Curr Genet (1995) 27: 270–279

ORIGINAL PAPER

Monique Turmel · Yves Choquet
Michel Goldschmidt-Clermont · Jean-David Rochaix
Christian Otis · Claude Lemieux

The *trans*-spliced intron 1 in the *psa*A gene of the *Chlamydomonas* chloroplast: a comparative analysis

Received: 29 June 1994 / 30 August 1994

Abstract In the secondary structure model that has been proposed for the *trans*-spliced intron 1 in the *Chlamydomonas reinhardtii psa*A gene, a third RNA species (*tsc*A RNA) interacts with the 5′ and 3′ intron parts flanking the exons to reconstitute a composite structure with several features of group-II introns. To test the validity of this model, we undertook the sequencing and modelling of equivalent introns in the *psa*A gene from other unicellular green algae belonging to the highly diversified genus *Chlamydomonas*. Our comparative analysis supports the model reported for the *C. reinhardtii psa*A intron 1, and also indicates that the 5′ end of the *tsc*A RNA and the region downstream from the *psa*A exon 1 cannot be folded into a structure typical of domain I as described for most group-II introns. It is possible that a fourth RNA species, yet to be discovered, provides the parts of domain I which are apparently missing.

Key words *Chlamydomonas* chloroplast DNA · *Trans*-spliced group-II intron · *tsc*A RNA · $tRNA^{Ile}$

Introduction

Group-II introns have been found in fungal and plant mitochondria, in chloroplasts (reviewed in Michel et al. 1989; Jacquier 1990; Bonen 1993; Saldanha et al. 1993), and more recently in cyanobacteria and purple bacteria (Ferat and Michel 1993). Despite their high degree of sequence divergence, they share a conserved secondary structure necessary for splicing. This structure consists of six helical domains (I to VI) radiating from a central wheel. As in nuclear mRNA introns, splicing is initiated by the formation of an intron lariat in which the 5′ end of the intron is ligated by a 2′–5′ phosphodiester bond to a residue (designated the branch site), usually an A, near the 3′ end of the intron. Group-II introns have also conserved 5′- and 3′-boundary sequences (GUGYG and AY), which are similar to those found in nuclear mRNA introns. Although no three-dimensional structure has been reported for group-II introns, the functions of domains I, V and VI have been elucidated. Domain V, the most highly conserved substructure, interacts with sequences within domain I to form the catalytic core, while domain VI, which is not required for ribozyme activity, contains the branch site and functions in positioning the 3′ splice site. Short sequences of domain I, designated exon-binding sites 1 and 2 (EBS1 and EBS2), also base-pair with the 3′ end of the upstream exon to define the 5′ splice site.

Group-II introns are of particular interest because their catalytic core appears to have given rise to the RNA components (snRNAs) of the spliceosomal splicing machinery of the eukaryotic nucleus (Sharp 1991). This hypothesis, that spliceosomal introns are fragmented group-II introns, is supported not only by the observations that these two classes of introns share similar boundary sequences and are excised from primary transcripts as branched molecules, but also by the resemblance between intermolecular base-pairing interactions of snRNAs with each other and with spliceosomal intron/exon sequences and intramolecular base-pairing in group-II introns (Bonen 1993). Further evidence comes from the findings of degenerate group-II introns that lack typical cognates of domains I, II, III or IV in some organisms as well as from the ability of group-II intron domains to function in *trans* both in vivo and in vitro (Bonen 1993; Saldanha et al. 1993). Thus far, *trans*-spliced group-II introns (a total of nine) have been identified in the chloroplast *rps12* and *psa*A genes of land plants (Fukusawa et al. 1986; Koller et al. 1987; Zaita et al. 1987; Kohchi et al. 1988a) and *Chlamydomonas rein-*

M. Turmel (✉) · C. Otis · C. Lemieux
Département de biochimie, Faculté des sciences et de génie, Université Laval, Québec G1K 7P4, Canada

Y. Choquet[1] · M. Goldschmidt-Clermont · J.-D. Rochaix
Departments of Molecular Biology and Plant Biology, University of Geneva, CH-1211 Geneva 4, Switzerland

[1] *Present address:* Institut de Biologie Physico-Chimique, Paris, France

EMBL/GenBank/DDBJ accession nos. L27497 and L27257–L27262 (inclusive)

Communicated by R.W. Lee

hardtii (Kück et al. 1987; Choquet et al. 1988; Goldschmidt-Clermont et al. 1991), respectively, and in the mitochondrial *nad1*, *nad2* and *nad5* genes of land plants (Chapdelaine and Bonen 1991; Knoop et al. 1991; Wissinger et al. 1991; Binder et al. 1992). These genes consist of widely spaced exons flanked by 5′- or 3′-segments of group-II introns with discontinuities in either domains I, III or IV. The exons at different loci are transcribed into separate precursor RNAs, which are *trans*-spliced presumably via base-pairing and other interactions between segments of the group-II intron(s) in association with *trans*-splicing factors.

Genetic analysis of the *C. reinhardtii psa*A gene has revealed that several nuclear-encoded factors are necessary for the two *trans*-splicing reactions leading to the mature transcript and that additional factors are specific to one or the other of these reactions (Choquet et al. 1988). Interestingly, the assembly of the *Chlamydomonas psa*A exons 1 and 2 specifically requires one chloroplast gene, *tsc*A, in addition to a minimum of seven nuclear genes (Goldschmidt-Clermont et al. 1990). Goldschmidt-Clermont et al. (1991) have recently reported that this chloroplast locus encodes a 424 ± 19-nucleotide (nt) RNA molecule that may link the two segments of the *psa*A intron 1, via base-pairing, to reconstitute a composite intron structure with several features typical of group-II introns. In the secondary structure model proposed by these authors, the 5′ part of the *tsc*A RNA interacts with intron residues flanking the *psa*A exon 1 to reconstitute the helix at the base of domain I; the internal regions encode domains II and III and a significant portion of the central wheel; while bases in the 3′ part interact with intron residues to form the helical domain IV. It should be noted that the complete structure of domain I was not reported because it could not be unambiguously determined (Goldschmidt-Clermont et al. 1991).

To test the validity of the secondary structure model reported for the composite *C. reinhardtii psa*A intron 1 (Goldschmidt-Clermont et al. 1991), and also to complete this model by including domain I, we undertook the sequencing and modelling of equivalent introns in the *psa*A gene from other unicellular green algae belonging to the highly diversified genus *Chlamydomonas*. Our comparative analysis supports the previously proposed model for the *trans*-spliced *C. reinhardtii psa*A intron 1 (Goldschmidt-Clermont et al. 1991), and also indicates that the 5′ end of the *tsc*A RNA and the region downstream from the *psa*A exon 1 cannot be folded into a structure typical of domain I as described for most group-II introns. Although our results are consistent with the hypothesis that a fourth piece of the *psa*A intron 1 yet to be discovered provides the parts of domain I which are apparently missing, it is possible that the domain I of this intron has an unconventional structure.

Materials and methods

Isolation of nucleic acids. Total cellular RNA and chloroplast DNA (cpDNA)-enriched fractions were prepared from various *Chlamydomonas* taxa as described by Turmel et al. (1993a). The taxa analyzed consist of wild-type *C. reinhardtii* Dangeard mt^+ (SAG 11-32b), *C. zebra* Korshikov (SAG 10.83), *C. gelatinosa* Korshikov (SAG 69.72), *C. iyengarii* Mitra (SAG 25.72), *C. starrii* Ettl (SAG 3.73), *C. komma* Skuja (SAG 26.72), *C. mexicana* Lewin mt^- (SAG 11-60a), *C. peterfii* Gerloff (SAG 70.72), *C. frankii* Pascher mt^- (SAG 19.72), *C. pallidostigmatica* King (SAG 9.83), *C. eugametos* Moewus mt^+ (UTEX 9), *C. pitschmannii* Ettl (SAG 14.73), *C. species* 66.72 (SAG 66.72), *C. geitleri* Ettl (SAG 6.73) and *C. humicola* Lucksch (SAG 11-9). In addition, the *C. reinhardtii* chloroplast mutant strain H13 which is deficient in *psa*A mRNA maturation (Goldschmidt-Clermont et al. 1990) was employed.

