Curr Genet (1999) 36: 345–353

ORIGINAL PAPER

Peik Haugen · Volker A.R. Huss · Henrik Nielsen
Steinar Johansen

Complex group-I introns in nuclear SSU rDNA of red and green algae: evidence of homing-endonuclease pseudogenes in the Bangiophyceae

Received: 26 May / 25 August 1999

Abstract The green alga *Scenedesmus pupukensis* and the red alga *Porphyra spiralis* contain large group-IC1 introns in their nuclear small subunit ribosomal RNA genes due to the presence of open reading frames at the 5′ end of the introns. The putative 555 amino-acid *Scenedesmus*-encoded protein harbors a sequence motif resembling the bacterial S9 ribosomal proteins. The *Porphyra* intron self-splices in vitro, and generates both ligated exons and a full-length intron RNA circle. The *Porphyra* intron has an unusual structural organization by encoding a potential 149 amino-acid homing-endonuclease-like protein on the complementary strand. A comparison between related group-I introns in the Bangiophyceae revealed homing-endonuclease-like pseudogenes due to frame-shifts and deletions in *Porphyra* and *Bangia*. The *Scenedesmus* and *Porphyra* introns provide new insights into the evolution and possible novel functions of nuclear group-I intron proteins.

Key words Group-I intron · Ribosomal DNA · Homing endonuclease · *Scenedesmus* · *Porphyra*

Peik Haugen · Steinar Johansen (✉)
Department of Molecular Cell Biology,
Institute of Medical Biology,
University of Tromsø, N-9037 Tromsø, Norway

Steinar Johansen
Norwegian Institute of Gene Ecology,
MH-building, N-9037, Tromsø, Norway

Volker A.R. Huss
Institut für Botanik und Pharmazeutische Biologie
der Universität, Erlangen, Germany

Henrik Nielsen
Department of Medical Biochemistry and Genetics,
Panum Institute, Blegdamsvej 3C,
DK-2200 Copenhagen N, Denmark

Communicated by H.T. Jacobs

Introduction

Some protists and fungi harbor group-I introns in their nuclear ribosomal RNA (rRNA) genes, which have to be spliced out from the precursor RNA in order to generate proper mature rRNA and functional ribosomes. Group-I intron splicing is a self-catalyzed reaction, initiated by a nucleophilic attack of an exogenous guanosine at the 5′ splice site. The splicing proceeds by two sequential trans-esterification reactions resulting in intron excision and exon ligation (reviewed in Cech and Herschlag 1996). The catalytic property of group-I intron RNA is highly dependent on a well-defined three-dimensional structure (Cate et al. 1996; Lehnert et al. 1996; Golden et al. 1998), consisting of ten paired segments (P1–P10) common to most group-I ribozymes and seven additional segments (P11–P17) present only in certain subgroups (Michel and Westhof 1990; Lehnert et al. 1996; Einvik et al. 1998).

Both reverse splicing at the RNA-level (Woodson and Cech 1989; Roman et al. 1999) and homing at the DNA-level (Muscarella and Vogt 1989; Johansen et al. 1997) may contribute to the mobility of nuclear group-I introns. Homing depends on the expression of intron-encoded homing endonucleases (Belfort and Roberts 1997). There are several classes of homing endonucleases known, including the His-Cys box class which is restricted to the nuclear group-I introns (Johansen et al. 1993). The His-Cys box is directly involved in zinc binding and in the active site of the endonuclease (Flick et al. 1998). The intron ORFs coding for the endonucleases exist as extension sequences within peripheral loop-regions of the group-I ribozyme structure (Table 1).

The P1 paired segment contains the 5′ splice site and the internal guide sequence which base pair to both the upstream and downstream exon sequences (Cech and Herschlag 1996). Some nuclear group-I introns contain additional sequences located as extensions beyond the P1 segment where they do not perturb the folding of the

Table 1 Nuclear group-I introns with ORFs

Species[a]	Position[b]	Location[c]	Size[d]	HC[e]	Acc[f]	Ref[g]
Chlorophyta (Pr):						
Scenedesmus pupukensis	S-40	P1	555	–	X91267	1
Rhodophyta (Pr):						
Porphyra spiralis	S-1506	P1	149	+	L26177	2
Porphyra tenera	S-516	P1	162	+	AB013176	3
Myxomycota (Pr):						
Didymium iridis	S-956	P2	244	+	X71792	4
Physarum polycephalum	L-1925	P1	163	+	L031813	5
Discomitochondria (Pr):						
Naegleria jamiesoni	S-516	P6	245	+	U80250	6
Naegleria sp. NG236	L-2563	P1	175	+	AJ001315	7
Ascomycota (F):						
Protomyces pachydermus	S-1506	P9	228	–	D85142	8

[a] Species are grouped into phyla of Protoctista (Pr) or Fungi (F) according to the classification of Margulis and Schwartz (1998)
[b] Insertion position of group-I introns corresponding to the *E. coli* small subunit (S) and large subunit (L) rRNA sequences (see Johansen et al. 1996) which immediately precedes the intron
[c] Paired segment (Pn) within the group-I ribozyme structure harbour ORF-containing peripheral sequences (see Fig. 1)
[d] Size of putative intron protein in amino acids
[e] His-Cys box motif present (+) or absent (–)
[f] EMBL/Genbank accession numbers
[g] References: (1) Kessler et al. 1997; this work; (2) Oliveira and Ragan 1994; this work; (3) Kunimoto et al. 1998; this work; (4) Johansen and Vogt 1994; Decatur et al. 1995; Vader et al. 1999; (5) Muscarella et al. 1990; (6) Einvik et al. 1997; Elde et al. 1999; (7) De Jonckheere and Brown 1998; (8) Nishida et al. 1998

ribozyme core. Such P1-extensions appear to have different molecular features: (1) direct repeat motifs have been described in the myxomycete *Physarum flavicomum* (Vader et al. 1994); (2) increase of the base-paired length has been noted in the ascomycete *Myriosclerotinia caricisampullacea* (Holst-Jensen et al. 1999); (3) intraspecific size polymorphism has been reported in the red alga *Porphyra spiralis* (Oliveira and Ragan 1994); (4) endonuclease-like ORFs have been found in rDNA introns of the protists *Physarum polycephalum*, *Naegleria* NG236, and *Porphyra tenera* (see Table 1). The well-characterized *P. polycephalum* ORF encodes the homing endonuclease I-*Ppo*I (Muscarella et al. 1990) and is directly involved in intron homing (Muscarella and Vogt 1989). I-*Ppo*I is a very rare example of a protein expressed from an RNA polymerase-I transcript (Lin and Vogt 1998).

