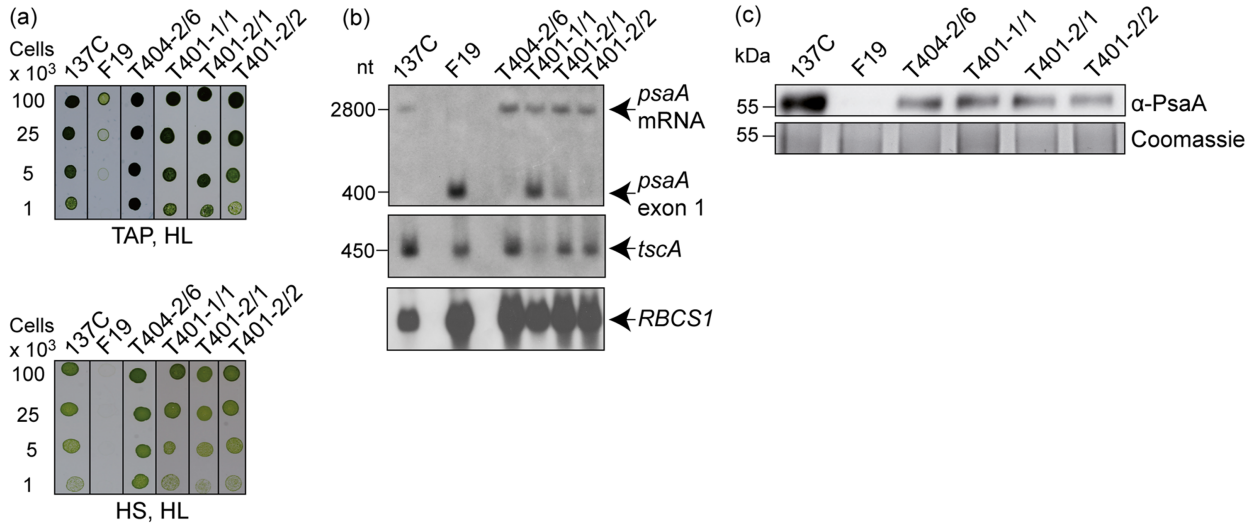
588 **FIG 1 Functional complementation of *trans*-splicing mutant F19. (a)** Test for
589 photoautotrophic growth of wild type 137C, *trans*-splicing mutant F19, and selected
590 transformants (T404-2/6, T401-1/1, T401-2/1, T401-2/2) obtained from transformation of
591 F19 with BAC 31O13. Cells were spread onto acetate containing (TAP), or acetate
592 lacking (HS) agar plates with cell densities as indicated and incubated under high light
593 (HL). 137C and complemented strains are able to grow on HS or TAP plates under HL in
594 contrast to mutant F19 that is photosynthetically deficient. **(b)** Northern analysis of 137C,
595 F19, and complemented transformant strains. Filters were consecutively hybridized with
596 probes specific for *psaA* exon 1, *tscA*, and *RBSC1* transcripts (as a loading control). **(c)**
597 Immunodetection of the PsaA protein in 137C, F19, and four transformants. Coomassie
598 stained gel served as a loading control.

599 **FIG 2 Complementation analysis of mutant F19.** The genomic region inserted in BAC
600 31O13 is indicated at the top and subclones are listed below. 31O13 carries 13 genes of
601 chromosome 10, in the figure abbreviated to the final six characters (e.g.
602 Cre10.g440050 = 440050). BAC vector sequences are indicated in black. Arrows
603 represent the predicted ORFs of the corresponding genes, whereas dashed lines
604 indicate deleted regions. The gene encoding *RAA8* (Cre10.g440000) is marked as a red
605 arrow. The observed functional complementation of F19 is indicated with "+", while "-"
606 indicates lack of complementation. The asterisk indicates a truncated version of *RAA8*.

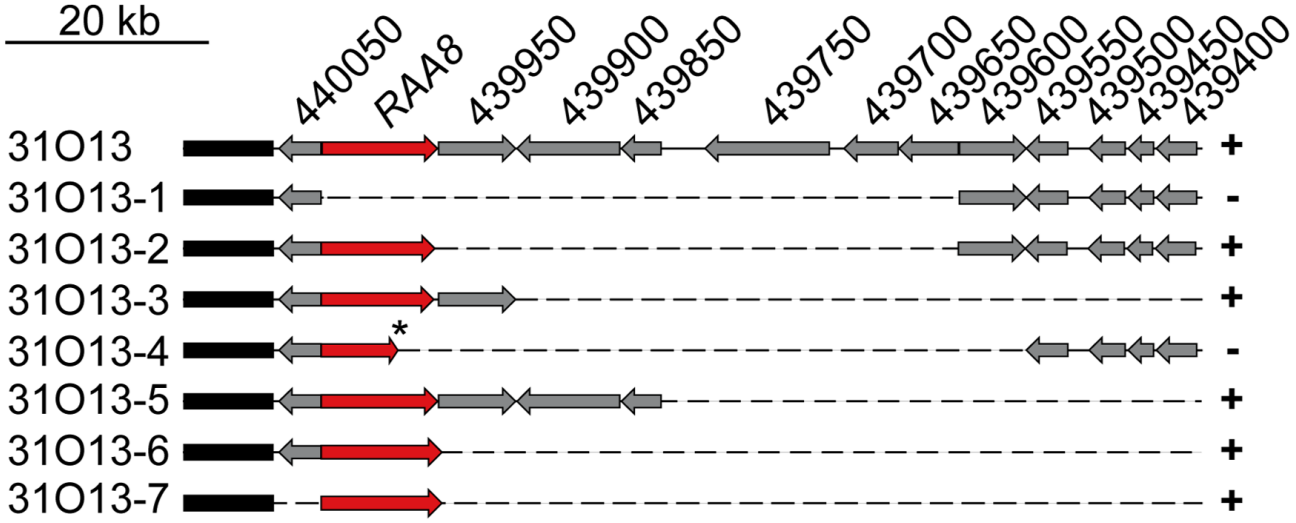
607 **FIG 3 Map of the *RAA8* locus (a)** Exon-intron organization of *RAA8.* Corresponding 5'
608 and 3' UTRs at both ends are marked as open boxes. The red bar shows the position of
609 the detected point mutation in mutant F19. **(b)** Sequence alignment of the partial
610 sequence of *RAA8* exon 3. The sequence shadowed in red indicates sequences that
611 deviate between the wild type and mutant sequence. A guanine to thymine transversion
612 in F19 leads to a stop codon (*).