Southern-blot hybridizations. Digests of cpDNA-enriched preparations were electrophoresed on 0.8% agarose gels and transferred to Hybond-N^{TM} nylon membranes (Amersham, Arlington Heights, Ill.) as recommended by the manufacturer. The membranes were hybridized under the conditions outlined by Woessner et al. (1986) with a 728-bp DNA fragment containing the *C. reinhardtii tsc*A gene [bases 347 to 1 066 of the sequence reported by Goldschmidt-Clermont et al. (1991)] and also with a 276-bp DNA fragment containing the last 33 bp of the *C. eugametos psa*A intron 1, the first 176 bp of the flanking *psa*A exon 2 from this alga, and 67 bp of pBluescript KS- vector sequence. The fragment containing *tsc*A was excised from a recombinant pBluescript KS-plasmid carrying a polymerase chain reaction (PCR)-amplified fragment (cloned into the *Sma*I site) using *Xba*I and *Bam*HI, two enzymes that cleave the sequences of the primers used for the PCR amplification. The fragment carrying the *psa*A exon 2 was PCR-amplified from a recombinant pBluescript KS- plasmid containing a 2 285-bp *Sau*3AI fragment with the primers 5′-TTGTAATACGACTCACTATAG-3′ (no. 85) and 5′-CCAGGTCATTTTTCACGTA-3′ (no. 256), which are complementary to the promoter T7 and an internal region of *psa*A exon 2, respectively. The two fragments were gel-purified and labelled with [α-^{32}P]dCTP (3 000 Ci/mmol) using the Multiprime DNA labelling system (Amersham, Arlington Heights, Ill.). The membranes were also hybridized under the conditions described by Boudreau et al. (1994) with ^{32}P-labelled oligonucleotides that are complementary to highly conserved regions of *tsc*A, the *psa*A exon 1, and *trn*I (CAU). The sequences of these oligonucleotides are as follows: for *tsc*A, 5′-GTAAATTCCAACTTCAGCTAG-3′ (no. 272) and 5′-AAACAATTAT(C/A)TGAGTACTAT-3′ (no. 273); for the *psa*A exon 1, 5′-ATGACAATTAGTACTCCAG-3′ (no. 275); and for *trn*I (CAU), 5′-GATTATGAGTCGTTTGCCT-3′ (no. 295). Oligos were 5′-end-labelled using [γ-^{32}P]ATP (3 000 Ci/mmol) and T4 polynucleotide kinase (Sambrook et al. 1989).

Cloning and sequence analysis of DNA. The sequence of the downstream region of exon 1 from *C. reinhardtii* (Goldschmidt-Clermont et al. 1991) was determined 500 bp further downstream using the Maxam-Gilbert sequencing method (Maxam and Gilbert 1980). The GenBank accession number of this sequence is L27497. Plasmid libraries of cpDNA-enriched preparations were constructed using *Hin*dIII and pBluescript KS- (Stratagene, La Jolla, Calif.) as described previously (Turmel et al. 1993a). Clones containing the sequences of interest were recovered from these libraries by colony hybridizations (Sambrook et al. 1989) using the aforementioned *tsc*A fragment probe and the oligonucleotides 5′-GCCCATTTTTCAAAACTTG-3′ (no. 190) and 5′-CCAGGTCATTTTTCACGTA-3′ (no. 256) which are complementary to the 3′ end of the *C. reinhardtii psa*A exon 1 and the 5′ end of the *psa*A exon 2 from this green alga. For sequencing of *tsc*A, subclones containing the *C. gelatinosa* and *C. zebra Hae*III restriction fragments of 700 and 610 bp, respectively, were used as templates. For sequencing of parts of the *psa*A exons 1 and 2, as well as of the intron-1 regions immediately flanking these exons in the two *Chlamydomonas* taxa, the *Hin*dIII clones retrieved by colony hybridizations were used as templates. Doubled-stranded DNA templates were sequenced as described previously (Turmel et al. 1993b). The sequences of the loci carrying *tsc*A, the *psa*A exon 1 and exon 2 are reported in this order in the following GenBank accession numbers: L27257–L27259 for *C. gelatinosa* and L27260–L27262 for *C. zebra*. Sequence analysis was carried out with the University of Wisconsin GCG software package

(Devereux et al. 1984), with the exception of the sequence alignments shown in Fig. 2 which were performed using CLUSTAL V (Higgins and Sharp 1988).

PCR amplifications. All PCR amplifications were carried out as described previously (Turmel et al. 1993b), except that the conditions for the 30 cycles of amplification were as follows: for cycles 1–9: 1 min at 94°C, 2 min at 50°C and 3 min at 72°C; for the remaining cycles: 1 min at 94°C, 2 min at 50°C and 3 min plus 10 additional s at 72°C. The following pairs of primers were used: (1) oligonucleotides nos. 272 and 273, and (2) oligonucleotides nos. 275 and 295 (see the section "Southern-blot hybridizations").

RNA analyses. Northern-blot hybridizations were carried out as outlined previously (Turmel et al. 1988). The *tscA* probe consisted of a mixture of restriction fragments from *C. reinhardtii*, *C. gelatinosa* and *C. zebra*. The *C. reinhardtii* fragment was the same as that used for Southern-blot hybridizations, while the *C. gelatinosa* and *C. zebra* fragments were obtained by cleaving the aforementioned recombinant pBluescript plasmids containing *Hae*III restriction fragments with *Bam*H1 and *Xba*I, two endonucleases that recognize sequences in the polylinker of the vector. Procedures for S1 nuclease mapping and primer extension were described by Erickson et al. (1984) and Ausubel et al. (1990), respectively.

Results

Screening of various *Chlamydomonas taxa* for the presence of a composite *psaA* intron comprising a chloroplast RNA component with sequence homology to the C. *reinhardtii tscA*

The two taxa carrying the composite *psa*A introns characterized in this study (*C. gelatinosa* and *C. zebra*) were selected following Southern-blot analysis of cpDNA-enriched preparations from 15 taxa that represent the various lineages distinguished in this green algal genus by comparative sequence analysis of chloroplast and nuclear rRNA genes (Buchheim et al. 1990; Turmel et al. 1993a) (see Fig. 1A). *Hin*dIII, *Ava*I and *Sty*I digests of all 15 green algal cpDNAs were hybridized with an oligonucleotide (19-mer) and a DNA fragment (728 bp) that were obtained from the *C. reinhardtii* cpDNA sequences containing the *psa*A exon 1 and *tsc*A, respectively, as well as with a 276-bp DNA fragment that was derived from the *C. eugametos* locus encoding the *psa*A exon 2. Note that *C. reinhardtii* and *C. eugametos* are very divergent taxa representing the two major lineages observed in *Chlamydomonas* (Buchheim et al. 1990; Turmel et al. 1993a) (see Fig. 1A). As the *C. eugametos psa*A exon sequences have not been reported yet, it is also worth mentioning that the three *psa*A exons of this green alga are colinear with their *C. reinhardtii* counterparts (M. T., C. O. and C. L., unpublished results) and, like these, reside at distinct sites on the chloroplast genome (Boudreau et al. 1994). Figures 1B–D shows the hybridization patterns of the three gene probes to the *Hin*dIII DNA digests of the 15 *Chlamydomonas* taxa. It can be seen that all taxa revealed positive hybridization signals with the 276-bp fragment specific to the *psa*A exon 2 (Fig. 1D) and that most of them also gave positive signals with the oligonucleotide specific to the *psa*A exon 1 (Fig. 1C). The latter oligonucleotide failed to hybridize to the *C. mexicana, C. peterfii and C. frankii* cpDNAs probably because its sequence is too divergent from the target sequences on these cpDNAs. Consistent with this idea, the *C. reinhardtii* oligonucleotide probe shows two nt differences (at positions 13 and 15) with the corresponding *C. eugametos* cpDNA region. Besides *C. reinhardtii*, only three taxa closely related to this alga, *C. gelatinosa*, *C. zebra* and *C. komma,* showed positive hybridization signals with the *tsc*A probe (Fig. 1B). Given that this probe spans the entire length of the *C. reinhardtii* cpDNA region encoding the *tsc*A RNA (424 ± 19 bp) plus at least 269 pb of flanking sequences, the absence of signals in the 12 remaining *Chlamydomonas* taxa may be attributed to the absence, or low sequence conservation, of *tsc*A. None of these 12 taxa consistently exhibited the same-sized fragments in the hybridizations of each of the three blots with the two *psa*A probes (data not shown for the *Ava*I and *Sty*I DNA blots), suggesting that exons 1 and 2 are distantly spaced on all of these green algal cpDNAs. Like their *C. reinhardtii* counterparts, the *C. gelatinosa*, *C. zebra* and *C. komma psa*A exons 1 and 2 appear to be unlinked to *tsc*A, as evidenced by the absence of common fragments recognized by the *psa*A and *tsc*A probes. Consistent with these conclusions, gene and physical mapping of the *C. gelatinosa* cpDNA indicated that *psa*A consists of three exons residing at distinct sites, none of which is closely linked to *tsc*A (E. Boudreau and M.T., unpublished results).

