Supplementary material to accompany the article:

Towards predicting self-splicing and protein-facilitated splicing of group I introns

Quentin Vicens, Paul J. Paukstelis, Eric Westhof, Alan M. Lambowitz and Thomas R. Cech

This supplementary material contains additional discussions, figures, tables and a list of the corresponding references.

Supplementary discussion

How close are we from predicting self-splicing activity based on the tertiary interactions at the core?

Understanding how the tertiary interactions directly helping substrate binding and its positioning within the catalytic site are favoring the catalytic process is not as straightforward as recognizing the role played by the long range interactions between peripheral elements. The six long range interactions which involve the junction regions J3/4, J4/5, J6/6a, J6/7, and J8/7 could be incorporated into the interaction networks of both active and inactive introns (Figures 5 and S1). All introns contained twelve base pairs between the G.U pair at the cleavage site and the L2 tetraloop (when a P2 element was present and not implicated in different types of interactions), thus maintaining the ability to position the substrate G.U pair in the active site (Michel & Westhof, 1990). The 12 introns also possessed a conserved catalytic site, as well as identical interactions between J4/5 and P1 and between J6/6a and P3 (Figures 5 and S1). A likely explanation to the difficulty in accounting for self-splicing activity from considering these interactions is that the relative stabilities and dynamic properties of these contacts would need to be monitored within the context of each intron. Some interactions may be less stable, i.e. more dynamic, than others, in certain introns but not all (Guo et al., 2006), thereby either facilitating folding or contributing to misfolding scenarios (Pan & Woodson, 1998; Pan & Woodson, 1999; Woodson & Rangan, 2008). However, one trend that is apparent from this survey is that the tertiary interactions of introns which are self-splicing have a higher fraction of base triples and A-minor interactions involving adenines and G—C pairs (compare P6 bp1,2—J6/7 1,2 and P1 bp-3,4—J8/7 1/2,2/3 for the inactive or poorly active P.b., A.p.LSU and A.p.SSU introns and the three active tRNA introns; Table S2). Adenines are the most common nucleotides found in single stranded regions and are involved in the maintenance of RNA tertiary structure through various universal interaction motifs (Doherty et al., 2000; Gutell et al., 2000; Lescoute & Westhof, 2006).

How well are the self-splicing behaviors accounted for by the criteria singled out in this study?

Attempts at determining self-splicing activity from each criterion individually (GC content, number and type of long range interaction in the periphery and at the core, etc.) may be vain. Instead, rationales for self-splicing behavior can be made from an

integrative analysis of these parameters. For example, the A.p.LSU is more GC rich that the C.b intron (41 % vs. 37 %), and was predicted to contain a tetraloop-receptor L9–P5 interaction. Yet, C.b. is significantly more active than A.p.LSU (C.b. and A.p.LSU are respectively 218 % and 30 % as active as the control A.s. intron). The more robust auto-catalytic activity observed for the C.b intron could be explained by the following observations: (1) the C.b. intron possesses consensus L9–P5 and L2—P8 tetraloop-base pair interactions as well as optimal interactions at the core (e.g. between the substrate helix P1 and J8/7); (2) a sarcin/ricin motif predicted in the peripheral element P7.1 could be involved in stabilizing the core by helix-helix packing, as observed in group I introns and other RNAs (Leontis & Westhof, 1998; Waldsich et al., 2002); and (3) the exon context in which the C.b. is found (T loop of a tmRNA) was preserved, while that of the A.p.LSU intron (ribosomal LSU) was truncated for the purpose of this study. Similarly, the absence of self-splicing for the N.c., D.p., and N.a. introns and the poor activity observed for the A.p.SSU, A.p.LSU, and C.n. intron, could be accounted for by the following: (1) these introns have low GC contents (25–34 %), sometimes much lower than the genome (N.c.); (2) they lack a P2 extension that would be involved in a long-range interaction with P8, or the L2 cannot establish such a long range interaction (e.g. CAAC L2 loop in A.p.SSU); (3) longer extensions are found in peripheral domains (e.g. P2 in N.c. and P8 in D.p.); and (4) mutations are likely found in core regions essential for activity, as for example the sequences of the self-splicing C.b. intron and of the poorly self-splicing C.n. intron are 40 % identical (Figure S2). In conlusion, this analysis suggests that even with the remaining blind spots in our understanding of group I intron structure and catalysis, a careful examination of these accessible parameters offers a rationale for the self-splicing behavior observed in vitro.

How well can the splicing behaviors of the two A.p. introns be explained on the basis of their proposed secondary structure?

Based on the secondary structure diagrams (Figure S1), the exact positioning of the P5ab and P5abd domains in the two A.p. introns is unclear. The presence of helix P5d and of single stranded residues between P5 and P5a in A.p.SSU suggests that the P5ab domain could be coaxial with P4–P6 as in the *Tetrahymena* intron (Cate et al., 1996). The apparent absence of unpaired residues between P5 and P5a of A.p.LSU (on one strand only) could indicate that P5 and P5a are in fact stacked. Such a structure could allow CYT-18 binding in a similar fashion to what was observed for other introns (Mohr

et al., 1994; Paukstelis et al., 2008). However, the 15 residues which are 100% conserved in domain P5b suggest that both domains are oriented similarly in both introns.

Supplementary figures

Figure S1: Interaction network diagrams of the 12 introns surveyed.

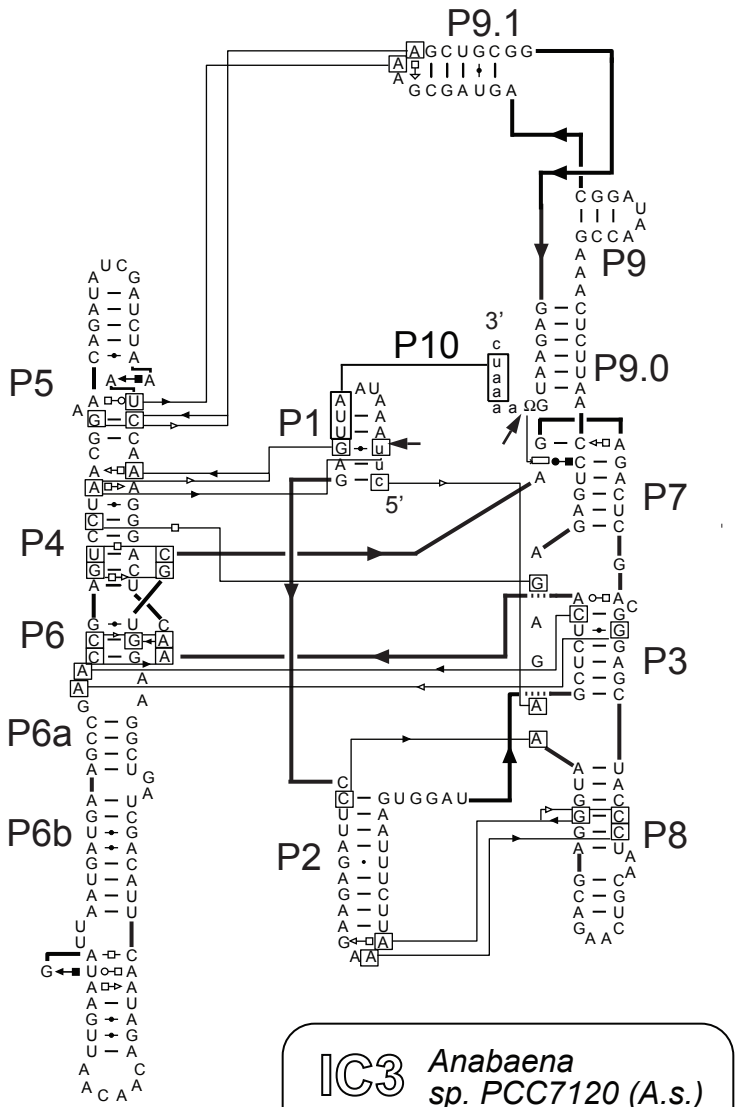
Diagrams are organized by their self-splicing activity (relative to the A.s. intron, see Table 1), as follows: A. self-splicing introns; B. poorly self-splicing introns; C. non-self-splicing introns. The representation is the same as that described in the legend of Figure 5. Dashed lines represent non consensus interactions. Question marks indicate ambiguous topologies and interactions.

Figure S2: Pairwise sequence alignments of the C.n., C.b., A.p.LSU and A.p.SSU introns.