Here we describe two unusual P1-extension sequences within nuclear rDNA group-I introns. The green algae *Scenedesmus pupukensis* contains a P1 extension with a potential ORF corresponding to a protein of 555 amino acids, which is approximately four-times larger than any other nuclear P1-ORF known. Some red algae of the class Bangiophyceae contain His-Cys box reading frames as P1-extensions located on the antisense strand compared with that encoding the group-I ribozyme and the ribosomal RNA. Comparative sequence analysis suggests that the Bangiophyceae group-I intron reading frames represent homing-endonuclease pseudogenes.

Materials and methods

Intron sequences

The *S. pupukensis* UTEX 2219 (= *Kermatia pupukensis*) SSU rDNA sequence (Kessler et al. 1997), including the intron inserted at position 40, was obtained from the Genbank/EMBL Database (X91267). The plasmid pCR1000-PSA-R was kindly provided by Dr. M.A. Ragan (Halifax, Canada) and contains the *P. spiralis* (var. *amplifolia* PSA-R) 1506-intron and some flanking exon sequences behind the T7 promoter (Oliveira and Ragan 1994). The *P. spiralis* intron ORF was re-sequenced on both strands using the Thermo Sequenase sequencing kit (Amersham Pharmacia Biotech) and [^{33}P] ddNTPs (GATC; 450 μCi/ml), and was found to be identical to the previously reported sequence (Genbank/EMBL Database accession number L26177) except for one guanosine insertion between intron positions 70 and 71. Computer analysis of nucleic-acid and amino-acid sequences was performed using the software package programs from the Genetics Computer Group (Madison, Wis.) and GeneCompare, Applied Maths Inc. (Kortrijk, Belgium).

In vitro transcription and splicing

Intron RNA transcription and splicing were performed essentially as described previously (Johansen and Vogt 1994). RNA was transcribed by T7 RNA polymerase off the *Bam*HI-linearized pCR1000-PSA-R using [^{35}S] CTP αS (10 μCi/μl; Amersham Pharmacia Biotech) for labeling, then purified and subjected to self-splicing conditions. RNA was incubated at 45 °C for 0–30 min in 40 mM Tris-HCl pH 7.5, 25 mM $MgCl_2$, 1 M KCl, 2 mM spermidine, 5 mM dithiothreitol, and 0.2 mM GTP. Splicing intermediates and products were analyzed on polyacrylamide gels as described by Decatur et al. (1995).

Circle junction determination

The circular RNA was eluted from polyacrylamide gels as previously described (Johansen and Vogt 1994) and subjected to reverse transcription using the First-Strand cDNA Synthesis kit (Amersham Pharmacia Biotech) and the primer OP-446 (Table 2). The transcripts were subsequently PCR-amplified by adding primer OP-54 (Table 2). PCR products were cloned into pUC18 using the Sure-Clone Ligation kit (Amersham Pharmacia Biotech) and sequenced in both directions using the Thermo Sequenase sequencing kit. Alternatively, the circle junction sequence was obtained by chemical sequencing of the PCR product. Here, the primer OP54 was 5′-endlabeled using [γ^{-32}P]ATP (10 μCi/μl; Amersham Pharmacia Biotech) and T4 polynucleotide kinase prior to PCR. The subsequent PCR product was then gel–purified and subjected to sequencing by the Maxam-Gilbert method according to Sambrook et al. (1989).

Table 2 Primer sequences

Name	5′-3′ sequence[a]	Position[b]
OP-54	ctcggctgggttcaagtcc	1044
OP-267	tcccacacgtgccacaac	184
OP-268	gcagtagatgcgacttttg	133-C
OP-269	tgctgtgcaggttttgcc	284-C
OP-270	ccgggtcccaacgggggat	407
OP-271	gtggttaggaccgtgtct	740-C
OP-380	aattaatacgactcactatagggaaccatggcgctcacca	26
OP-381	aattaatacgactcactatagggtagaatgtcctgcagaga	458-C
OP-392	gaggaaggagaagtcgt	−9
OP-446	gttgaacacggctctt	525-C

[a] Underlined nucleotides represent an added T7 promoter sequence
[b] Primers are numbered relative to the position of the 3′ nucleotide in the *Porphyra* intron sequence (OP-54, OP-267, OP-268, OP-381, OP-392, OP-446) presented in Figs 1 and 3, or the *Physarum* PpLSU3 intron sequence (GenBank accesion number L03183: OP-269, OP-270, OP-271, OP-380). C, complementary strand. The 3′ nucleotide of OP-392 corresponds to a position 9 nt upstream of the 5′ splice site

Plasmid construction

The *P. spiralis* intron ORF was amplified from pCR1000-PSA-R with the primers OP-381 and OP-392 (Table 2) and cloned into pUC18 using the SureClone Ligation kit. A PCR approach was used to construct a recombined ORF consisting of the *Porphyra* His-Cys box swapped into a I-*Ppo*I ORF sequence context. Here, the I-*Ppo*I ORF regions upstream of and downstream from the 29 amino-acid His-Cys box were amplified from pPHR33 (Muscarella and Vogt 1989) using the primer sets OP-380/OP-269 and OP-270/OP-271, respectively (Table 2). The *Porphyra* His-Cys box was amplified from pCR1000-PSA-R using the primers OP-267 and OP-268 (Table 2). The three PCR products were blunt-end ligated and inserted into pUC18 in a series of cloning steps, and the construct with the desired arrangement was confirmed by DNA sequencing.

ORF expression and activity assay

The *P. spiralis* intron ORF and the recombinant ORF were transcribed by T7 RNA polymerase and translated in vitro using the Rabbit Reticulocyte Lysate System kit, labelled with L-[^{35}S] Cysteine (10 μCi/μl; Amersham Pharmacia Biotech). Both ORFs and I-*Ppo*I were tested for cleavage activities according to the I-*Ppo*I assay described by Muscarella et al. (1990) on corresponding intron-less SSU rDNA and LSU rDNA sequences, respectively. Furthermore, the *Porphyra* intron ORF activity was tested under a variety of conditions, including different temperatures (15 °–65 °C) and at different pHs (7–9).