Like their *C. reinhardtii* counterpart, the *C. zebra* and *C. gelatinosa tsc*A RNAs potentially interact with intron sequences flanking the *psa*A exons 1 and 2

Figure 2 presents an alignment of the *C. reinhardtii*, *C. gelatinosa* and *C. zebra* sequences for the *psa*A exon 1 and the 5′ part of intron 1 (panel A), for *tsc*A (panel B), and for the 3′ part of the *psa*A intron 1 as well as the flanking exon 2 region (panel C). This alignment was derived on the basis of both primary sequence and secondary structure. As expected, the 5′ coding region of *psa*A is split at exactly the same position in all three green algae, i.e., 86 bp following the ATG initiation codon. Exons 1 and 2 are highly conserved in sequence, with such conservation extending to the region upstream of exon 1. Unlike these exons, the *tsc*A genes and the intron-1 regions show a number of deletions/additions and segments of substantial sequence divergence.

The consensus secondary structure model for the composite *C. reinhardtii, C. gelatinosa* and *C. zebra psa*A introns 1 is shown in Fig. 3. As expected, it displays several features of group-II introns, the most notable being the central wheel from which protude six helical domains. As reported also for the secondary structure proposed for the *C. reinhardtii psa*A intron (Goldschmidt-Clermont et al. 1991), the consensus model features many evolutionarily conserved residues typical of introns in subgroup-IIB1. Internal regions of the *tsc*A RNA encode domains II and III, while bases in the 5′ and 3′ terminal regions interact with intron sequences near the *psa*A exons 1 and 2 to re-

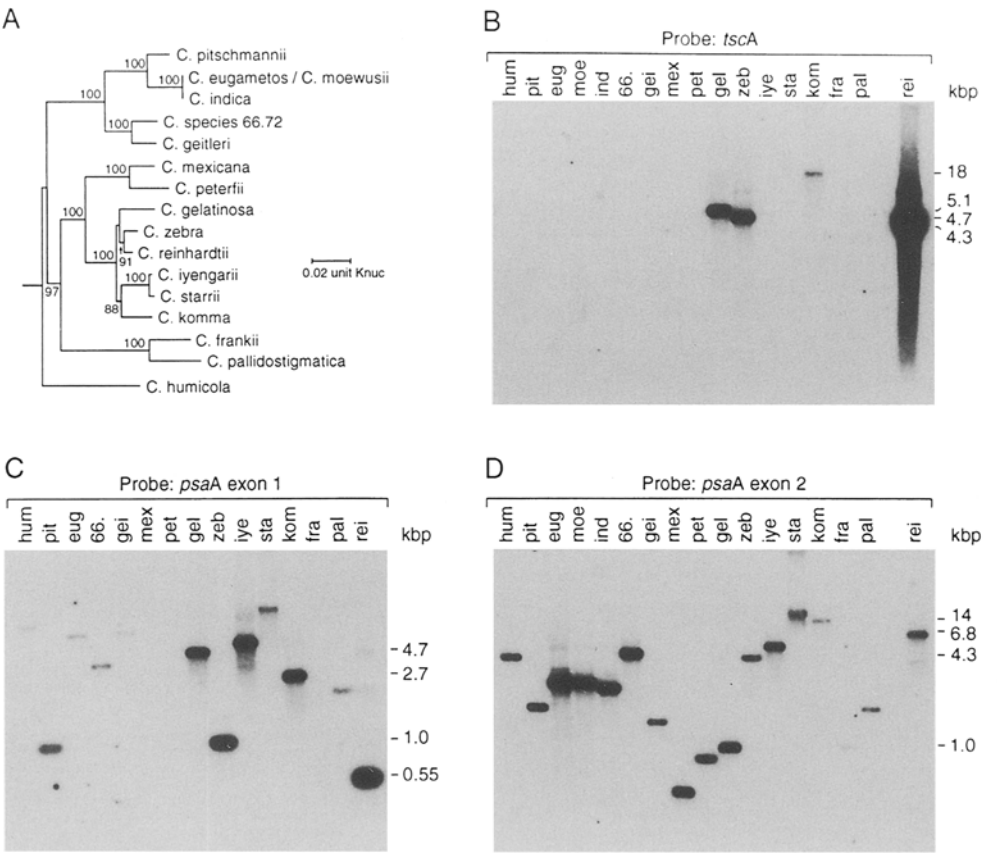

Fig. 1 Southern-blot hybridizations of *Hin*dIII-digested cpDNAs from 15 *Chlamydomonas* taxa whose phylogenetic relationships are known (**A**) with probes specific to *tsc*A (**B**), the *psa*A exon 1 (**C**) and the *psa*A exon 2 (**D**). The phylogenetic relationships among the 15 taxa were recently determined by comparative sequence analysis of the chloroplast large subunit rRNA gene (Turmel et al. 1993a). The three probes were hybridized under moderate-stringency conditions. Note that the Southern blot that was hybridized with the *tsc*A and the *psa*A exon-2 probes is different from that used for the hybridization with the *psa*A exon-1 probe and that, in *C. reinhardtii*, the *tsc*A probe recognized two fragments of 4.3 and 5.1 kbp. Full names of taxa are abbreviated by listing the first three characters of the species name. Sizes of hybridizing fragments are indicated for the four taxa that revealed positive signals with the *tsc*A probe (gel, zeb, kom and rei)

constitute part of the helical domain I and domain IV, respectively. Of the five domains completely represented in the consensus secondary structure model, domains IV and V exhibit the highest degree of sequence conservation. Domain II contains the smallest number of conserved residues, but the largest number of proven base-pairings. The most important deletion/addition differences (62–216 nt) between the three algal *psa*A introns map to the regions of domains II, III and VI which prolong the helices represented in Fig. 3. The largest domains II, III and VI among these introns are found in *C. reinhardtii* (139 nt), *C. zebra* (252 nt) and *C. gelatinosa* (260 nt) respectively, while the smallest domain II (38 nt) lies in *C. gelatinosa* and the smallest domains III (180 nt) and VI (44 nt) reside in *C. reinhardtii* (see Fig. 2). Despite these substantial variations, only small differences were observed in the size of the mature *tsc*A RNAs of the three *Chlamydomonas* taxa (424 ± 19 nt in *C. reinhardtii*, 382 ± 3 nt in *C. gelatinosa*, 462 ± 3 nt in *C. zebra*).