Sequences paired in the secondary structure are highlighted in color. Identical residues are boxed in grey and indicated by the symbol *.

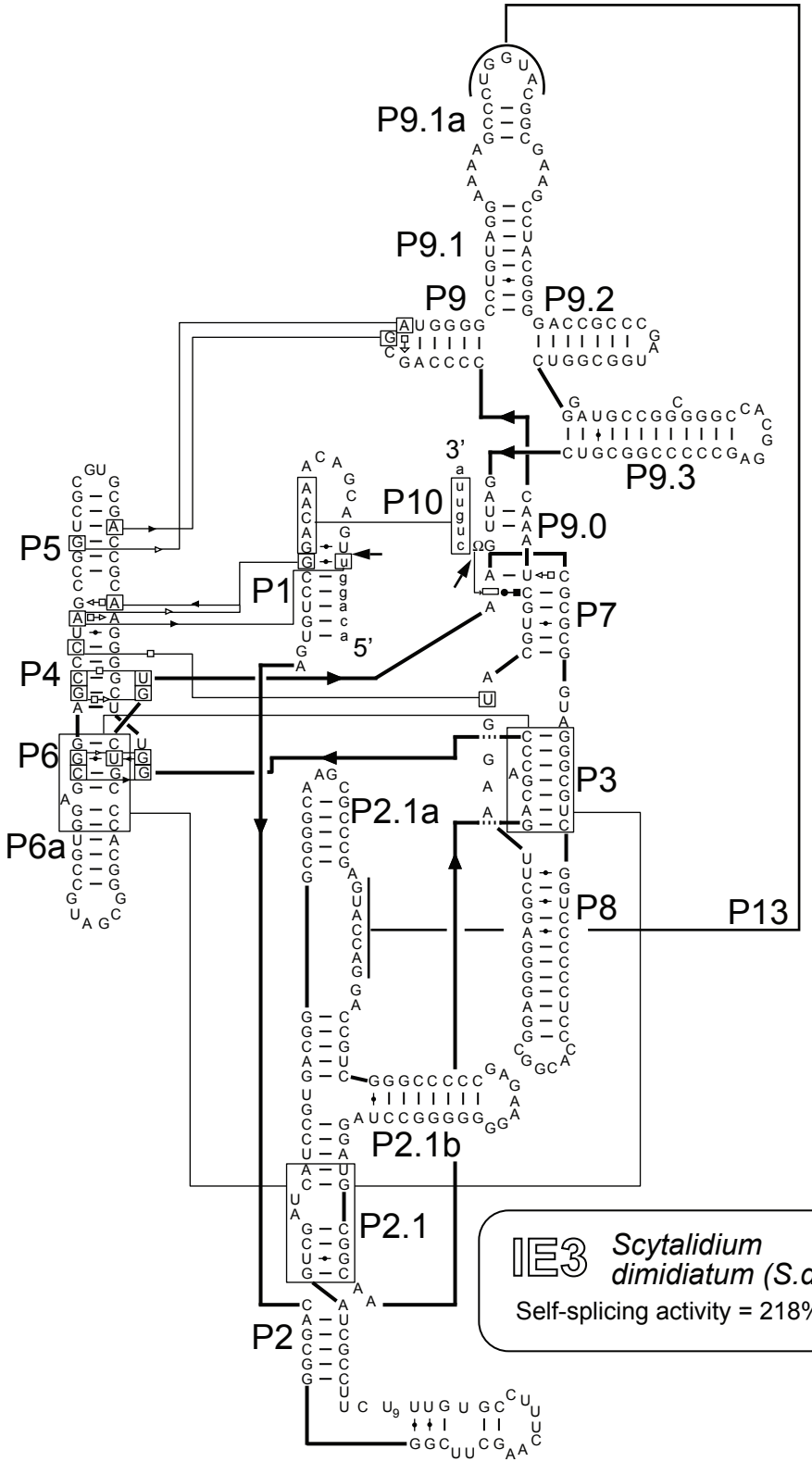
Figure S1

A



IC3 *Anabaena sp. PCC7120 (A.s.)*

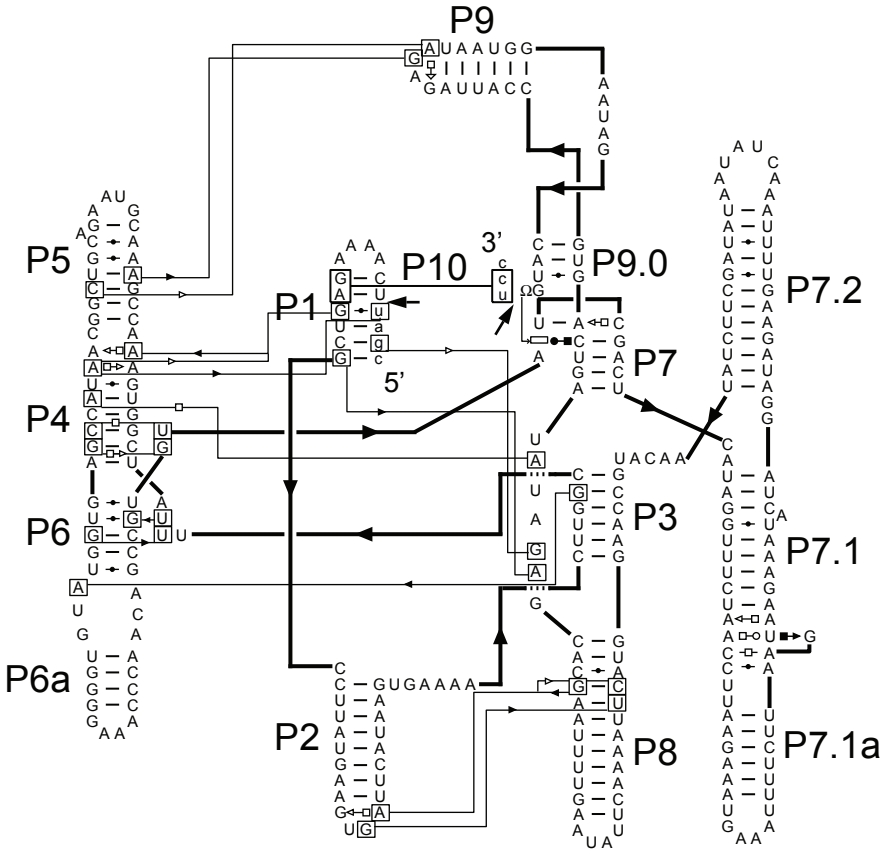
Self-splicing activity = 100%



IE3 *Scytalidium dimidiatum (S.d.)*

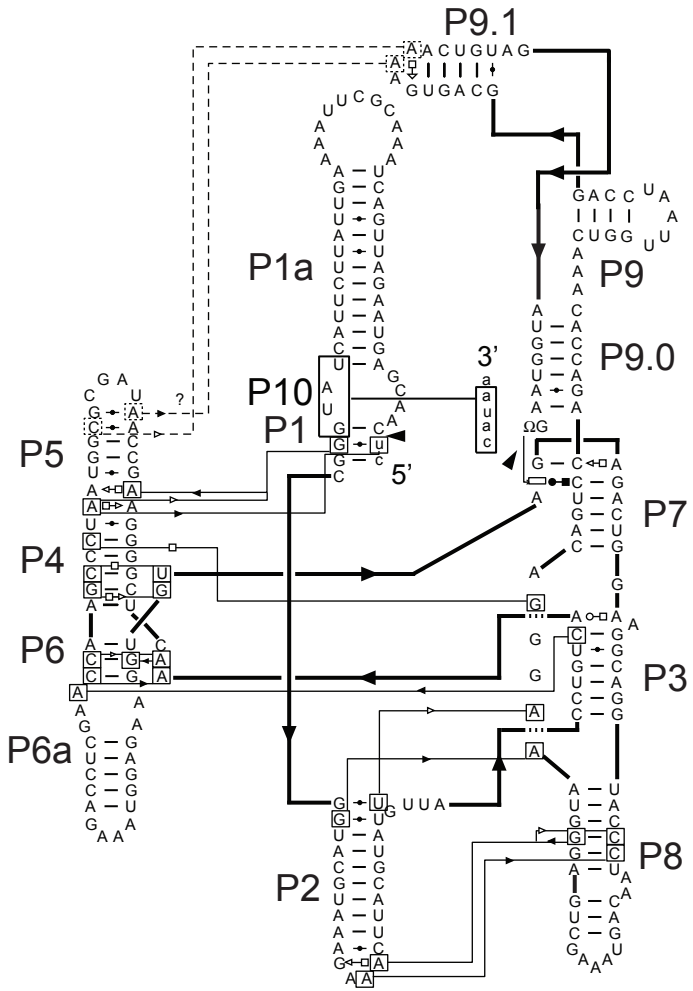
Self-splicing activity = 218%

Figure S1 cont.