Results and discussion

SSU rDNA group-I introns in *S. pupukensis* and *P. spiralis*

The green alga *S. pupukensis* and the red alga *P. spiralis* contain large group-I introns in their nuclear SSU rRNA genes inserted beyond positions 40 and 1506 (according to the *Escherichia coli* SSU rDNA sequence numbering), respectively (Table 1). Whereas position 1506 in SSU rDNA is a common intron insertion site in protists and fungi (Bhattacharya et al. 1994; Shinohara et al. 1996; Nishida et al. 1998), no introns have previously been reported at position 40. Secondary structure models of the *Scenedesmus* and *Porphyra* introns are presented in Fig. 1. These structures are based upon known general and special features among the group-I intron structures (Michel and Westhof 1990; Cech et al. 1994; Lehnert et al. 1996; Golden et al. 1998) as well as sequence comparisons of homologous Rhodophyta introns (Table 3; Oliveira and Ragan 1994). Both the *Scenedesmus* and *Porphyra* ribozymes (Fig. 1) are typical for the group-IC1 introns, a subgroup very well studied at the RNA structure level (Michel and Westhof 1990; Cate et al. 1996; Lehnert et al. 1996; Golden et al. 1998). The unusual length of both the *Scenedesmus* and *Porphyra* introns (Table 1) is explained by large P1 extension sequences of 1895 nt and 482 nt, respectively.

The *Porphyra* intron self-splices in vitro and generates full-length intron RNA circles

To analyze the catalytic activity of the *Porphyra* intron RNA, linearized pCR1000-PSA-R plasmid (Oliveira and Ragan 1994) containing the intron and some flanking exon sequences was transcribed using T7 RNA polymerase. The corresponding RNA was incubated under splicing conditions (1 M KCl, 25 mM Mg^{2+}, 45 °C), and a representative time-course experiment is presented in Fig. 2 A. The majority of precursor RNA was spliced after 30 min of incubation, generating free intron RNA (Int) and ligated exon RNA (LE). This observation corroborates the findings reported by Oliveira and Ragan (1994).

A circular RNA has a different relative migration pattern in polyacrylamide gels compared to a linear RNA (see Einvik et al. 1997), and we presumed the slow-migrating RNA species generated during splicing to represent intron circles. To analyze the intron circle junction, the RNA fragment was eluted from the polyacrylamide gel, amplified by RT-PCR and then cloned and sequenced. Four independent clones were all identical in sequence. A representative result is shown in Fig. 2 B (left) implying the formation of full-length (FL) intron circles during splicing. To ascertain if the FL intron circle is the major circular RNA species made, the RT-PCR product was subjected to a chemical direct-sequencing approach (see Materials and methods). As expected, a FL intron circle junction was observed (Fig. 2 B, right).

The formation of FL intron RNA circles is rare among most nuclear group-I introns, but has been observed during the splicing of ORF-containing introns both in vitro and in vivo (Decatur et al. 1995; Einvik et al. 1997; Vader et al. 1999). Studies of *Naegleria* twin-ribozyme introns (Einvik et al. 1997) suggest that FL circle formation of group-IC1 intron RNA is directly dependent on the presence of large peripheral extension sequences. Whereas FL circles were made both from the