Unlike the conserved bases found in internal regions of the *C. reinhardtii, C. gelatinosa* and *C. zebra tsc*A RNAs, those in the regions most proximal to the 5′ and 3′ termini of these RNAs are not extensively involved in base-pairing interactions in the secondary structure model shown in Fig. 3. Indeed most (25) of the 28 strictly conserved nt in the 5′ terminal region and all ten absolutely conserved nt in the 3′ terminal region cannot base-pair with intron sequences flanking the exons. Using oligonucleotide probes complementary to these two highly conserved regions (see Fig. 2), we have screened the 15 *Chlamydomonas* cpDNA-enriched preparations we initially surveyed for the presence of this gene and found that, besides the four taxa previously identified, *C. starrii* and *C. iyengarii* disclosed hybridization signals (Fig. 4). All of these six taxa, with the exception of *C. iyengarii*, gave a hybridization signal with each of the two probes. The *C. iyengarii* cpDNA preparation failed to hybridize with the oligonucleotide specific to the 5′ region. In PCR reactions using the same oligonucleotides as primers, however, all six taxa yielded products each consisting of a single band in agarose gels (Fig. 5). The size of the PCR-amplified fragment derived from *C. iyengarii* (0.33 kbp) is similar to those derived from *C. reinhardtii* (0.38 kbp), *C. gelatinosa* (0.33 kbp) and *C. zebra* (0.42 kbp), whereas the *C. starrii* (0.68 kbp) and *C. komma* (1.4 kbp) fragments are significantly larger.

Northern-blot hybridization of total cellular RNA from the 15 aforementioned *Chlamydomonas* taxa with fragment probes specific to the *C. reinhardtii, C. gelatinosa* and *C. zebra tsc*A RNAs revealed the presence of a single transcript in *C. zebra* and the wild-type strain of *C. reinhardtii,* and of three transcripts in *C. gelatinosa* (Fig. 6). The sizes of the *C. reinhardtii* and *C. zebra tsc*A transcripts and of the smallest *C. gelatinosa* transcript are consistent with the values deduced from the positions of the 5′ and 3′ termini of *tsc*A RNAs as determined by primer extension and S1 nuclease protection. It is possible that the smallest

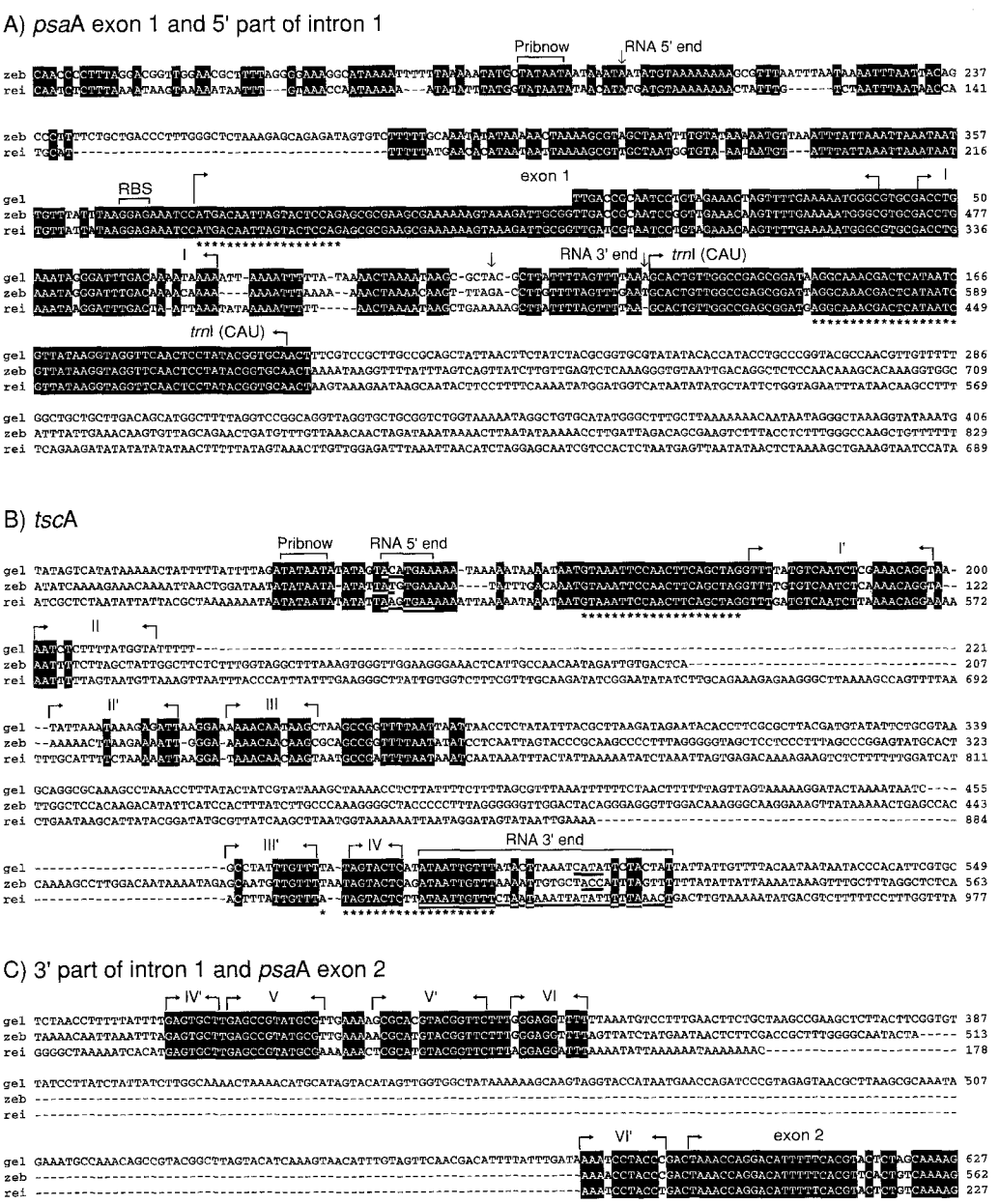

Fig. 2 Alignments of the *C. gelatinosa, C. zebra* and *C. reinhardtii* cpDNA sequences spanning the *psa*A exon 1 and the adjacent 5′ part of intron 1 (**A**), *tsc*A (**B**), and also the region containing the 3′ part of the *psa*A intron 1 and the 5′ part of exon 2 (**C**). Sequences were aligned using CLUSTAL V and minor modifications were made by eye to minimize insertion/deletion events. *Black boxes* mark sequence identities, while *hyphens* represent gaps inserted to maximize the alignment. Note that identities are not indicated for the *C. zebra* and *C. reinhardtii tsc*A sequences between the regions labelled II and II′ because these sequences could not be aligned unambiguously and that positions showing identical residues in the remaining regions were highlighted only if they were found in a window of ten positions where at least four positions share common residues. *Roman numerals* denote the six major domains of the composite *psa*A intron 1 (see Fig. 3). A *Roman numeral without a'* designates the 5′ branch of a domain, while a *Roman numeral accompanied with a'* designates the 3′ branch. The boundaries of all branches coincide with those of the corresponding helices in the secondary structure model shown in Fig. 3. Upstream of the region containing the exon 1, the positions of a putative ribosome binding site (RBS) and Pribnow box are indicated. The 5′ and 3′ termini of a precursor RNA of this exon in a *C. reinhardtii* mutant lacking the mature *psa*A mRNA are denoted by *small arrows*. The position of the 5′ terminus was taken from Choquet et al. (1988), while the 3′ terminus was mapped during the present study (see Fig. 8). The 5′ and 3′ termini of the mature *tsc*A RNA from each alga are *underlined*. The *C. gelatinosa* and *C. zebra tsc*A termini were mapped during this study, while the positions of the *C. reinhardtii tsc*A termini were taken from Goldschmidt-Clermont et al. (1991). *Asterisks* below the alignments denote the positions of the oligonucleotides used as probes or primers. For *C. gelatinosa* and *C. zebra*, all DNA sequences were taken from this study; for *C. reinhardtii*, they were taken from Kück et al. (1987), Goldschmidt-Clermont et al. (1991), and this study. Full names of taxa are abbreviated as in Fig. 1