613 **FIG 4 The *RAA8* gene encodes a 269 kDa protein that exhibits six predicted C-**
614 **terminal OPR motifs. (a)** Scheme of the primary structure of the Raa8 polypeptide and
615 two truncated versions. *In silico* analyses predicted several low complexity regions (gly-
616 rich, pro-rich, gln-rich, ala-rich, ser-rich), RNA binding residues, a chloroplast transit
617 peptide (cTP), and OPR motifs. Both, F19 and 31O13-4, encode a truncated protein. **(b)**
618 Six OPR motifs in Raa8 were predicted by the MEME program (28). The alignment was
619 visualized using ClustalX (30) with a default color scheme assigned for each residue that
620 meets the amino acid profile of the alignment at that position. The consensus sequence
621 defined by MEME is indicated below. Numbers in front of each OPR motif indicate the
622 positions of the first amino acid. Numbers behind each motif represent the *p*-value
623 defined by MEME using characterized OPR proteins like Raa1 and Rat2.

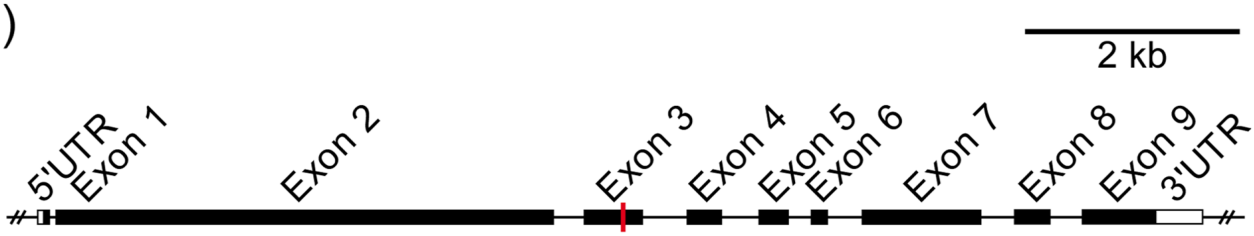
1



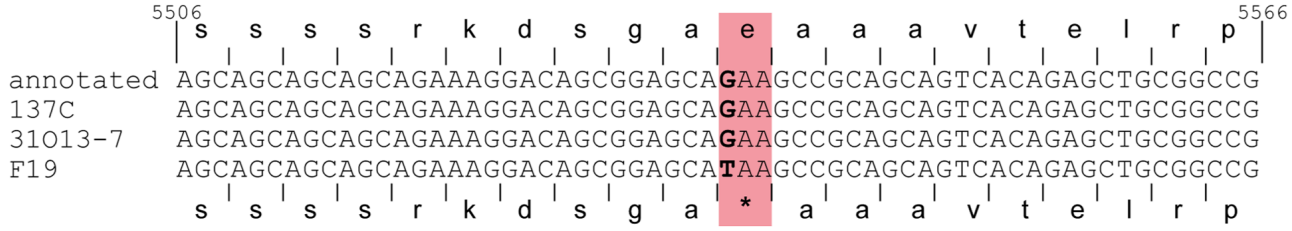
2



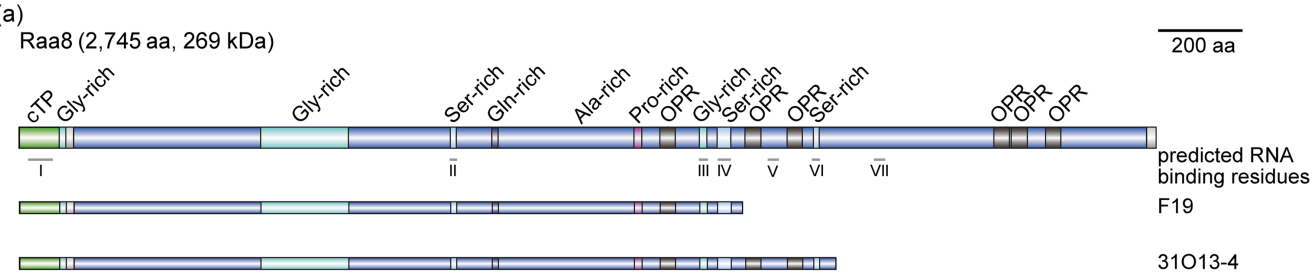
3 (a)



(b)



4 (a)



(b)