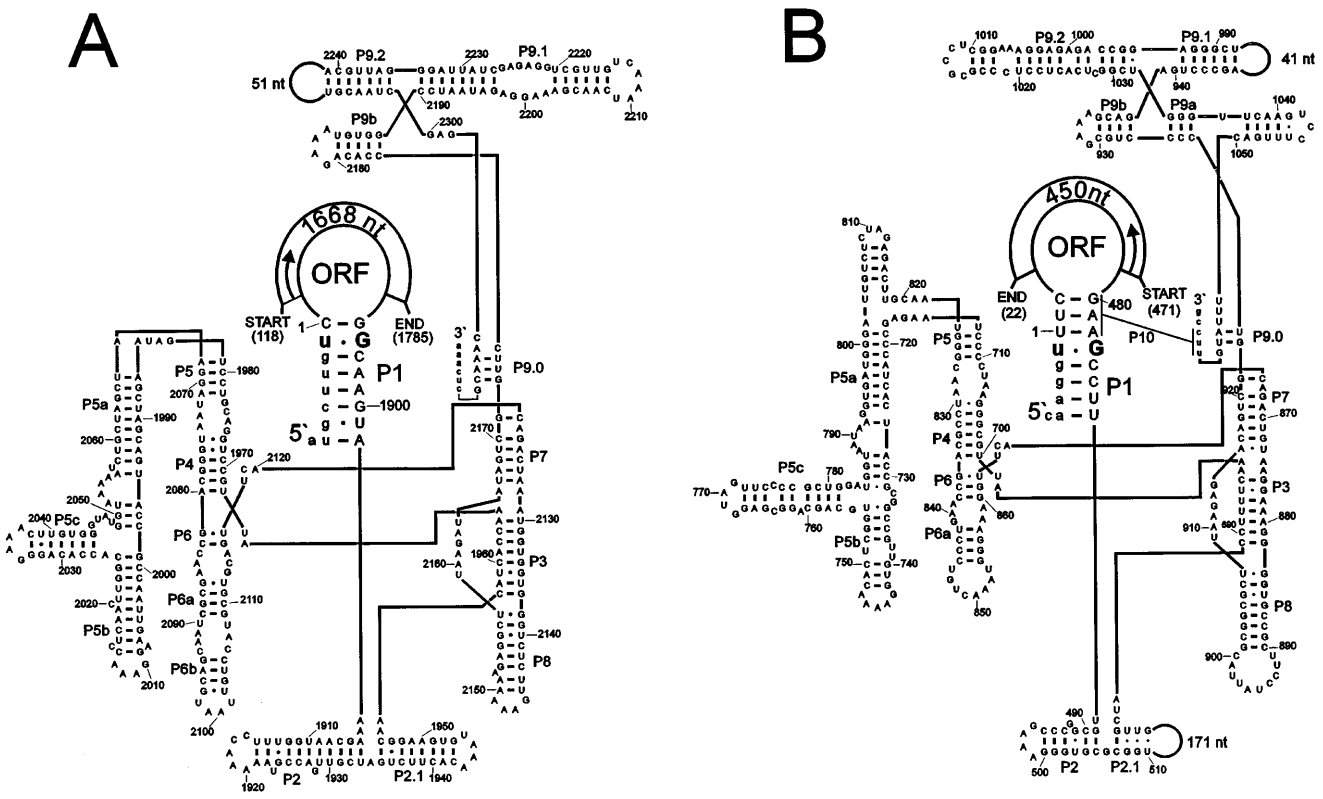


Fig. 1 A,B Secondary structure model of the nuclear group-I introns of *S. pupukensis* (**A**) and *P. spiralis* (**B**). Paired RNA segments (P1–P10) are indicated according to Cech et al. (1994), and the intron sequences are shown in *uppercase* and exon sequences in *lowercase letters*. Both introns are *numbered* with the first nucleotide of the intron as nucleotide number 1. *5′* and *3′* intron splice sites; *ORF* open reading frame

Table 3 Nuclear group-I introns in SSU rDNA position 1506 in the Rhodophyta

Species	Intron size[a]	P1 size[b]	HC[c]	Acc[d]
Porphyra spiralis	1056	482	+	L26177
Porphyra spiralis	908	235	+	L26176
Porphyra spiralis	744	170	–	L26175
Porphyra tenera	960	518	+	AB013175
Porphyra tenera	783	342	+	AB013176
Porphyra dentata	572	52	–	AB013183
Porphyra haitanensis	554	61	–	AB013181
Porphyra katadae	588	83	–	AB013184
Porphyra pseudolinealis	581	108	–	AB013185
Porphyra yesoensis	526	80	–	AB017083
Porphyra yesoensis	526	77	–	AB017081
Porphyra yesoensis	514	79	–	AB017078
Porphyra yesoensis	509	77	–	AB013178
Porphyra sp. 1	654	150	–	AB013182
Porphyra sp. 2	967	521	+	AB017077
Bangia atropurpurea	1038	514	+	L36066
Hildenbrandia ruba	501	13	–	L19345
Hildenbrandia ruba	511	18	–	AF076995

[a] Size of introns in base pairs
[b] Size of P1-extensions in base pairs, confined by the conserved u:G pair in the P1 segment
[c] His-Cys box motif present (+) or absent (–)
[d] EMBL/Genbank accession numbers

Naegleria wild-type and deletion-mutants containing 930-nt and 630-nt P6 extensions, respectively, circles lacking the first two intron-encoded nucleotides were observed in the introns without a P6 extension (Einvik et al. 1997).

The *Scenedesmus* intron encodes a 555 amino-acid putative protein

The 1895-nt P1 extension sequence of the *Scenedesmus* group-I intron contains an open reading frame (ORF) potentially encoding a protein of 555 amino acids (Fig. 3 A), which is much larger than any other known ORF from a nuclear group-I intron (see Table 1). No known sequence motifs corresponding to an intron endonuclease, maturase, or reverse transcriptase could be recognized. This feature appears similar to the 687-nt ORF in a recently reported nuclear group-I intron from the fungus *Protomyces* (Nishida et al. 1998; Table 1). Some mitochondrial group-I introns harbor ORFs encoding putative structural proteins such as ribosomal proteins (Burke and RajBhandary 1982; Cummings et al. 1990; Hausner et al. 1999) or NADH dehydrogenase protein subunits (Cummings et al. 1990; Beagley et al. 1996), some of them which are comparable in size to the *Scenedesmus* intron protein. A TFASTA database search revealed similarities to several bacterial and chloroplast S9 ribosomal proteins at the N-terminal region of the *Scenedesmus* intron protein sequence (Fig. 3 B), and thus resembles the recently reported *Cryphonectria* mitochondrial rRNA group-I intron

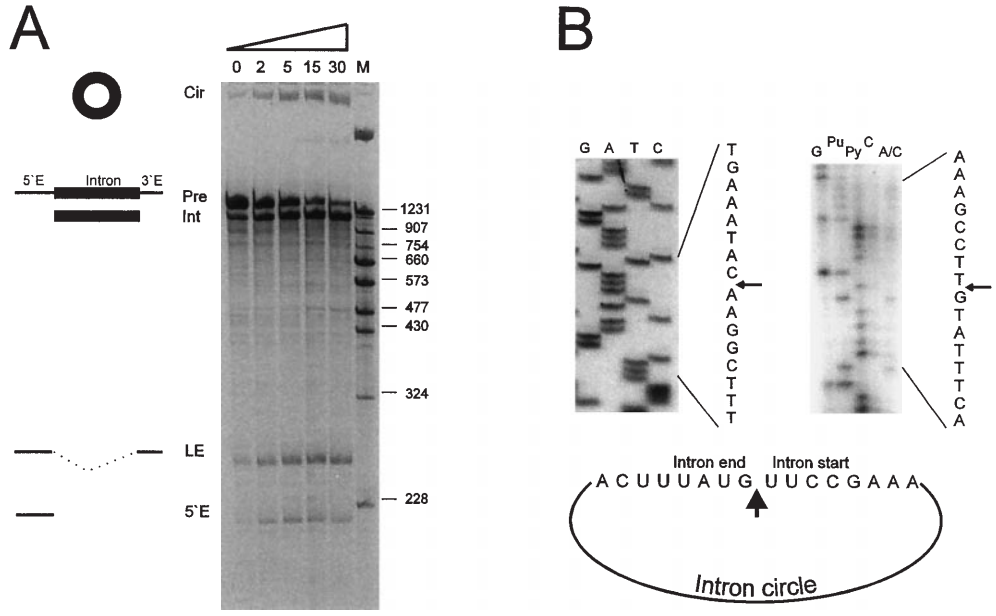


Fig. 2 A,B Gel analysis of in vitro self-splicing products and RNA circle junctions of the *Porphyra* intron. **A** time course self-splicing reaction incubated for 0–30 min and analyzed on an 8 M urea-5% polyacrylamide gel. The observed RNAs are the primary transcript (*Pre*: 1312 nt); excised intron (*Int*: 1057 nt; including the exogenous guanosine); ligated exons (*LE*: 256 nt); 5′ exon (*5′E*: 218 nt). *Cir* intron circle; *M* RNA size marker. The 3′ exon RNA (38 nt) was run off the gel. **B** region corresponding to the circle junction of isolated circular intron RNA was amplified by inverse RT-PCR and sequenced. Schematic presentation of the full-length intron junction (*arrow*) sequence is presented below. Left: enzymatic sequencing of a cloned intron junction product. Right: chemical direct-sequencing of an intron junction product. *Pu* purine (G or A); *Py* pyrimidine (C or T); *A/C* A or C. *Horizontal arrows* indicate the circle junction

which encodes a putative S5 ribosomal protein fused at its C-terminus to a maturase-like protein (Hausner et al. 1999). Perhaps the group-I intron in *Scenedesmus* nuclear rDNA codes for a similar structural ribosomal protein.