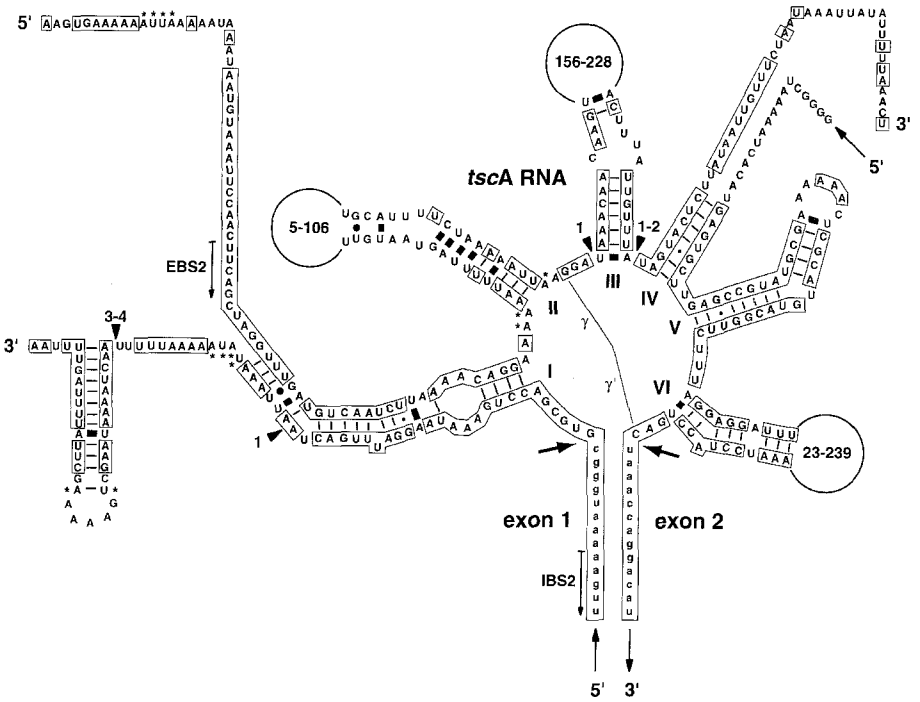

Fig. 3 Secondary structure model of the composite *C. reinhardtii psa*A intron 1 based on comparative analysis of DNA sequences from this green alga and the closely related *C. gelatinosa* and *C. zebra*. An alignment of these green algal sequences is presented in Fig. 2. *Roman numerals* designate the six major secondary-structure domains of group-II introns. Positions showing identical residues among the three *Chlamydomonas psa*A introns are boxed. Positions that are missing in *C. gelatinosa* and/or *C. zebra* are denoted by *asterisks*, whereas those that represent additions in the latter two algae relative to *C. reinhardtii* are denoted by *solid triangles*, with the accompanying *numbers* indicating the size of the insertions in nt. Base-pairings involving conserved residues are denoted by *thin bars or small dots*. Base-pairings proven by comparative sequence analysis are denoted by *thick bars or large dots*, whereas those that are disproven are not indicated. Note that several residues in the highly variable regions denoted by *circles* can participate in Watson-Crick base-pairing interactions. *Numbers in the circles* are the observed minimum and maximum lengths of these regions. *Blocked arrows* denote a potential EBS2-IBS2 pairing. Another long-range tertiary interaction, the γ-γ' pairing, is indicated. *Thick arrows* point to intron-exon junctions. *Thin arrows* indicate the 5′ and 3′ strand polarity of the separate transcripts containing exon 1 and exon 2

transcript of *C. gelatinosa* results from processing of the larger transcripts (1.1 and 0.78 kb).

The sequences downstream from the *C. reinhardtii*, *C. zebra* and *C. gelatinosa psa*A exons 1 cannot fold into the complex structure described for domain I of group-II introns

Most of the structural components (subdomains A, B, C and D) that are usually present in domain I of group-II introns cannot be identified in the consensus secondary structure model of the *C. reinhardtii*, *C. zebra* and *C. gelatinosa psa*A introns 1 (see Fig. 3). Abbreviated versions of domain I have been described only for the second intron in the chloroplast *clp*P gene of *Marchantia polymorpha*

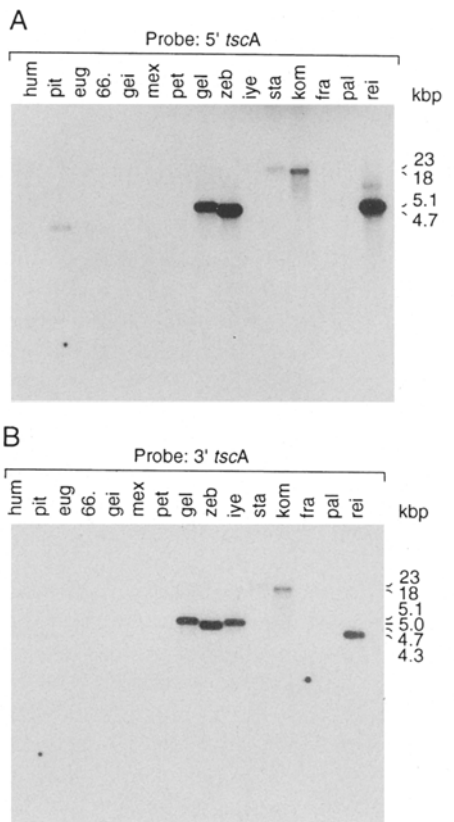

Fig. 4 Screening of 15 *Chlamydomonas* taxa for the presence of *tsc*A by Southern-blot hybridizations with two oligonucleotides that are complementary to highly converved sequences mapping at the 5′ (**A**) and 3′ (**B**) ends of this gene. The positions of these oligonucleotides are indicated in Fig. 2. Sizes of the hybridizing fragments are indicated. Full names of taxa are abbreviated as in Fig. 1. Sequence analysis of the *C. pitschmannii* fragment that revealed a positive signal with the 5′ *tsc*A probe disclosed no significant homology with the *C. reinhardtii, C. gelatinosa* and *C. zebra tsc*A genes

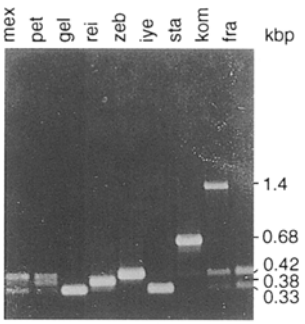

Fig. 5 PCR amplification of an internal region of *tsc*A from 15 *Chlamydomonas* taxa. The oligonucleotides used as primers in this experiment were employed as probes in the Southern-blot hybridizations presented in Fig. 4

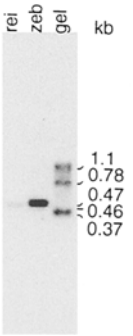

Fig. 6 Analysis of total cellular RNA from *C. reinhardtii*, *C. gelatinosa* and *C. zebra* by Northern-blot hybridization with a *tsc*A probe. The probe consisted of a mixture of restriction fragments from these three green algae (for more details see the Materials and methods section). Full names of taxa are abbreviated as in Fig. 1