IA3 *Clostridium botulinum (C.b.)*

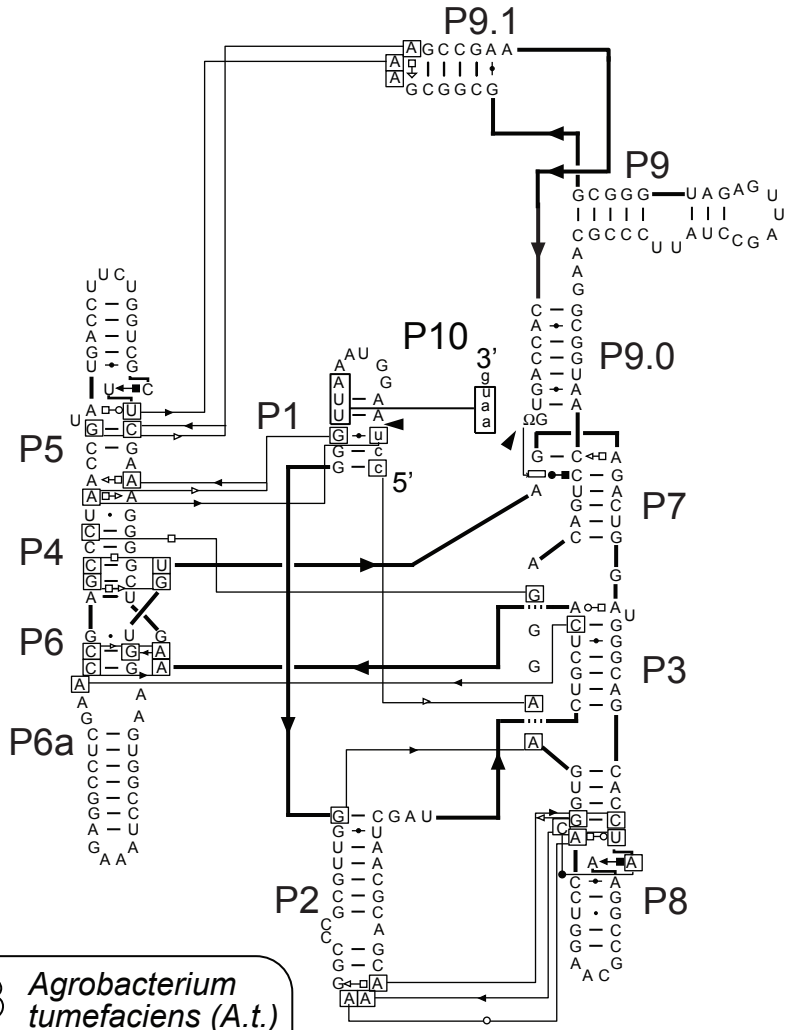
Self-splicing activity = 200%



IC3 *Scytonema hofmanni (S.h.)*

Self-splicing activity = 117%

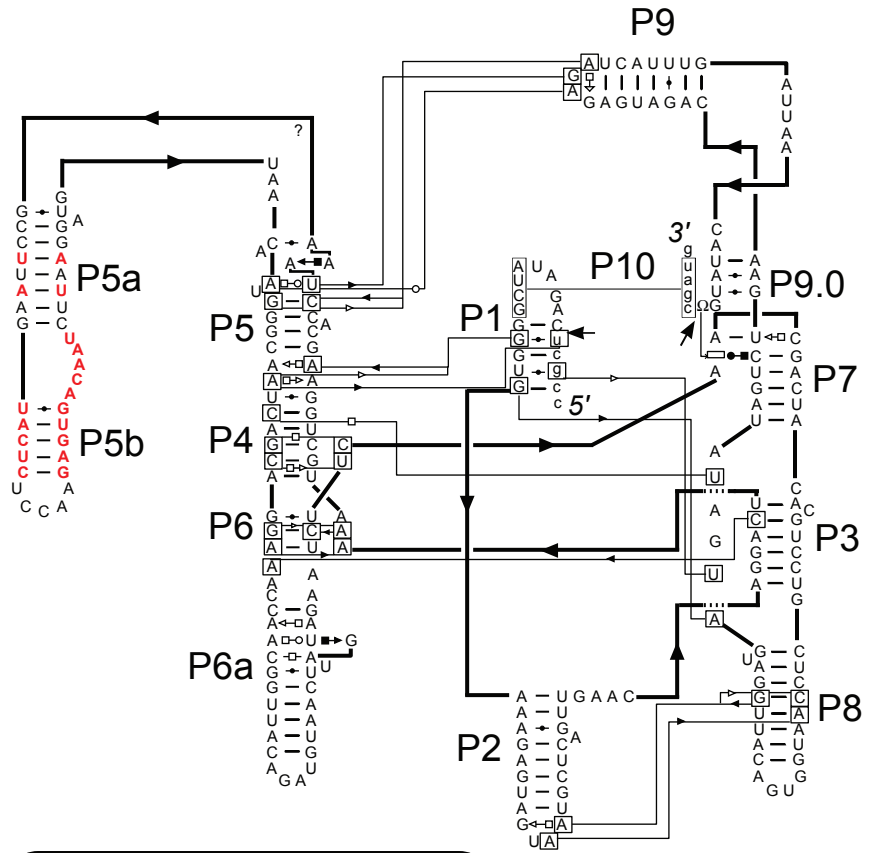
Figure S1 cont.