The *Porphyra* intron encodes a homing endonuclease-like protein on the opposite strand

We have previously noted that the *P. spiralis* group-I intron appears to carry an ORF within the P1-extension, but on the complementary strand to that encoding the SSU rRNA and the group-I ribozyme (Vader et al. 1994). Re-sequencing of the P1-extension revealed an ORF corresponding to a 149 amino-acid His-Cys box protein (Fig. 3 C), a hallmark of nuclear homing endonucleases (Johansen et al. 1993). The antisense location of protein genes within nuclear rDNA has been reported in only a few eukaryotes. Recently we noted an antisense His-Cys box ORF motif within a group-I intron of the fungus *Nectria galligena* (Johansen and Haugen 1999), and the protists *Giardia lamblia* and *Entamoeba histolytica* have previously been reported to express antigen and hemolysin, respectively, from the complementary rRNA coding strand (Upcroft et al. 1990; Jansson et al. 1994).

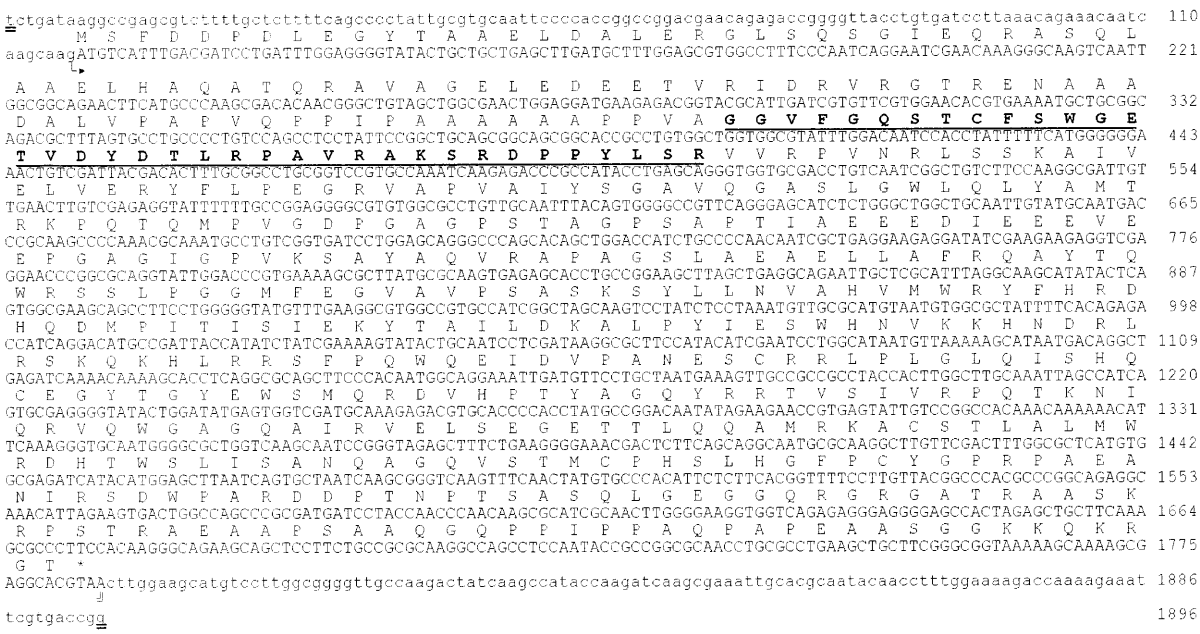
To test for endonuclease activity in the *Porphyra* intron protein, the ORF was over-expressed in vitro in a reticulocyte extract (see Materials and methods). We were unable to detect any cleavage activity on an intronless SSU rDNA target using various conditions that work well for the I-*Ppo*I enzyme (data not shown). In a different approach, the conserved 29 amino-acid His-Cys box motif from the *Porphyra* sequence was exactly swapped at the DNA-level into a I-*Ppo*I sequence context. Sixteen of the twenty nine residues in the His-Cys box motif are identical between these proteins (see alignment in Fig. 5), suggesting that the *Porphyra* motif might be able to complement the corresponding I-*Ppo*I motif. However, no cleavage activity was observed for the recombined protein at conditions where the I-*Ppo*I endonuclease was very active (data not shown).

The *Porphyra* intron ORF appears to be a homing-endonuclease pseudogene

A database homology search identified 18 similar Rhodophyta group-I introns at the same insertion site (1506) in nuclear SSU rDNA (Table 3). Interestingly, all introns from *Bangia* and *Porphyra* (class Bangiophyceae) contain P1-extensions of various lengths (52–521 nt). The six largest P1-extensions, one in *Bangia* and five in *Porphyra* species, contain His-Cys box motifs coded on the complementary strand (Table 3). A schematic presentation of the P1 extensions is shown in Fig. 4. The *P. spiralis* intron size-variants reported by Oliveira and Ragan (1994) are due to deletions in the ORF region. Here we note a similar example of the two variant intron forms of *P. tenera* (Fig. 4).

A.