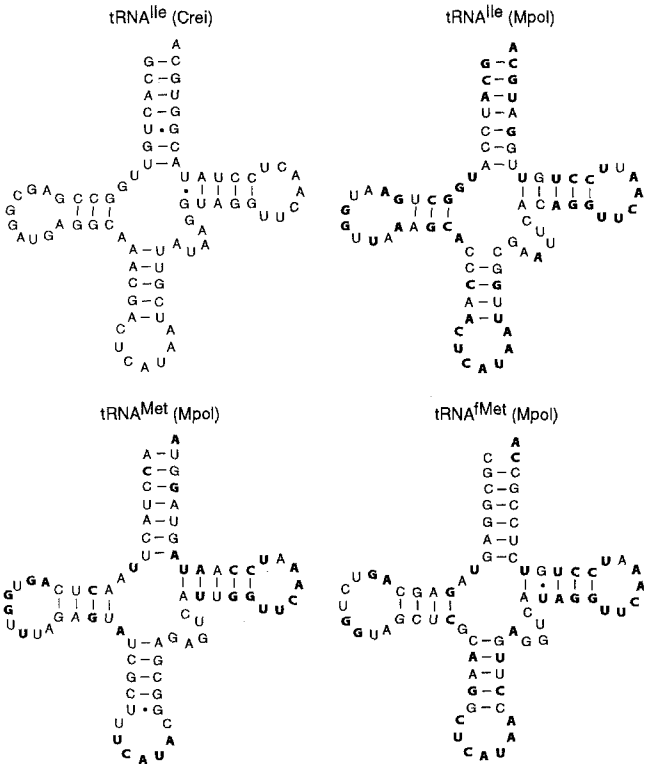

Fig. 7 Comparison of the *C. reinhardtii* chloroplast (Crei) tRNA^{Ile} (CAU) with the *M. polymorpha* (*Mpol*) chloroplast tRNA^{Ile} (CAU), initiator tRNA^{fMet} (CAU) and elongator tRNA^{Met} (CAU). Conserved bases (*boldface*) are indicated. The tRNA sequences of *Marchantia* were taken from Ohyama et al. (1986), while that of *C. reinhardtii* was taken from this study

(Kohchi et al. 1988b) and for group-II introns in the *Euglena gracilis* cpDNA (Hallick et al. 1993). To our surprise, 82 to 86 bases downstream from the *psa*A exon 1 in the three *Chlamydomonas* taxa examined, we have found a highly conserved sequence that can be folded into a tRNA with the anticodon CAU (Fig. 2). This tRNA was identified as the tRNA^{Ile} (CAU) by sequence comparison with the three *M. polymorpha* tRNAs displaying the same anticodon (Fig. 7). The *C. reinhardtii* choroplast tRNA^{Ile} (CAU) is 63%, 48% and 42% identical in sequence with the *Marchantia* chloroplast tRNA^{Ile} (CAU), initiator tRNA^{fMet} (CAU) and elongator tRNA^{Met} (CAU), respectively. As the codon and amino-acid specificities of the *Escherichia coli* tRNA^{Ile} (CAU) are determined by a single post-transcriptional modification (C → lysidine) at the first position of the anticodon (Muramatsu et al. 1988), this highly conserved position in the corresponding tRNA from chloroplasts is likely to be similarly modified to allow specific recognition of adenosine in the third position of the codon. Note that the loci encoding the initiator tRNA^{fMet} (CAU) and elongator tRNA^{Met} (CAU) have been recently mapped on the *C. reinhardtii* cpDNA and found to be unlinked to the *trn*I (CAU) gene (Boudreau et al. 1994). Five conserved bases (GUUCA, four of which are unpaired) in the TΨC arm of the tRNA^{Ile} (CAU) are complementary to the last five bases of the *psa*A exon 1 and thus could constitute a potential EBS1 site if the *psa*A exon 1 and the *trn*I gene are cotranscribed as a precursor. An exon 1 transcript could be detected in a *C. reinhardtii* class-C mutant deficient in *psa*A exon1-exon 2 splicing. The 5′ end of this transcript (approximately 300 nt in size) lies 130 ± 3 nt upstream of the initiation codon (Choquet et al. 1988), whereas the 3′ terminus lies 84 ± 5 nt downstream from the exon-intron junction near the 5′ end of the *trn*I (CAU) gene as shown by the S1 analysis in Fig. 8. A fainter signal is also observed 20 nt upstream corresponding to an A-rich terminal loop which may be an artefact. A potential EBS2 site is located near the 5′ end of the *tsc*A RNA that is complementary to the intron-binding site 2 (IBS2) within exon 1 (Fig. 3). No other potential EBS1 and EBS2 sites are conserved downstream from the *trn*I gene among the three algae.

To determine if the linkage of the *psa*A exon 1 to *trn*I (CAU) is prevalent in *Chlamydomonas*, we hybridized Southern blots of *Hin*dIII digests from the 15 *Chlamydomonas* cpDNAs analyzed previously with an oligonucleotide specific to this tRNA gene (Fig. 9B). Moreover, we PCR-amplified the region between the *psa*A exon 1 and *trn*I (CAU) in these taxa (Fig. 9A). Hybridization signals were observed for all of the taxa, with the exception of *C. humicola*; however, only the closely related *C. zebra, C. gelatinosa, C. komma, C. starrii* and *C. iyengarii* revealed signals to the same *Hin*dIII fragments that hybridized with the *psa*A exon 1 probe (see Fig. 1). Note that, in *C. reinhardtii*, the presence of a *Hin*dIII site in the region between the *psa*A exon 1 and *trn*I (CAU) did not allow us to confirm the close linkage of these genes. Our PCR amplifications from the *C. reinhardtii, C. zebra, C. gelatinosa, C.*

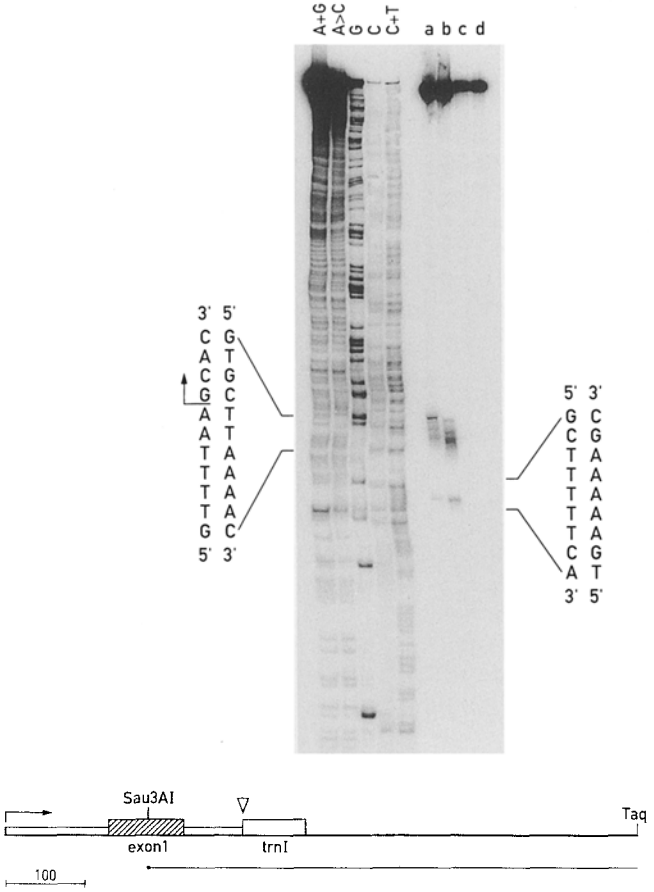

Fig. 8 S1 nuclease analysis of the 3′ end of the *psa*A exon-1 transcript of *C. reinhardtii*. Upper: a 650-bp *Sau*3AI-*Taq*I fragment 3′ end-labelled at its *Sau*3AI site (located within the *psa*A exon 1) was annealed with 200 µg of total cellular RNA of mutant H13, treated with 5 U (*lane a*) or 15 U (*lane b*) of S1 nuclease and electrophoresed on a sequencing gel; *c*, same as *b* except that yeast tRNA was used; *d*, probe without S1 nuclease treatment. Sequencing ladders are shown on the left with a display of the nt sequences around the two S1 sites. The *broken arrow* indicates the 5′ end of the *trn*I (CAU) gene. Lower: map of the *psa*A exon1-*trn*I (CAU) region. The 3′ end-labelled probe is shown. The *arrow* indicates the direction of transcription and the *wedge* marks the principal S1 site