IC3 *Agrobacterium tumefaciens (A.t.)*

Self-splicing activity = 94%

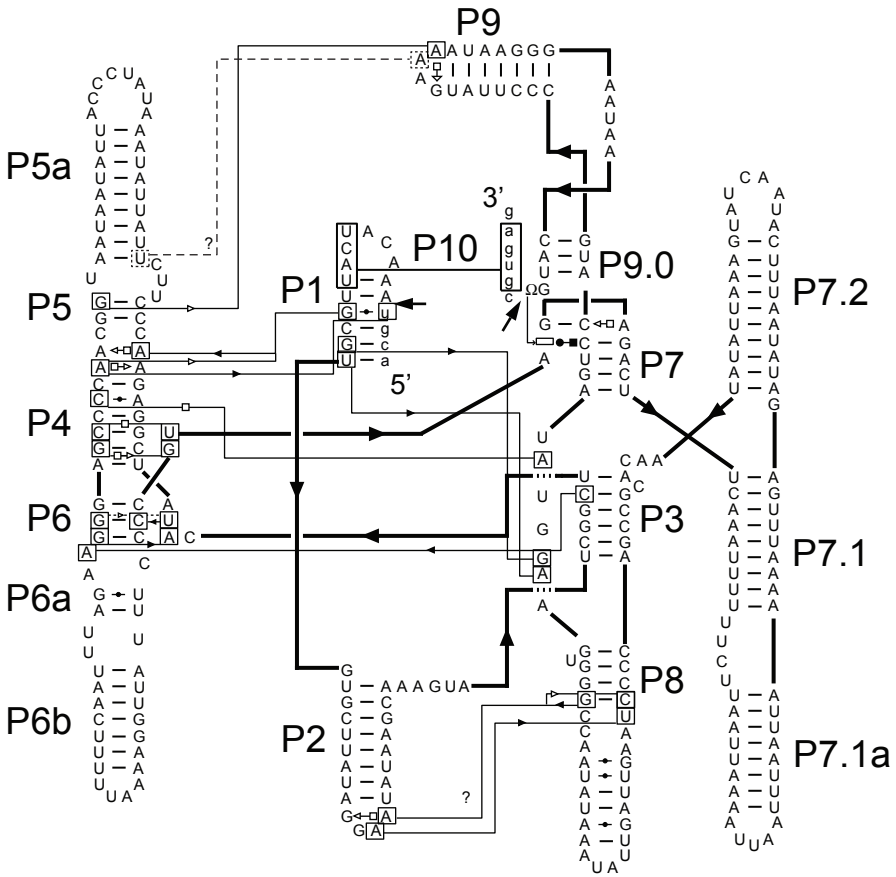
B



IB1 *Amoebidium parasiticum (A.p.LSU)*

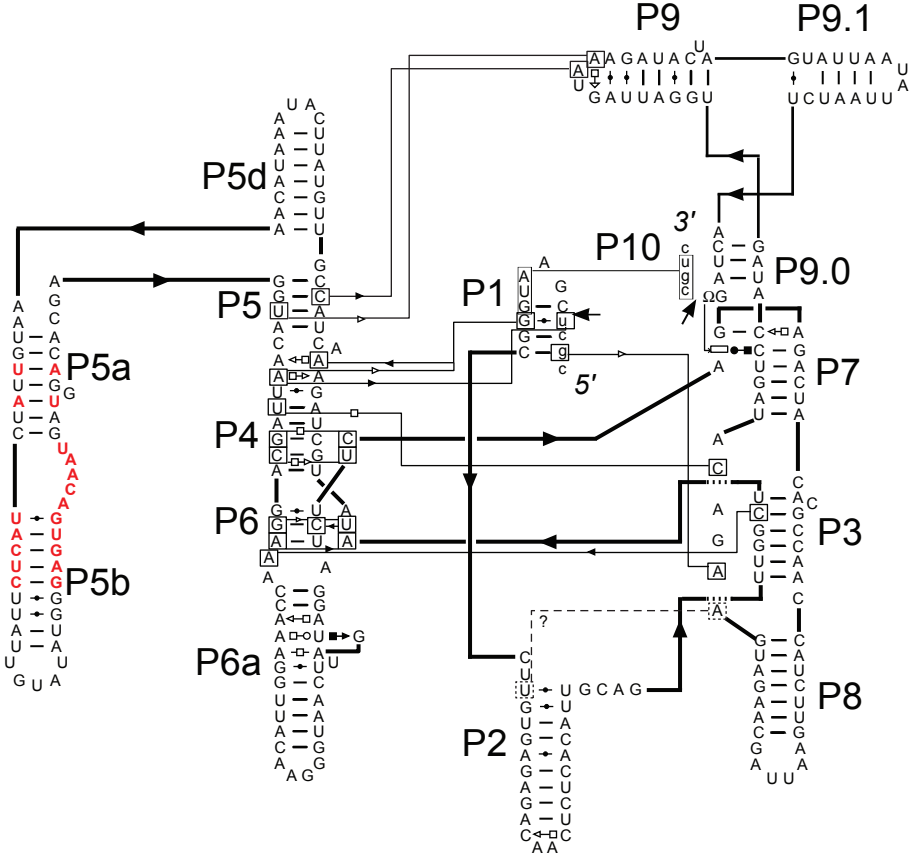
Self-splicing activity = 30%

Figure S1 cont.



IA3 *Chlamydomonas nivalis (C.n.)*

Self-splicing activity = 24%



IB1 *Amoebidium parasiticum (A.p.SSU)*

Self-splicing activity = 10%

Figure S1

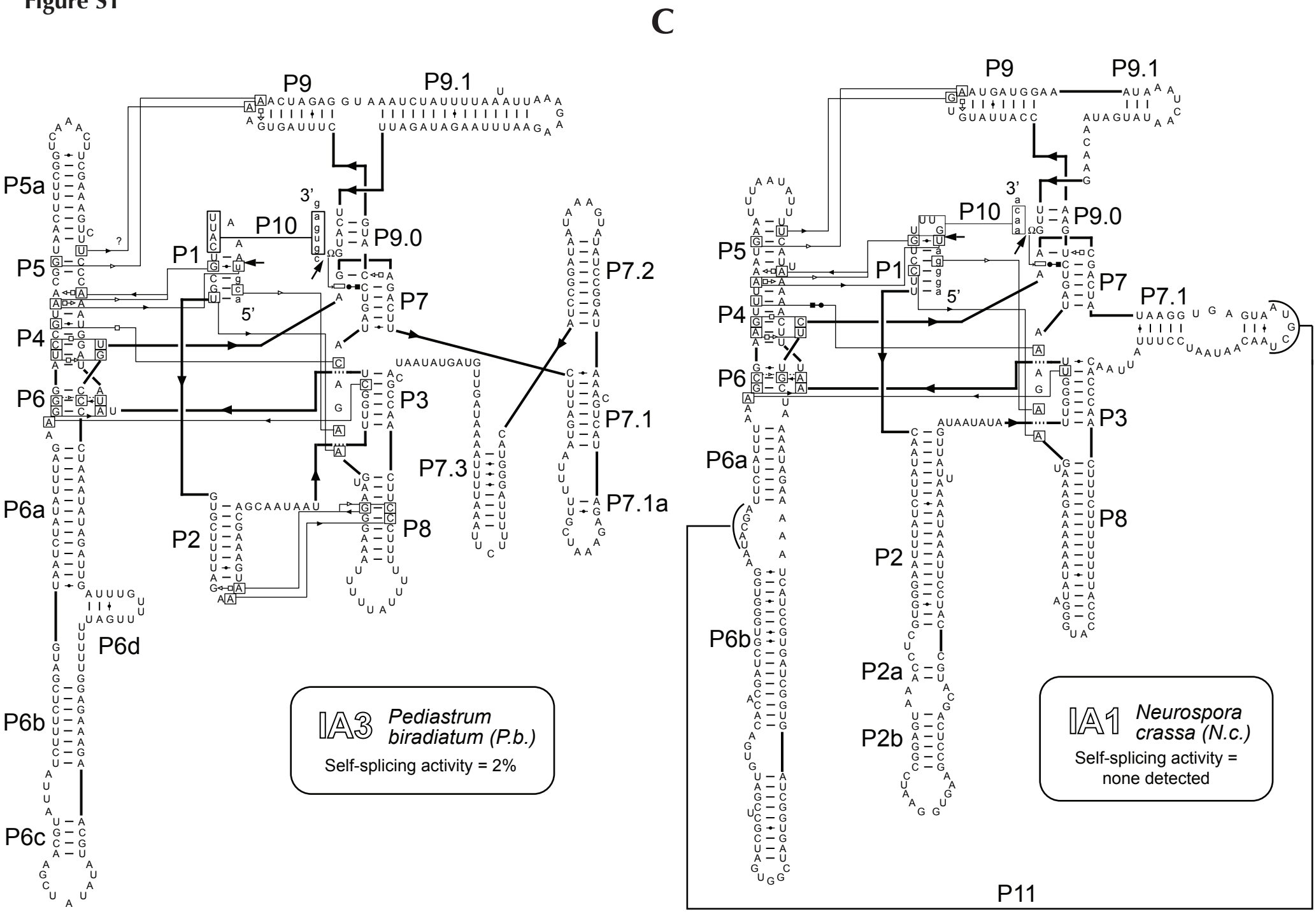
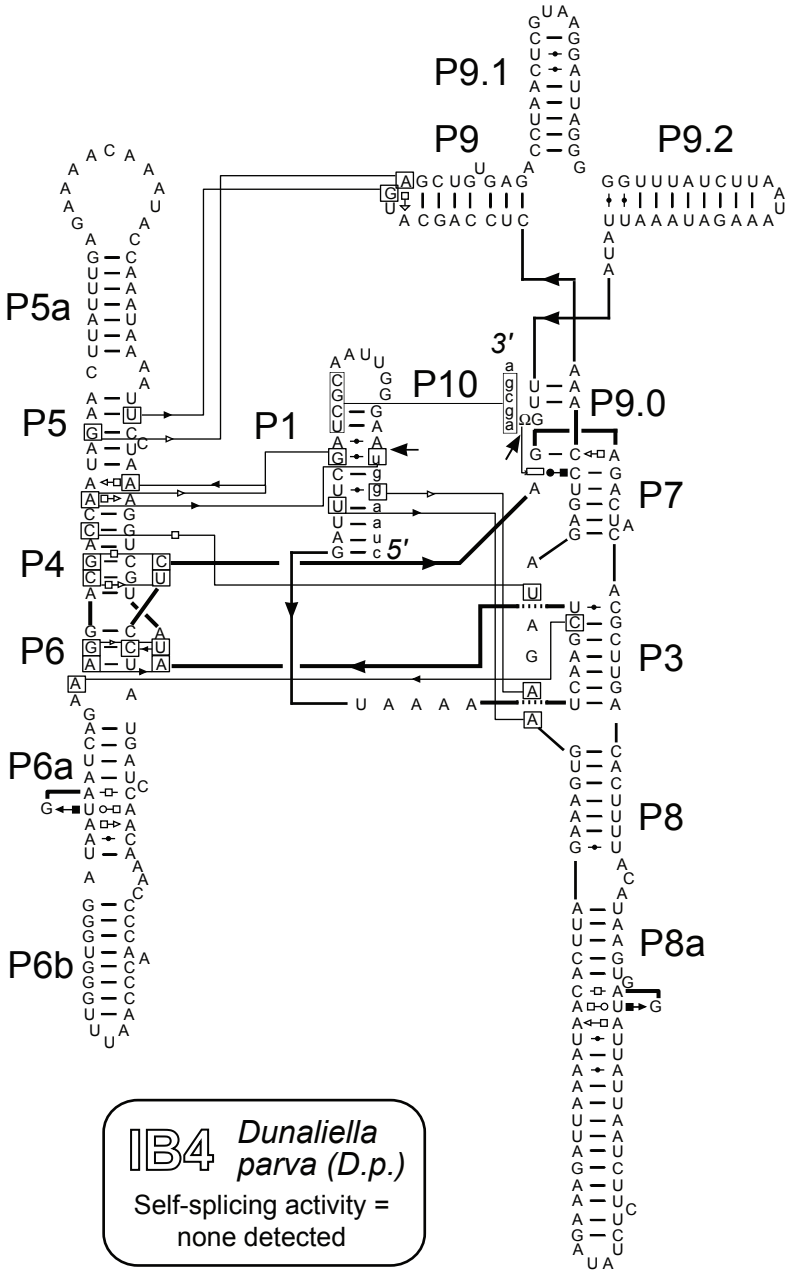