S. pupukensis



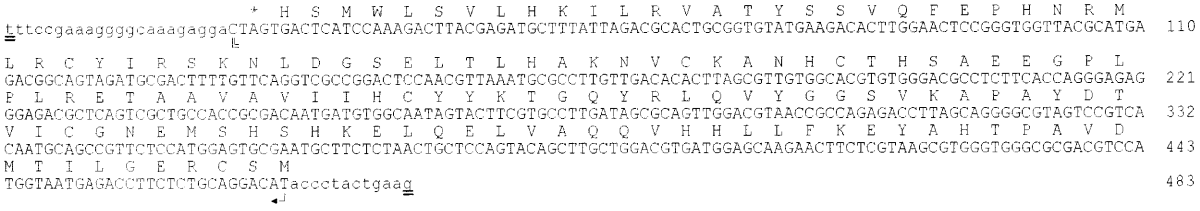
B.

S. pupukensis	G	G	V	F	G	Q	S	T	C	F	S	W	G	E	T	V	D	Y	D
H. influenza	.	.	.	I	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E. coli	.	.	.	I	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P. purpurea	.	.	.	L	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G. theta	.	.	.	L	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. vulgaris	.	.	.	L	S	A	.	.	.	.	.	.	.	.	.	.	.	.	.

S. pupukensis	T	L	R	P	A	V	R	A	K	S	R	D	P	P	Y	L	S	R	V
H. influenza	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E. coli	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P. purpurea	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G. theta	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. vulgaris	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

C.

P. spiralis



An alignment of the putative amino-acid sequences from the four largest P1-extensions (Table 3) is presented in Fig. 5. The following interesting features with implication for protein function and evolution are noted. First, all but *P. spiralis* lack an AUG start codon in their corresponding reading frame. Second, a common N-terminal sequence of the putative intron proteins can not be identified from the comparisons. Third, an alignment of 94 positions at the C-terminal end can only be achieved after introducing a reading-frame shift in three of the four intron ORF sequences (see Fig. 4). Interestingly, a minus one frame shift which we confirmed by DNA sequencing, is present close to the 3′-end of the *P. spiralis* intron ORF (Fig. 5). This frame shift results in a truncated C-terminal end lacking three essential zinc-binding residues, corresponding to C132,

Fig. 3 A–C The P1 extension sequences of the *S. pupukensis* (**A**) and *P. spiralis* (**C**) introns. The ORF sequences in **A** and **C** are shown in *uppercase*, flanking sequences in *lowercase letters*, with the corresponding amino-acid sequences given above the ORFs. Nucleotides corresponding to the conserved U:G pair in P1, which define the 5′-splice site (SS), are *underlined*. The nucleotide numbering is according to the intron sequences in Fig. 1. Note that the *P. spiralis* ORF is located on the complementary strand. **B** sequence motif in the *S. pupukensis* intron ORF (*bold and underlined* in **A**) compared to S9 ribosomal protein sequences from the bacteria *H. influenza* (accession number AAC23092 in the NCBI protein sequence data base) and *E. coli* (R3EC9), and the algal chloroplasts of *P. purpurea* (AAC08177), *G. theta* (AAC35726) and *C. vulgaris* (BAA57995). Identical positions are indicated by *dots* and introduced gaps by *dashes*. Regions which are highly similar between the intron ORF and bacterial proteins are *boxed*

H134 and C138 of I-*Ppo*I (Fig. 5). This truncation may explain the lack of cleavage activity in the in vitro translated protein.

Horizontal transfer of endonuclease ORFs has been suggested in both fungi mitochondria and T4 phage group-I introns (Loizos et al. 1994; Sellem and Belcour 1997). To test if a similar transfer may have occurred among the Bangiophyceae ORFs, separate comparisons of the four endonuclease-like protein sequences from the *Porphyra* and *Bangia* 1506-introns (Fig. 5), their corresponding intron cores, and SSU rDNA sequences, were performed. The same pattern of relationship was observed between the rRNA and intron core sequences, suggesting a vertical transmission of the 1506-introns (data not shown). However, the *Bangia* and *P. spiralis* ORFs were found to be more closely related (69% amino-acid sequence identity) than the two *Porphyra* ORFs (*Porphyra tenera* and *P. spiralis*; 52%). These results indicate horizontal transfer of the endonuclease ORFs independent of their corresponding intron cores during evolution of the Bangiophyceae.

Nuclear homing endonucleases have been reported to generate double-strand breaks at intron-lacking rDNA alleles (Muscarella and Vogt 1989; Johansen et al. 1997; Elde et al. 1999) and are lethal when expressed in yeast (Muscarella and Vogt 1993; Lin and Vogt 1998). Endonuclease expression in mobile versions of Bangiophyceae 1506-introns would challenge the integrity of chromosomes. Subsequently, they may have been selected against during evolution by introducing truncations and frame shifts, as observed in present-day Bangiophyceae intron ORFs.

Concluding remarks

The *S. pupukensis* and *P. spiralis* introns represent two unusual examples of ORF-containing group-I introns in nuclear rDNA. Both algal introns interrupt the same gene (SSU rRNA gene), belong to the same subgroup (IC1), and harbour ORFs in the same location (P1 loop region). Most nuclear group-I intron ORFs correspond to His-Cys box homing endonucleases (see Table 1) expressed in vivo from excised intron RNA (Lin and Vogt 1998; Vader et al. 1999). However, the *Scenedesmus* ORF appears to encode a large structural ribosomal-like protein and the *Porphyra* ORF seems to be a His-Cys box homing-endonuclease pseudogene located on the complementary strand. Whereas gain of an intron ORF may be due to an insertional event, corresponding loss appears circumstantial involving a series of frameshift and partial deletion steps.