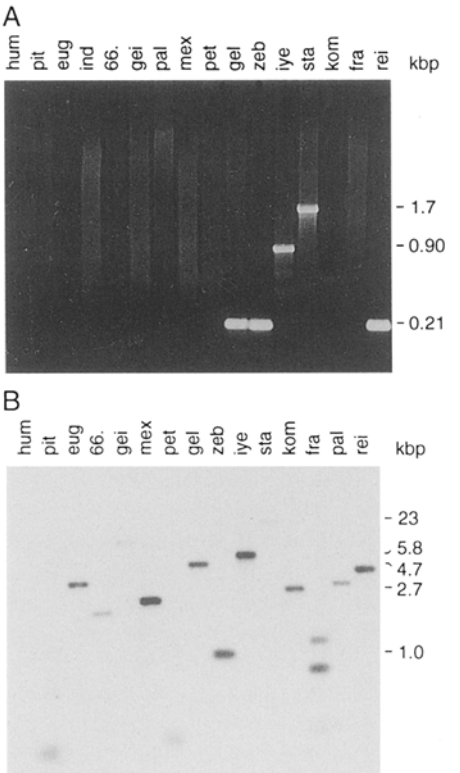

Fig. 9 Degree of linkage between the *psa*A exon 1 and *trn*I (CAU) as established by PCR amplification (**A**) and Southern-blot hybridization (**B**) in 15 *Chlamydomonas* taxa. The positions of the PCR primers are indicated in Fig. 2. The Southern-blot of *Hin*dIII digests which was employed for the hybridization with the *psa*A exon 1 probe (see Fig. 1) was hybridized with the *trn*I-specific oligonucleotide used as a primer for the PCR amplifications. Full names of taxa are abbreviated as in Fig. 1. Sizes of the hybridizing fragments that were recognized by both the *trn*I and *psa*A exon-1 probes (i.e., those of *gel*, *zeb*, *iye*, *sta* and *kom*) are indicated

starrii and *C. iyengarii* cpDNA preparations yielded fragments with sizes ranging from 0.21 to 1.7 kbp. The *C. komma* preparation gave no PCR product, most probably because the distance between the target sites of the primers (estimated to be < 2.7 kbp based on Southern analysis) slightly exceeds the maximal size (approximately 2.5 kbp) of the region that is routinely amplified under the conditions employed.

Discussion

During this study, we have employed both Southern-blot analysis and PCR amplification to screen several *Chlamydomonas* taxa for the presence of a *trans*-spliced *psa*A intron that is similar in sequence and structure to the first intron in the *C. reinhardtii psa*A gene. Our finding that probes specific to the *psa*A exons 1 and 2 hybridized to distinct restriction fragments in all of the taxa that revealed hybridization signals with these probes is consistent with our recent mapping and sequencing data for the *C. eugametos psa*A exons which suggest that this gene was probably split in pieces prior to the emergence of *Chlamydomonas* (Boudreau et al. 1994; M. T., C. O. and C. L., unpublished results). As all of the chloroplast *psa*A genes characterized so far in other algae and in land plants do not feature any *trans*-spliced introns nor *cis*-spliced introns at the same locations as those found in *C. reinhardtii*, insertion and splitting of group-II introns probably occurred relatively recently in a common ancestor of *Chlamydomonas* taxa. Only the taxa most closely allied to *C. reinhardtii* revealed a chloroplast locus encoding an RNA component with sequence similarity to the *C. reinhardtii tsc*A RNA. Our failure to identify a sequence similar to the *C. reinhardtii tsc*A locus in two taxa belonging to the *C. reinhardtii* lineage (*C. mexicana* and *C. peterfii*), as well as in all those belonging to the *C. eugametos/C. moewusii* lineage, does not necessarily indicate that such an RNA species is

lacking from each of these green algal *psa*A introns as its primary sequence may be too divergent to be detected by hybridization or PCR amplification. Analysis of the intron sequences flanking the *C. eugametos* exons 1 and 2 has been undertaken to determine if these sequences can fold into a structure with all of the features of group-II introns.

Our detailed characterization of the *C. gelatinosa* and *C. zebra tsc*A loci as well as of the *psa*A exons 1 and exons 2 and their flanking intron segments has allowed us to identify the potential base-pairings and unpaired residues which have been preserved in the secondary structure model proposed for the *C. reinhardtii psa*A intron 1 (Goldschmidt-Clermont et al. 1991). This comparative analysis has confirmed the validity of this model postulating that internal regions of the *tsc*A RNA encode domains II and III and that terminal regions reconstitute the disrupted helices of domains I and IV via base-pair interactions with the 5′ and 3′ intron segments, respectively. As reported for other group-II introns (Michel et al. 1989), the most extensive variations in sequence and length among the *C. reinhardtii*, *C. gelatinosa* and *C. zebra psa*A introns 1 have been located at the tips of helical regions (domains II, III, IV and VI), whereas the most conserved regions have been found predominantly within the helical regions. Michel et al. (1989) have suggested that the peripheral regions can tolerate large insertions because they are presumably at the periphery of the intron molecule, whereas the helical regions cannot retain such insertions because they are buried within the molecule.

It is intriguing to find a significant number of strictly conserved bases that are unpaired, or poorly paired, immediately adjacent to the helices of domains I and IV within the terminal regions of the *C. reinhardtii*, *C. gelatinosa* and *C. zebra tsc*A RNAs (see Fig. 3). Using oligonucleotides specific to these regions, we have succeeded in PCR-amplifying the *C. starrii* and *C. iyengarii tsc*A loci, but failed to identify these DNA regions by hybridizing Southern blots of cpDNA digests with a fragment probe specific to the *C. reinhardtii tsc*A under low-stringency conditions. Given their high level of sequence conservation, these terminal regions of the *tsc*A RNA must have important functions. It is possible that they participate in the expression of *tsc*A via association with specific proteins or in the assembly of the mature *psa*A transcript via interaction with proteins, or else with a fourth RNA component (see below) or another region of the *psa*A intron. For instance, 5 of the 28 absolutely conserved bases at the 5′ terminus of *tsc*A can potentially base-pair with the AAGUU exon-1 sequence upstream of the 5′ splice site to form the long-range EBS2-IBS2 interaction (see Fig. 3) which has been proposed to stabilize the intron-5′ exon complex and contribute to positioning the 5′ splice site (see Michel et al. 1989; Saldanha et al. 1993). In contrast to the EBS1-IBS1 pairing which is required for accurate selection of the 5′ splice site, the EBS2-IBS2 interaction seems to be missing in several group-II introns (Michel et al. 1989). As specific mutations can be efficiently introduced into the *C. reinhardtii* cpDNA via transformation (Boynton et al. 1988, see Rochaix 1992), functional analysis of the very conserved regions at the termini of this green alga *tscA* RNA should be undertaken to gain an insight into their roles.

The composite, *trans*-spliced *psa*A introns of *C. reinhardtii, C. gelatinosa* and *C. zebra* are also very unusual in that the 5′ end of the *tsc*A RNA and the region upstream of the *psa*A exon 1 cannot fold into the conventional structure described for domain I of group-II introns. Modelling of these regions has revealed no recognizable cognates of subdomains A, B, C, and D. The putative EBS1 sequence in the $tRNA^{Ile}$ (CAU) encoded near the 3′ end of the *psa*A exon 1 is unlikely to participate in splicing of the *psa*A intron 1 for the following reasons. First, our S1 analysis has provided no evidence for co-transcription of the *psa*A exon 1 and *trn*I (CAU). This observation does not eliminate the possibility that co-transcription of these genes occurs and is followed by very rapid processing; however, interaction of the putative EBS1 sequence with the complementary region upstream of the *psa*A exon 1 would require that intron splicing precedes processing. Second, the highly variable distance (82 bp to >2.5 kbp) separating the *psa*A exon 1 and *trn*I (CAU) in the six *Chlamydomonas* cpDNAs in which we have identified a composite *psa*A intron is likely to cause a problem in bringing the putative EBS1-IBS1 helix into position within the catalytic site. Third, because the putative EBS1 sequence in the $tRNA^{Ile}$ (CAU) plays an important role in tRNA structure/function and is found in virtually all tRNAs, an additional role in intron splicing seems very unlikely. In *C. eugametos*, the finding that the region downstream from the *psa*A exon 1 and the 5′ end of the putative *tsc*A RNA cannot fold into the typical structure described for domain I of group-II introns would also argue against the participation of the $tRNA^{Ile}$ (CAU) in intron splicing, as the chloroplast *trn*I (CAU) of this green alga lies about 30 kbp from the *psa*A exon 1 (Boudreau et al. 1994).