C
P9
P9.1
P5a
P10
P9.0
P1
P5
P7
P7.2
P4
P6
P3
P7.1
P7.3
P7.1a
P6a
P2
P8
P6d
P6b
P6c
IA3 Pediastrum biradiatum (P.b.)
Self-splicing activity = 2%
P5
P1
P10
P9.0
P7
P7.1
P4
P6
P3
P6a
P2
P8
P6b
P2a
P2b
P11
IA1 Neurospora crassa (N.c.)
Self-splicing activity = none detected

Figure S1 cont.



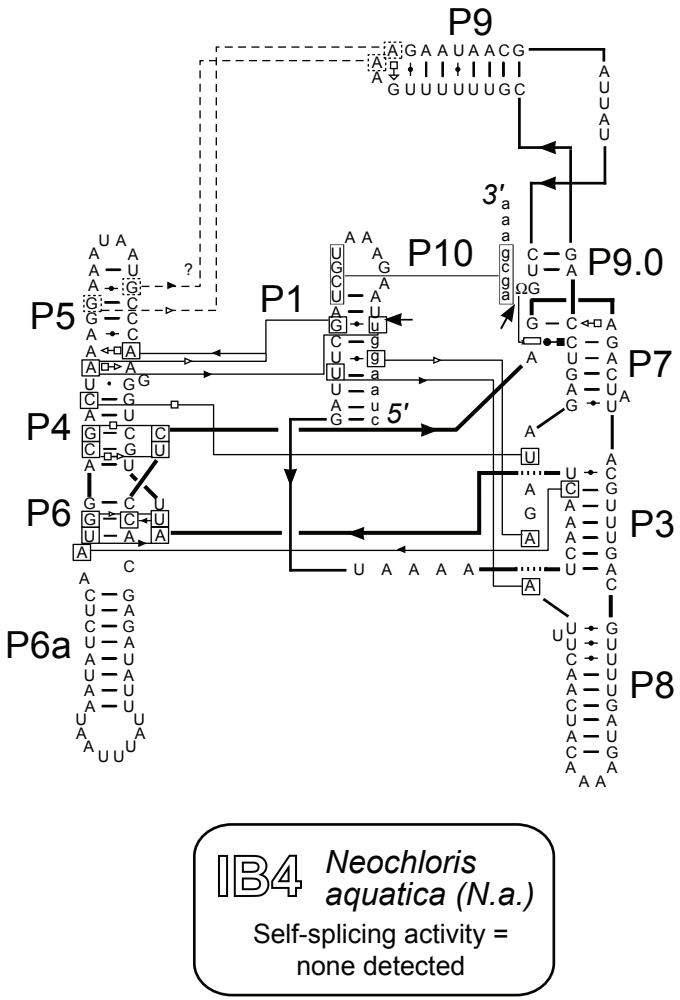


Figure S2

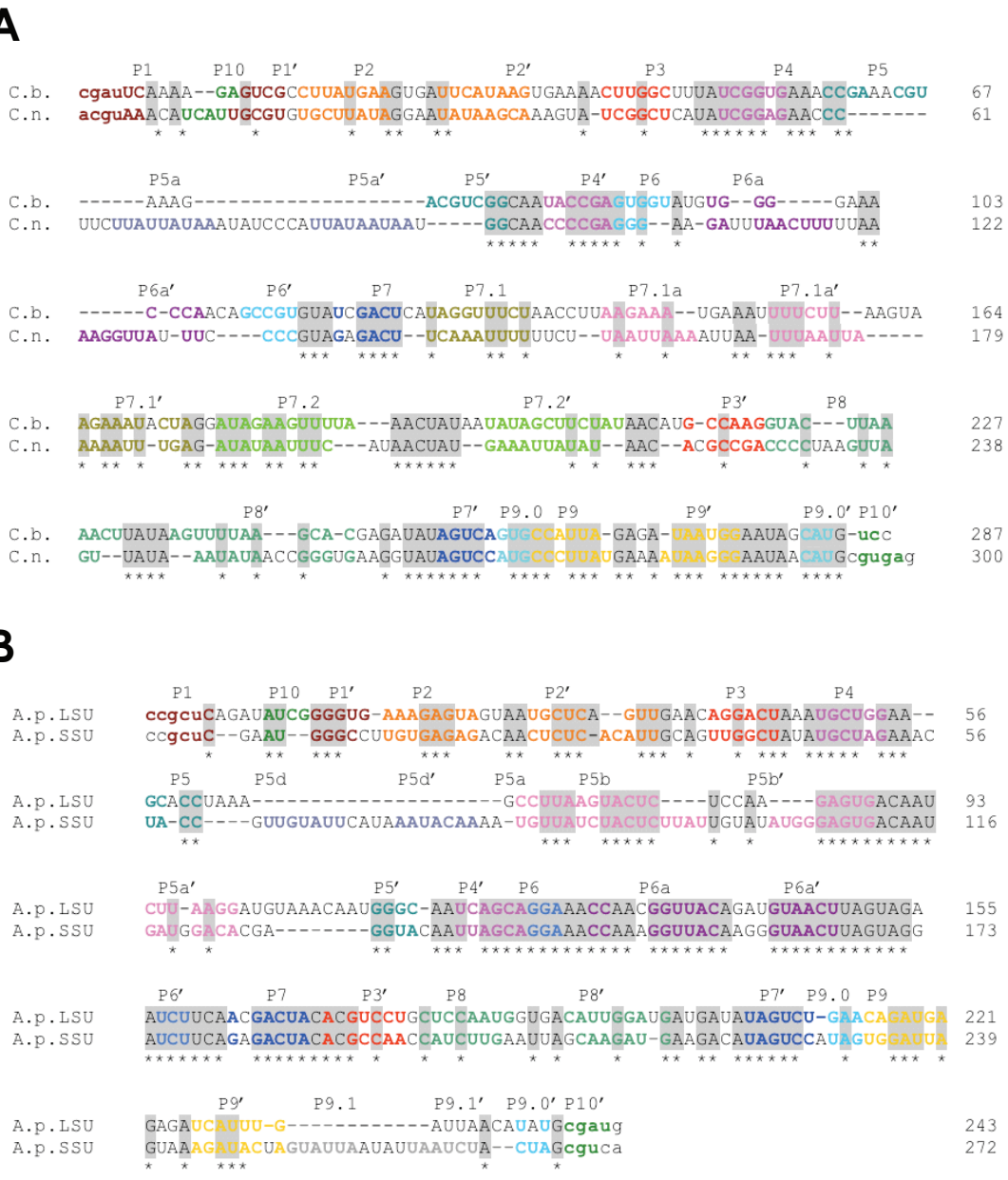


Table S1 **Statistics compiled for 38 group I introns (incl. 26 from the literature).**