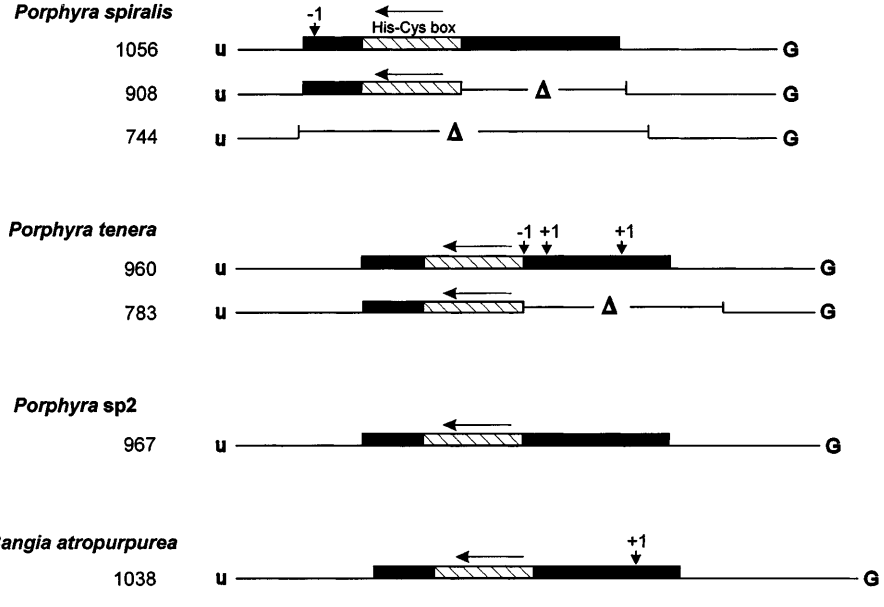


Fig. 4 Schematic presentation of P1-extensions in different size versions of *Porphyra spiralis*, *Porphyra tenera*, *Porphyra* sp. 2 and *Bangia atropurpurea* group-I introns (see Table 3). The extensions are confined by the conserved U:G pair which defines the 5′-SS. Reading frames are denoted by *black boxes*, except for the His-Cys box motifs (*hatched boxes*; see Fig. 5). Frame shifts (–1 or +1) are indicated above the reading frames, and deletions (Δ) are defined by *short vertical lines* in the diagrams

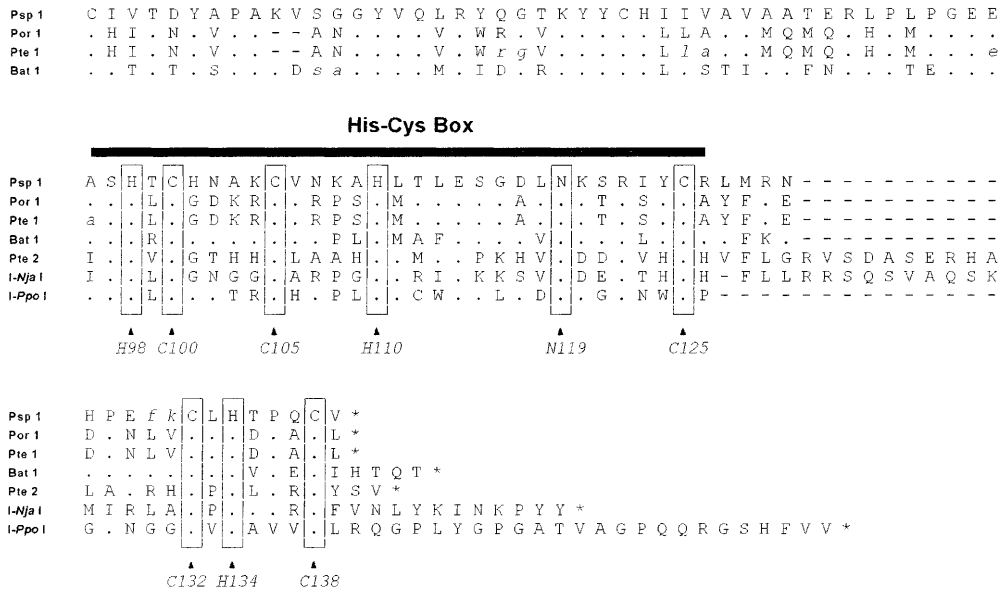


Fig. 5 Endonuclease-like sequences from Bangiophyceae S-1506 group-I introns. The 93 amino-acid C-terminal sequence of the *Porphyra spiralis* intron ORF (Psp 1) is aligned to the corresponding sequences from *Porphyra tenera* (Pte 1), *Porphyra* sp. (Por 1) and *Bangia atropurpurea* (Bat 1). Identical positions are indicated by *dots*, deletions by *dashes*, and residues inferred from reading-frame shift corrections are shown in *lowercase letters*. The His-Cys box motif (*boxed*) is compared to the S-516 intron ORF from *P. tenera* (Pte 2), and to the homing endonucleases I-*Nja*I and I-*Ppo*I. Conserved residues proposed to be directly involved in zinc binding (C100, C105, H110, C125, C132, H134, C138) and the active site (H98, N119) of the I-*Ppo*I endonuclease (Flick et al. 1998) are indicated. See Tables 1 and 3 for sequence references

Acknowledgments We thank Mark A. Ragan for providing the pCR1000-PSA-R plasmid. This work was supported by grants from The Norwegian Research Council (S.J.), The Norwegian Cancer Society (S.J.) and The Aakre Foundation for Cancer Research (S.J.).

References

Bhattacharya D, Surek B, Rüsing M, Damberger S, Melkonian M (1994) Group-I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). Proc Natl Acad Sci USA 91:9916–9920

Beagley CT, Okada NA, Wolstenholme DR (1996) Two mitochondrial group-I introns in a metazoan, the sea anemone *Metridium senile*: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. Proc Natl Acad Sci USA 93:5619–5623

Belfort M, Roberts RJ (1997) Homing endonucleases: keeping the house in order. Nucleic Acids Res 25:3379–3388

Burke JM, RajBhandary UL (1982) Introns within the large rRNA gene of *N. crassa* mitochondria: a long open reading frame and a consensus sequence possibly important in splicing. Cell 31:509–520

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996) Crystal structure of a group-I ribozyme domain: principles of RNA packing. Science 273:1678–1685

Cech TR, Herschlag D (1996) Group-I ribozymes: substrate recognition, catalytic strategies, and comparative mechanistic analysis. Nucleic Acids Mol Biol 10:1–17

Cech TR, Damberger SH, Gutell RR (1994) Representation of the secondary and tertiary structure of group-I introns. Nature Struct Biol 1:273–280

Cummings DJ, McNally KL, Domenico JM, Matsuura ET (1990) The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. Curr Genet 18:375–402

Decatur WA, Einvik C, Johansen S, Vogt VM (1995) Two group-I ribozymes with different functions in a nuclear rDNA intron. EMBO J 14:4558–4568

De Jonckheere JF, Brown S (1998) Three different group-I introns in the nuclear large subunit ribosomal DNA of the amoeboflagellate *Naegleria*. Nucleic Acids Res 26:456–461

Einvik C, Decatur WA, Embley TM, Vogt VM, Johansen S (1997) *Naegleria* nucleolar introns contain two group-I ribozymes with different functions in RNA splicing and processing. RNA 3:710–720