Because our comparative analysis has failed to resolve a potential secondary structure of domain I, it is possible that, in the *Chlamydomonas psa*A introns examined, this domain has extensively diverged from that of its counterpart in other organelle genes and that the missing RNA subdomains are not required. It seems more reasonable, however, to propose that a fourth RNA species encoded by a distinct chloroplast locus specifies the missing RNA subdomains. This hypothesis is particularly attractive in light of the idea that group-II intron domains might have evolved into snRNAs in the evolution of nuclear pre-mRNA introns (Sharp 1991). Even though genetic analyses of *C. reinhardtii* PSI mutants defective in *psa*A mRNA maturation have failed to provide hints for the requirement of more than one chloroplast locus (Choquet et al. 1988; Herrin and Schmidt 1988; Goldschmidt-Clermont et al. 1990; Roitgrund and Mets 1990), we cannot reject this possibility given that chloroplast mutations in a locus as small as that predicted for the potential *tsc*B gene might have easily escaped detection due to low frequency of such events. Consistent with the involvement of a second *trans*-acting factor for splicing of the *psa*A exons 1 and 2 would be the finding that the *C. eugametos psa*A intron 1 lacks a typical domain I. Alternatively, one or several proteins encoded

by chloroplast or nuclear genes could replace the function normally provided by the region of domain I which appears to be missing from the *C. reinhardtii, C. gelatinosa* and *C. zebra psa*A introns 1. Some of the seven genetically identified nuclear loci that have been reported to be specifically involved in *psa*A exon 1-exon 2 splicing (Goldschmidt-Clermont et al. 1990) could specifiy such proteins.

Acknowledgements We thank Drs. Linda Bonen and Murray N. Schnare for their helpful comments on the manuscript. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (GP0003293 to M. T. and GP0002830 to C.L.), "Le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche" (93-ER-0350), and the Swiss National Research Fund (31-26345.89). M. T. and C. L. are Scholars in the Evolutionary Biology Program of the Canadian Institute for Advanced Research.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1990) Current protocols in molecular biology. J Wiley and Sons, New York

Binder S, Marchfelder A, Brennicke A, Wissinger B (1992) RNA editing in *trans*-splicing intron sequences of *nad2* mRNAs in *Oenothera* mitochondria. J Biol Chem 276:7615–7623

Bonen L (1993) *Trans*-splicing of pre-mRNA in plants, animals, and protists. FASEB J 7:40–46

Boudreau E, Otis C, Turmel M (1994) Conserved gene clusters in the highly rearranged chloroplast genomes of *Chlamydomonas moewusii* and *Chlamydomonas reinhardtii*. Plant Mol Biol 24: 585–602

Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB, Sanford JC (1988) Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. Science 240:1534–1538

Buchheim MA, Turmel M, Zimmer EA, Chapman RL (1990) Phylogeny of *Chlamydomonas* (Chlorophyta) based on cladistic analysis of nuclear 18S rRNA sequence data. J Phycol 26:689–699

Chapdelaine Y, Bonen L (1991) The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a *trans*-splicing model for this gene-in-pieces. Cell 65:465–472

Choquet Y, Goldschmidt-Clermont M, Girard-Bascou J, Kück U, Bennoun P, Rochaix J-D (1988) Mutant phenotypes support a *trans*-splicing mechanism for the expression of the tripartite *psa*A gene in the *C. reinhardtii* chloroplast. Cell 52:903–913

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Erickson JM, Rahire M, 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

Ferat J-L, Michel F (1993) Group-II self-splicing introns in bacteria. Nature 364:358–361

Fukuzawa H, Kohchi T, Shirai H, Ohyama K, Umesono K, Inokuchi H, Ozeki H (1986) Coding sequences for chloroplast ribosomal protein S12 from liverwort, *Marchantia polymorpha*, are separated far apart on different DNA strands. FEBS Lett 198:11–15

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

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

Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E (1993) Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res 21:3537–3544

Herrin DL, Schmidt GW (1988) *Trans*-splicing of transcripts for the chloroplast psaA1 gene. J Biol Chem 263:14601–14604

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 79: 237–244

Jacquier A (1990) Self-splicing group-II and nuclear pre-mRNA introns: how similar are they? Trends Biochem Sci 15:351–354

Knoop V, Schuster W, Wissinger B, Brennicke A (1991) *Trans* splicing integrates an exon of 22 nucleotides into the *nad5* mRNA in higher plant mitochondria. EMBO J 10:3483–3493

Kohchi T, Umesono K, Ogura Y, Komine Y, Nakahigashi K, Komano T, Yamada Y, Ozeki H, Ohyama K (1988a) A nicked group-II intron and *trans*-splicing in liverwort, *Marchantia polymorpha*, chloroplasts. Nucleic Acids Res 16:10025–10036

Kohchi T, Ogura Y, Umesono K, Yamada Y, Komano T, Ozeki H, Ohyama K (1988b) Ordered processing and splicing in a polycistronic transcript in liverwort chloroplasts. Curr Genet 14: 147–154

Koller B, Fromm H, Galun E, Edelman M (1987) Evidence for in vivo *trans*-splicing of pre-mRNAs in tobacco chloroplasts. Cell 48:111–119

Kück U, Choquet Y, Schneider M, Dron M, Bennoun P (1987) Structural and transcription analysis of two homologous genes for the P700 chlorophyll *a*-apoproteins in *Chlamydomonas reinhardtii*: evidence for in vivo *trans*-splicing. EMBO J 8:2185–2195

Maxam AM, Gilbert W (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol 65:499–560

Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group-II catalytic introns – a review. Gene 82:5–30

Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T, Yokoyama S (1988) Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature 336:179–181

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha*. Nature 322:572–574

Rochaix J-D (1992) Post-transcriptional steps in the expression of chloroplast genes. Annu Rev Cell Biol 8:1–28

Roitgrund C, Mets LJ (1990) Localization of two novel chloroplast genome functions: *trans*-splicing of RNA and protochlorophyllide reduction. Curr Genet 17: 147–153

Saldanha R, Mohr G, Belfort M, Lambowitz A (1993) Group-I and group-II introns. FASEB J 7:15–24

Sambrook J, Fritch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sharp PA (1991) "Five easy pieces". Science 254:663

Turmel M, Lemieux B, Lemieux C (1988) The chloroplast genome of the green alga *Chlamydomonas moewusii:* localization of protein-coding genes and transcriptionally active regions. Mol Gen Genet 214:412–419

Turmel M, Gutell RR, Mercier J-P, Otis C, Lemieux C (1993a) Analysis of the chloroplast large subunit ribosomal RNA gene from 17 *Chlamydomonas* taxa: three internal transcribed spacers and 12 group-I intron insertion sites. J Mol Biol 232:446–467

Turmel M, Mercier J-P, Côté M-J (1993b) Group-I introns interrupt the chloroplast *psa*B and *psb*C and the mitochondrial *rrn*L gene in *Chlamydomonas*. Nucleic Acids Res 21:5242–5250

Wissinger B, Schuster W, Brennicke A (1991) *Trans* splicing in *Oenothera* mitochondria: *nad1* mRNAs are edited in exon and *trans*-splicing group-II intron sequences. Cell 65:473–482

Woessner JP, Gillham NW, Boynton JE (1986) The sequence of the chloroplast *atpB* gene and its flanking regions in *Chlamydomonas reinhardtii*. Gene 44:17–28

Zaita N, Torazawa K, Shinozaki K, Sugiura M (1987) *Trans*-splicing in vivo: joining of transcripts from the "divided" gene for ribosomal protein S12 in the chloroplasts of tobacco. FEBS Lett 210:153–156