Organism name	Intron location, gene type[1] (position[2]), subgroup[3]	Self-splicing activity[4]	Intron length (nt)	GC content of intron (%)	GC content of the genome or of the coding sequences(%)	Difference in GC content (Intron – Genome) (%)	Reference for the assays of intron self-splicing activity	Reference for the genome or the coding sequences	Number of codons
Introns that are self-splicing									
Enterobacteriophage T4	phage, td (n.d.), IA2	Y	1017	43	35	8	Chu FK et al., Cell (1986), 45:157-166	Codon Usage Database at http://www.kazusa.or.jp/codon/	114247
Bacteriophage Twort	phage, orf142-I2, IA2	Y	242	37	31	6	Landthaler M & Shub DA, Proc. Natl. Acad. Sci. USA (1999), 96:7005-7010	Codon Usage Database at http://www.kazusa.or.jp/codon/	41431
Bacteriophage SPO1	phage, dn (n.d.), IA2	Y	883	42	41	1	Goodrich-Blair H et al., Cell (1990), 63:417-424	Codon Usage Database at http://www.kazusa.or.jp/codon/	6007
Bacteriophage S3b	phage, orf288 (617), IA2	Y	1013	41	43	-2	Foley S et al., J. Virol. (2000), 74: 611-618	Foley S et al., J. Virol. (2000), 74: 611-618	n.a.
Bacillus anthracis	nucleus, recA (n.d.), IA2	Y	327	41	36	5	Ko M et al., J. Bacteriol. (2002), 184:3917-3922	Codon Usage Database at http://www.kazusa.or.jp/codon/	1551935
Clostridium botulinum	nucleus, tmrna (338), IA3	Y	287	36	29	7	this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	1010741
Tetrahymena thermophila	nucleus, rnl (1925), IC1	Y	413	44	22	22	Kruger K et al., Cell (1982), 31:147-157	Eisen JA et al., PLoS Biol. (2006), 4:1620-1642	145922
Dunaliella parva	nucleus, rns (943), IC1	Y	381	48	55–58 for related Dunaliella	(-7)–(-10)	Wilcox LW et al., Mol. Biol. Evol. (1992), 9:1103-1118	Codon Usage Database at http://www.kazusa.or.jp/codon/	n.a.
Dunaliella parva	nucleus, rns (1512), IC1	Y	421	52	55–58 for related Dunaliella	(-3)–(-5)	Wilcox LW et al., Mol. Biol. Evol. (1992), 9:1103-1118	Codon Usage Database at http://www.kazusa.or.jp/codon/	n.a.
Pneumocystis carinii	nucleus, rns (1506), IC1	Y	390	42	35	7	Sogin ML & Edman JC, NAR (1989), 17:5349-5359	Codon Usage Database at http://www.kazusa.or.jp/codon/	113187
Physarum polycephalum	nucleus, rnl (1925), IC1	Y	941	54	50	4	Ruoff B et al., NAR (1992), 20:5899-5906	Codon Usage Database at http://www.kazusa.or.jp/codon/	44747
Azoarcus sp. BH72	nucleus, trnI -CAU- (36), IC3	Y	205	60	68	-8	Reinhold-Hurek B & Shub DA, Nature (1992), 357:173-176	Codon Usage Database at http://www.kazusa.or.jp/codon/	1337417
Anabaena sp. PCC7120	nucleus, trnL -UAA- (34), IC3	Y	249	43	41	2	Xu et al., Science (1990), 250: 1566-1570; this work	Kaneko T et al., DNA Res. (2001), 8:205-213	n.a.
Scytonema hofmanni	nucleus, trnM (34), IC3	Y	244	45	47	-2	this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	497
Agrobacterium tumefaciens	nucleus, trnR -CCU- (36), IC3	Y	237	58	58	0	Reinhold-Hurek B & Shub DA, Nature (1992), 357:173-176; this work	Wood DW et al., Science (2001), 294:2317-2323	n.a.
Synechocystis PCC 6803	nucleus, trnM (34), IC3	Y	655	46	48	-2	Bonocora MM & Shub DA, Mol. Microbiol. (2001), 39:1299-1306	Codon Usage Database at http://www.kazusa.or.jp/codon/	1161949
Synechococcus elongatus PCC 6301	nucleus, trnL -UAA- (34), IC3	Y	240	47	56	-9	Sugita M et al., DNA Res. (1995), 2:71-76	Codon Usage Database at http://www.kazusa.or.jp/codon/	799138
Neurospora crassa	mitochondrion, CB (n.d.), ID	Y	1251	29	35	-6	Wallweber GJ et al., RNA (1997), 3:114-131	Codon Usage Database at http://www.kazusa.or.jp/codon/	13464
Candida albicans	nucleus, rnl (1923), IE2	Y	379	57	36	21	Mercure S et al., NAR (1993), 21:6020-6027	Codon Usage Database at http://www.kazusa.or.jp/codon/	635617
Scytalidium dimidiatum	nucleus, rns (1199), IE3	Y	393	68	54-58 for related Scytalidium	10–14	this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	n.a.
Introns that are poorly self-splicing									
Saccharomyces cerevisiae	mitochondrion, cb-bI5, IA1	P	738	16	38	-22	Gampel A & Tzagoloff A, Mol. Cell. Biol. (1987), 7:2545-2551; Gampel A & Cech TR, Gen. Dev. (1991), 5:1870-1880	Wood V et al., Nature (2002), 415:871-880	n.a.
Pediastrum biradiatum	chloroplast, rnl (2593), IA3	P	421	30	n.a.	n.a.	this work	n.a.	n.a.
Chlamydomonas nivalis	chloroplast, rnl (2593), IA3	P	300	31	30–45 for related Chlamydomonas	(-14)–1	this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	n.a.
Chlorella vulgaris	chloroplast, rnl (2593), IA3	P	815	37	35	2	Kapoor M et al., Curr. Genet. (1997), 31:503-510	Codon Usage Database at http://www.kazusa.or.jp/codon/	30087
Amoebidium parasiticum	mitochondrion, rnl (2500), IB1	P	243	41	35	6	this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	4842
Amoebidium parasiticum	mitochondrion, rns (1403), IB1	P	272	37	35	2	this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	4842
Introns that are not self-splicing									
Neurospora crassa	mitochondrion, rnl (2449), IA1	N	388	33	35	-2	Garriga G et al., Cell (1984), 38:631-641; this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	13464

Neurospora crassa	mitochondrion, cb (n.d.), IA1	N	1276	25	35	-10	Wallweber GJ et al., RNA (1997), 3:114-131	Codon Usage Database at http://www.kazusa.or.jp/codon/	13464
Emericella nidulans	mitochondrion, rnl (2449), IA1	N	1678	17	30	-13	Hur M et al. Curr. Genet. (1997), 32:399-407	Codon Usage Database at http://www.kazusa.or.jp/codon/	7912
Saccharomyces cerevisiae	mitochondrion, rnl (2449), IA1	N	1143	20	38	-18	Tabak HF et al., Cell (1984), 39:623-629	Wood V et al., Nature (2002), 415:871-880	n.a.
Schizosaccharomyces pombe	mitochondrion, ox1 (n.d.), IB1	N	258	26	36	-10	Schäfer B et al., Mol. Gen. Genet. (1991), 225:159-167	Wood V et al., Nature (2002), 415:871-880	n.a.
Neurospora crassa	mitochondrion, ND1 (n.d.), IB2	N	1118	27	35	-8	Wallweber GJ et al., RNA (1997), 3:114-131	Codon Usage Database at http://www.kazusa.or.jp/codon/	13464
Schizosaccharomyces pombe	mitochondrion, ox1 (n.d.), IB2	N	421	28	36	-8	Schäfer B et al., Mol. Gen. Genet. (1991), 225:159-167	Wood V et al., Nature (2002), 415:871-880	n.a.
Schizosaccharomyces pombe	mitochondrion, ox1 (n.d.), IB4	N	1081	25	36	-11	Schäfer B et al., Mol. Gen. Genet. (1991), 225:159-167	Wood V et al., Nature (2002), 415:871-880	n.a.
Neochloris aquatica	chloroplast, rnl (1931), IB4	N	183	32	33	-1	this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	223
Dunaliella parva	chloroplast, rnl (1931), IB4	N	318	34	32	2	this work	Codon Usage Database at http://www.kazusa.or.jp/codon/	514
Symkania negevensis	nucleus, rnl (n.d.), IB4	N	658	36	40 for related Chlamydiales	-4	Everett KD et al., J. Bacteriol. (1999), 181:4734-4740	Horn M et al., Science (2004), 304:728-730	n.a.
Emericella nidulans	mitochondrion, CB (n.d.), IB4	N	1057	26	30	-4	Hur M et al. Curr. Genet. (1997), 32:399-407	Codon Usage Database at http://www.kazusa.or.jp/codon/	7912