Einvik C, Nielsen H, Westhof E, Michel F, Johansen S (1998) Group-I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. RNA 4:530–541

Elde M, Haugen P, Willassen NP, Johansen S (1999) I-*Nja*I, a nuclear intron-encoded homing endonuclease from *Naegleria*, generates a pentanucleotide 3′ cleavage-overhang within a 19 base-pair partially symmetric DNA recognition site. Eur J Biochem 259:281–288

Flick KE, Jurica MS, Monnat RJ Jr, Stoddard BL (1998) DNA binding and cleavage by the nuclear intron-encoded endonuclease I-*Ppo*I. Nature 394:96–101

Golden BL, Gooding AR, Podell ER, Cech TR (1998) A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. Science 282:259–264

Hausner G, Monteiro-Vitorello CB, Searles DB, Maland M, Fulbright DW, Bertrand H (1999) A long open reading frame in the mitochondrial LSU rRNA group-I intron of *Cryphonectria parasitica* encodes a putative S5 ribosomal protein fused to a maturase. Curr Genet 35:109–117

Holst-Jensen A, Vaage M, Schumacher T, Johansen S (1999) Structural characteristics and possible horizontal transfer of group-I introns between closely related plant pathogenic fungi. Mol Biol Evol 16:114–126

Jansson Å, Gillin F, Kagardt U, Hagblom P (1994) Coding of hemolysins within the ribosomal RNA repeat on a plasmid in *Entamoeba histolytica*. Science 263:1440–1443

Johansen S, Haugen P (1999) A complex group-I intron in *Nectria galligena* rDNA. Microbiology 145:516–517

Johansen S, Vogt VM (1994) An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group-I ribozyme and a novel ribozyme that cooperate in self-splicing. Cell 76:725–734

Johansen S, Embley TM, Willassen NP (1993) A family of nuclear homing endonucleases. Nucleic Acids Res 21:4405

Johansen S, Muscarella DE, Vogt VM (1996) Insertion elements in ribosomal DNA. In: Zimmermann RA, Dahlberg AE (eds) Ribosomal RNA: structure, evolution, processing, and function in biosynthesis. CRC Press, Boca Raton, Florida, pp 89–108

Johansen S, Elde M, Vader A, Haugen P, Haugli K, Haugli F (1997) In vivo mobility of a group-I twintron in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. Mol Microbiol 24:737–745

Kessler E, Schäfer M, Hümmer C, Kloboucek A, Huss VAR (1997) Physiological, biochemical, and molecular characters for the taxonomy of the subgenera of *Scenedesmus* (Chlorococcales, Chlorophyta). Bot Acta 110:244–250

Kunimoto M, Yamamoto Y, Kaminishi Y, Mizukami Y, Kito H (1998) A discrimination of *Porphyra* species (Rhodophyta) based on the sequences of the small subunit ribosomal RNA gene. DDBJ/ EMBL/ GenBank database accession number AB013176

Lehnert V, Jaeger L, Michel F, Westhof E (1996) New loop-loop tertiary interactions in self-splicing introns of subgroups IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme. Chem Biol 3:993–1009

Lin J, Vogt VM (1998) I-*Ppo*I, the endonuclease encoded by the group-I intron PpLSU3, is expressed from an RNA polymerase-I transcript. Mol Cell Biol 18:665–675

Loizos N, Tillier ERM, Belfort M (1994) Evolution of mobile group-I introns: recognition of intron sequences by an intron-encoded endonuclease. Proc Natl Acad Sci USA 91:11983–11987

Margulis L, Schwartz KV (1998) Five kingdoms. An illustrated guide to the phyla of life on earth, 3rd edn. WH Freeman, New York

Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. J Mol Biol 216:585–610

Muscarella DE, Vogt VM (1989) A mobile group-I intron in the nuclear rDNA of *Physarum polycephalum*. Cell 56:443–454

Muscarella DE, Vogt VM (1993) A mobile group-I intron from *Physarum polycephalum* can insert itself and induce point mutations in the nuclear ribosomal DNA of *Saccharomyces cerevisiae*. Mol Cell Biol 13:1023–1033

Muscarella DE, Ellison EL, Ruoff BM, Vogt VM (1990) Characterization of I-*Ppo*, an intron-encoded endonuclease that mediates homing of a group-I intron in the ribosomal DNA of *Physarum polycephalum*. Mol Cell Biol 10:3386–3396

Nishida H, Tajiri Y, Sugiyama J (1998) Multiple origins of fungal group-I introns located in the same position of the nuclear SSU rRNA gene. J Mol Evol 46:442–448

Oliveira MC, Ragan MA (1994) Variant forms of a group-I intron in nuclear small-subunit rRNA genes of the marine red alga *Porphyra spiralis* var. *amplifolia*. Mol Biol Evol 11:195–207

Roman J, Rubin MN, Woodson SA (1999) Sequence specificity of in vivo reverse splicing of the *Tetrahymena* group-I intron. RNA 5:1–13

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sellem CH, Belcour L (1997) Intron open reading frames as mobile elements and evolution of a group-I intron. Mol Biol Evol 14:518–526

Shinohara ML, LoBuglio KF, Rogers SO (1996) Group-I intron family in the nuclear ribosomal RNA small subunit genes of *Cenococcum geophilum* isolates. Curr Genet 29:377–387

Upcroft JA, Healey A, Mitchell R, Boreham PFL, Upcroft P (1990) Antigen expression from the ribosomal DNA repeat unit of *Giardia intestinalis*. Nucleic Acids Res 18:7077–7081

Vader A, Næss J, Haugli K, Haugli F, Johansen S (1994) Nucleolar introns from *Physarum flavicomum* contain insertion elements that may explain how mobile group-I introns gained their open reading frames. Nucleic Acids Res 22:4553–4559

Vader A, Nielsen H, Johansen S (1999) In vivo expression of the nucleolar group-I intron-encoded I-*Dir*I homing endonuclease involves the removal of a spliceosomal intron. EMBO J 18:1003–1013

Woodson SA, Cech TR (1989) Reverse self-splicing of the *Tetrahymena* group-I intron: implication for the directionality of splicing and for intron transposition. Cell 57:335–345