Footnotes:

1 *rnl:* large subunit rRNA; *rns:* small subunit rRNA; *tmrna:* transfer-messenger RNA; *trnL* (anticodon): $tRNA^{Leu}$; *trnM:* $tRNA^{Met}$; *trnR* (anticodon): $tRNA^{Arg}$; nd1: NADH dehydrogenase subunit 1; cb: cytochrome ; dn: DNA polymerase; ox1: cytochrome oxidase subunit 1

2 Nucleotide number in the corresponding *Escherichia coli* small or large subunit rRNA molecule or in the host tRNA/tmRNA, except otherwise stated (n.d.: not determined)

3 Based on the classification by Michel & Westhof (1990), expanded by Suh (1999) and Li & Zhang (2005)

4 N: no activity (in the absence of any protein facilitator); P: poor activity; Y: self-splicing activity

n.a. not available

Table S2 **Long range interactions observed in the 12 introns surveyed.**

	12 bp between G.U and L2	L9—P5		L2—P8		P4 bp2,3—J6/7 1,2	P6 bp1,2—J3/4 2,3	J4/5—P1	J6/6a 1—P3 bp2	P1 bp - 3,4—J8/7 1/2,2/3	P4 bp5—J8/7 5/6	Other motifs and interactions	Particular observations
		L9—P5bp2, 3	L9—P5 receptor	L2—P8 bp4,5	L2—P8 receptor								
A.s.	Y	—	GAAA—11nt	GAAA—CC:GG	—	U:A—C; G:C—G	C:G—A; C:G—A	A:A—G; A:A—U	A—C:G	G:C—A; C:G—A	C:G—G	related sarcin/ricin motif in P6a	—
S.d.	n.a.	GCGA—CA:UG	—	n.a.	n.a.	C:G—U; G:C—G	G:U—G; C:G—G	G:A—G; A:A—U	n.a.	n.a.	C:G—U	P13; P3—P2.1—P6 triple helix interaction	Extended P2, P9 insertions
C.b.	Y	GAGA—GA:UC	—	GUGA—CU:AG	—	C:G—U; G:C—G	U:G—U; G:C—U	A:A—G; A:A—U	A—G:C	C:G—G; G:C—A	A:U—A	related sarcin/ricin motif in P7.1	Extended P7 insertion
S.h.	Y	GAAA—AA:GC (?)	—	GAAA—CC:GG	—	C:G—U; G:C—G	C:G—A; C:G—A	A:A—G; A:A—U	A—C:G	G:U—A; G:U—A	C:G—G	n.d.	Extended P1a
A.t.	Y	—	GAAA—11nt	—	GAAA—11nt	C:G—U; G:C—G	C:G—A; C:G—A	A:A—G; A:A—U	A—C:G	G:C—A; G:C—A	C:G—G	n.d.	—
A.p.LSU	Y	—	GAGA—11nt	GUAA—CA:UG	—	G:C—C; C:G—U	G:C—A; A:U—A	A:A—G; A:A—U	A—C:G	U:G—U; G:C—A	C:G—U	related sarcin/ricin motif in P6a	Extended P5ab
C.n.	Y	GAAA—CU:AG (?)	—	GGAA—CU:CG (?)	—	C:G—U; G:C—G	G:C—U; G:C—A	A:A—G; A:A—U	A—C:G	G:C—G; U:A—A	C:A—A	n.d.	Extended P7 insertion
A.p.SSU	Y	GUAA—AC:GU	—	—	—	G:C—C; C:G—U	G:C—U; A:U—A	A:A—G; A:A—U	A—C:G	C:G—A; ?	U:A—C	related sarcin/ricin motif in P6a	Extended P5abd, P9
P.b.	Y	GAAA—CU:UG (?)	—	GAAA—CC:GG	—	C:G—U; U:A—G	G:C—U; G:C—A	A:A—G; A:A—U	A—C:G	C:G—A; U:A—A	G:U—C	n.d.	Extended P6, P7 insertions
N.c.	n.a.	GUGA—CU:AG	—	—	—	G:C—C; A:U—U	C:G—A; G:C—A	A:A—G; A:A—U	A—U:A	C:G—A; C:G—A	U:A—A	P11	Extended P2, P6
D.p.	N	AUGA—CU:AG	—	—	—	G:C—C; C:G—U	G:C—U; A:U—A	A:A—G; A:A—U	A—C:G	U:G—A; U:A—A	C:G—U	related sarcin/ricin motif in P6a, P8a	Extended P5a, P6, P8, P9 insertions
N.a.	N	GAAA—CG:AG (?)	—	—	—	G:C—C; C:G—U	G:C—U; U:A—A	A:A—G; A:A—U	A—C:G	U:G—A; U:A—A	C:G—U	n.d.	—
Reference		Michel & Westhof, 1990; Costa & Michel, 1995; Costa & Michel, 1997; Geary et al., 2008		Michel & Westhof, 1990; Costa & Michel, 1995; Costa & Michel, 1997; Geary et al., 2008		Flor et al. 1989; Green et al. 1990; Michel et al. 1990; Michel & Westhof, 1990		Strobel & Cech 1995; Strobel et al. 1998	Rangan et al. 2003	Adams et al. 2004; Guo et al. 2004; Golden et al. 2005; Vicens & Cech 2006	Tanner & Cech 1997	Leontis & Westhof 1998; Leontis et al 2002; Waldsich et al 2002; Li & Zhang 2005	

Table S3 Statistics compiled for the 14 fungal introns surveyed in Haugen et al. 2004

Organism	Intron length (nt)	GC content (%)	Self-splicing
Acaspora complanata	248	62	N
Ascolacicola austriaca	379	60	Y
Baeomyces placophyllus	333	58	N
Barrmaelia oxyacanthae	299	54	Y
Buellia georgei	213	70	N
Delitschia didyma	341	56	Y
Geosmithia virida	340	58	Y
Lecanora dispersa	213	66	N
Myriosclerotinia scirpicola	293	40	N
Penicillium oblatum	352	60	Y
Physcia stellaris	215	50	N
Pyrenula cruenta	302	56	N
Rhynchostoma minutum	356	not calculated	N
Rinodina cacuminum	208	61	N
Average GC content for self-splicing introns:		57.6	
Average GC content for non self-splicing introns		57.9	

Table S4 Additional informations and references for the introns used in this study.

Intron abbreviation	GenBank accession number	Provider	Reference	Nature of sample received from provider	Plasmid used in this study	Restriction enzymes used for cloning	Linearization enzyme	RNA polymerase
A.s.	AJ228705	n/a	Xu et al. 1990; Zaug et al. 1993	Plasmid containing intron and complete exons (pAtRNA-1)	pAtRNA-1	n/a	Ear I	T7
S.d.	AF258603	Marie Machouart (CHU Brabois, Vandoeuvre-les-Nancy, France) & Frederic Lorenzo (Université Paris VII, Paris, France)	Machouart-Dubach et al. 2002	PCR amplicon containing intron and portions of the exons	subcloned into pUC19	Eco RI / Bam HI	Ear I	T7
C.b.	L49156	Claude Lemieux (Université de Laval, Québec, Canada)	Pombert et al. 2006	PCR amplicon containing intron	subcloned into pUC19	Eco RI / Bam HI	Fok I	T7
S.h.	U10481	David Shub (University at Albany, State University of New York, Albany, NY)	Biniszkiewicz et al. 1994; Paquin et al. 1997	Plasmid containing intron and complete exons (pSH6)	subcloned into pUC19	Bam HI/Pst I	Bsa I	T7
A.t.	X66220	David Shub (University at Albany, State University of New York, Albany, NY)	Reinhold-Hurek et al. 1992; Paquin et al. 1999	Plasmid containing intron and complete exons (pBGA12.17)	subcloned into pUC19	Bam HI/Pst I	Bsa I	T7
A.p.LSU	AF538044	Gertraud Burger (Université de Montréal, Québec, Canada)	Lang et al. 2002; Burger et al. 2003	cDNA of intron and complete exons cloned into pFBSV	subcloned into pUC19	Eco RI / Bam HI	Fok I	T7
C.n.	L42990	Claude Lemieux (Université de Laval, Québec, Canada)	Pombert et al. 2006	PCR amplicon containing intron	subcloned into pUC19	Eco RI / Bam HI	Fok I	T7
A.p.SSU	AF538042	Gertraud Burger (Université de Montréal, Québec, Canada)	Lang et al. 2002; Burger et al. 2003	cDNA of intron and complete exons cloned into pFBSV	subcloned into pUC19	Eco RI / Bam HI	Ear I	T7
P.b.	AM412317	Kelly Williams (Virgina Bioinformatics Institute, Blacksburg, VA)	Williams 2002	Plasmid containing intron and complete exons (p237AK)	p237AK	n/a	Eco RI	T7
N.c.	J01427	n/a	Garriga and Lambowitz 1984; Mohr et al. 2001	Plasmid containing intron and portions of the exons (pBD5a)	pBD5a	n/a	Ban I	T3
D.p.	L43540	Claude Lemieux (Université de Laval, Québec, Canada)	Pombert et al. 2006	PCR amplicon containing intron	subcloned into pUC19	Eco RI / Bam HI	Fok I	T7
N.a.	L49155	Claude Lemieux (Université de Laval, Québec, Canada)	Pombert et al. 2006	PCR amplicon containing intron	subcloned into pUC19	Eco RI / Bam HI	Fok I	T7

Supplementary references

Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., Strobel, S. A. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature*. **430**:45-50.

Biniszkiewicz, D., Cesnaviciene, E. & Shub, D. A. 1994. Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. *EMBO J*. **13**: 4629-4635.

Burger, G., Forget, L., Zhu, Y., Gray, M. W. & Lang, B. F. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc. Natl. Acad. Sci. USA*. **100**: 892-897.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., Doudna, J. A. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*. **273**:1678-1684.

Doherty, E. A., Batey, R. T., Masquida, B., Doudna, J. A. 2000. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**:339-343.

Costa, M., Michel, F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J*. **14**:1276-1285.

Costa, M., Michel, F. 1997. Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution. *EMBO J*. **16**:3289-3302.

Flor, P. J., Flanegan, J. B., Cech, T. R. 1989. A conserved base pair within helix P4 of the *Tetrahymena* ribozyme helps to form the tertiary structure required for self-splicing. *EMBO J*. **8**:3391-3399.

Garriga, G. & Lambowitz, A. M. 1984. RNA splicing in neurospora mitochondria: self-splicing of a mitochondrial intron *in vitro*. *Cell*. **39**: 631-641.

Geary, C., Baudrey, S., Jaeger, L. 2008. Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res*. **36**:1138-1152.

Golden, B. L., Kim, H., Chase, E. 2005. Crystal structure of a phage Twort group I ribozyme-product complex. *Nat. Struct. Mol. Biol.* **12**:82-89.

Green, R., Ellington, A. D., Szostak, J. W. 1990. In vitro genetic analysis of the Tetrahymena self-splicing intron. *Nature*. **347**:406-408.

Guo, F., Gooding, A. R., Cech, T. R. 2004. Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol. Cell*. **16**:351-362.

Guo, F., Gooding, A. R., Cech, T. R. 2006. Comparison of crystal structure interactions and thermodynamics for stabilizing mutations in the Tetrahymena ribozyme. *RNA*. **12**:387-395.

Gutell, R. R., Cannone, J. J., Shang, Z., Du, Y., Serra, M. J. 2000. A story: unpaired adenosine bases in ribosomal RNAs. *J. Mol. Biol.* **304**:335-354.

Haugen, P., Runge, H. J. & Bhattacharya, D. 2004. Long-term evolution of the S788 fungal nuclear small subunit rRNA group I introns. *RNA*. **10**: 1084-1096.

Lang, B. F., O'Kelly, C., Nerad, T., Gray, M. W. & Burger, G. 2002. The closest unicellular relatives of animals. *Curr. Biol.* **12**: 1773-1778.

Leontis, N. B. & Westhof, E. 1998. A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. Mol. Biol.* **283**: 571-583.

Leontis, N. B., Stombaugh, J., Westhof, E. 2002. Motif prediction in ribosomal RNAs Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie.* **84**:961-973.

Lescoute, A., Westhof, E. 2006. The A-minor motifs in the decoding recognition process. *Biochimie.* **88**:993-999.

Li, Z., Zhang, Y. 2005. Predicting the secondary structures and tertiary interactions of 211 group I introns in IE subgroup. *Nucleic Acids Res.* **33**:2118-2128.

Machouart-Dubach, M., Lacroix, C., Vaury, C., Feuilhade de Chauvin, M., Bellanne, C., Derouin, F. & Lorenzo, F. 2002. Nucleotide structure of the *Scytalidium hyalinum* and *Scytalidium dimidiatum* 18S subunit ribosomal RNA gene: evidence for the insertion of a group IE intron in the rDNA gene of S. dimidiatum. *FEMS Microbiol. Lett.* **208**: 187-196.

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.

Mohr, G., Rennard, R., Cherniack, A. D., Stryker, J. & Lambowitz, A. M. 2001. Function of the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase in RNA splicing. Role of the idiosyncratic N-terminal extension and different modes of interaction with different group I introns. *J. Mol. Biol.* **307**: 75-92.

Mohr, G., Caprara, M. G., Guo, Q., Lambowitz, A. M. 1994. A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the *Tetrahymena* ribozyme. *Nature.* **370**:147-150.

Pan, J., Woodson, S. A. 1998. Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J. Mol. Biol.* **280**:597-609.

Pan, J., Woodson, S. A. 1999. The effect of long-range loop-loop interactions on folding of the *Tetrahymena* self-splicing RNA. *J. Mol. Biol.* **294**:955-965.

Paquin, B., Kathe, S. D., Nierzwicki-Bauer, S. A. & Shub, D. A. 1997. Origin and evolution of group I introns in cyanobacterial tRNA genes. *J. Bacteriol.* **179**: 6798-6806.

Paquin, B., Heinfling, A. & Shub, D. A. 1999. Sporadic distribution of $tRNA^{Arg}_{CCU}$ introns among alpha-purple bacteria: evidence for horizontal transmission and transposition of a group I intron. *J. Bacteriol.* **181**: 1049-1053.

Paukstelis, P. J., Chen, J. H., Chase, E., Lambowitz, A. M., Golden, B. L. 2008. Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature.* **451**:94-97.

Pombert, J. F., Lemieux, C. & Turmel, M. 2006. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biol.* **4**: 3.

Rangan, P., Masquida, B., Westhof, E., Woodson, S. A. 2003. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. USA* **100**:1574-1579.

Reinhold-Hurek, B. & Shub, D. A. 1992. Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature*. **357**: 173-176.

Strobel, S. A., Cech, T. R. 1995. Minor groove recognition of the conserved G.U pair at the *Tetrahymena* ribozyme reaction site. *Science* **267**:675-679.

Strobel, S. A., Ortoleva-Donnelly, L., Ryder, S. P., Cate, J. H., Moncoeur, E. 1998. Complementary sets of non-canonical base pairs mediate RNA helix packing in the group I intron active site. *Nature Struct. Biol.* **5**:60-65.

Tanner, M. A., Cech, T. R. 1997. Joining the two domains of a group I ribozyme to form the catalytic core. *Science* **275**:847-849.

Vicens, Q., Cech, T. R. 2006. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **31**:41-51.

Waldsich, C., Masquida, B., Westhof, E., Schroeder, R. 2002. Monitoring intermediate folding states of the *td* group I intron in vivo. *EMBO J*. **21**:5281-5291.

Williams, K. P. 2002. The tmRNA Website: invasion by an intron. *Nucleic Acids Res*. **30**: 179-182.

Woodson, S. A., Rangan, P. 2008. Folding mechanisms of group I ribozymes. In: Lilley, D. M., Eckstein, F., eds. *Ribozymes and RNA catalysis*. Cambridge, UK: RSC Publishing. pp 295-314.

Xu, M. Q., Kathe, S. D., Goodrich-Blair, H., Nierzwicki-Bauer, S. A. & Shub, D. A. 1990. Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science*. **250**: 1566-1570.

Zaug, A. J., McEvoy, M. M. & Cech, T. R. 1993. Self-splicing of the group I intron from Anabaena pre-tRNA: requirement for base-pairing of the exons in the anticodon stem. *Biochemistry*. **32**: 7946-7953.