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

Quentin Vicens, Paul J. Paukstelis, Eric Westhof, et al.

RNA 2008 14: 2013-2029 originally published online September 3, 2008
Access the most recent version at doi:10.1261/rna.1027208

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

QUENTIN VICENS,[1] PAUL J. PAUKSTELIS,[2] ERIC WESTHOF,[3] ALAN M. LAMBOWITZ,[2] and THOMAS R. CECH[1]

[1]Howard Hughes Medical Institute, University of Colorado, Department of Chemistry and Biochemistry, Boulder, Colorado 80309-0215, USA
[2]Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, and Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas at Austin, Austin, Texas 78712, USA
[3]Institut de Biologie Moléculaire et Cellulaire du CNRS, Université Louis Pasteur, F-67084 Strasbourg Cedex, France

ABSTRACT

In the current era of massive discoveries of noncoding RNAs within genomes, being able to infer a function from a nucleotide sequence is of paramount interest. Although studies of individual group I introns have identified self-splicing and nonself-splicing examples, there is no overall understanding of the prevalence of self-splicing or the factors that determine it among the >2300 group I introns sequenced to date. Here, the self-splicing activities of 12 group I introns from various organisms were assayed under six reaction conditions that had been shown previously to promote RNA catalysis for different RNAs. Besides revealing that assessing self-splicing under only one condition can be misleading, this survey emphasizes that in vitro self-splicing efficiency is correlated with the GC content of the intron (>35% GC was generally conductive to self-splicing), and with the ability of the introns to form particular tertiary interactions. Addition of the *Neurospora crassa* CYT-18 protein activated splicing of two nonself-splicing introns, but inhibited the second step of self-splicing for two others. Together, correlations between sequence, predicted structure and splicing begin to establish rules that should facilitate our ability to predict the self-splicing activity of any group I intron from its sequence.

Keywords: catalytic RNA; group I intron; protein-facilitated splicing; ribozyme; self-splicing

INTRODUCTION

Since the discovery of self-splicing of the *Tetrahymena thermophila* rRNA intron 25 yr ago (Kruger et al. 1982), numerous group I introns have been identified in diverse eucaryotes and bacteria (Cannone et al. 2002). Group I introns possess at their core a conserved architecture formed by 10 paired segments (P1–P10) organized in three domains: a structural domain (P4–P6) that folds independently in the *Tetrahymena* rRNA intron (Murphy and Cech 1993), a substrate domain (P1, P10), and a catalytic domain (P3, P7, and P8) (Michel and Westhof 1990; Golden et al. 1998). These introns are classified into 14 subgroups (IA1–3, IB1–4, IC1–3, ID, IE1–3) based on conservation of the core domains and on the presence of characteristic peripheral elements (e.g., P2, P7.1, P7.2, and P9) (Michel and Westhof 1990; Lehnert et al. 1996; Suh et al. 1999; Li and Zhang 2005). In general, the peripheral elements form long-range tertiary interactions that stabilize the core domains to promote catalysis (Jaeger et al. 1991; Costa and Michel 1995; Lehnert et al. 1996). For certain group I introns, protein cofactors like CYT-18 (a tyrosyl-tRNA synthetase from *Neurospora crassa* mitochondria) and CBP2 (the cytochrome *b* pre-mRNA processing protein 2 from yeast mitochondria) cooperate with and sometimes supersede peripheral elements in order to facilitate folding of the RNA into a catalytically active structure (Lambowitz and Perlman 1990; Mohr et al. 1994; Weeks and Cech 1996; Webb et al. 2001; Paukstelis et al. 2008). The crystal structures of three well-studied introns (Adams et al. 2004; Guo et al. 2004; Golden et al. 2005), one of them now also solved in complex with a splicing-competent C-terminally truncated CYT-18 protein (Paukstelis et al. 2008) further illuminate how diverse architectures can stabilize a conserved catalytic core (Vicens and Cech 2006).

Nevertheless, for most of the >2300 group I introns sequenced to date, little information is available regarding their activity in vitro or in vivo (Hammann and Westhof 2007; Lang et al. 2007). Moreover, self-splicing is typically

Reprint requests to: Quentin Vicens, Howard Hughes Medical Institute, University of Colorado, UCB 215, Department of Chemistry and Biochemistry, Boulder, CO 80309-0215, USA; e-mail: quentin.vicens@colorado.edu; fax: (303) 492-6194.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1027208.

assayed under one or only a limited number of conditions, which are often laboratory-specific. Whether different conditions would give the same diagnosis about the activity of a particular intron is often unclear. Here, 12 group I introns of various origins and chosen to represent different subgroups were assayed for self-splicing and CYT-18-facilitated splicing under a set of reaction conditions that had been previously shown to promote catalysis of different RNAs. First, the comparison of the various conditions tested indicated that approaches that consider only one condition to assess splicing can be flawed. Second, analysis of the newly tested and 26 additional introns from the literature reveals that a GC content greater than 35% is generally conducive to self-splicing. Third, the presence or absence of certain structural elements and long range interactions could account for the observed level of activity for most introns, although definite verdicts would require investigating a larger sample of introns at a finer detail. In contrast, predicting the requirement of a protein facilitator for activity is more difficult, as exemplified by the CYT-18-dependent splicing found for one intron that was unanticipated based on its secondary structure, and by the protein-dependent inhibition of self-splicing characterized for two introns.

RESULTS

The 12 group I intron RNAs tested in this study originated from divergent organisms (bacteria, fungi, plants, and protozoa), represented four of the five structural subgroups (IA, IB, IC, and IE), contained 30%–68% GC, and were 183–421 nucleotides long (Table 1). These introns were also embedded in different contexts of exons (tRNA, large ribosomal subunit, tmRNA), which were either kept at their natural lengths or shortened to 75 nucleotides (Table 1). All introns were tested under six different reaction conditions, most of which had been shown previously to promote catalysis of different RNAs. These included conditions that supported self-splicing of the *Anabaena* IC3 intron (Zaug et al. 1993) and the *Didymium iridis* IE1 intron (Johansen and Vogt 1994; Nielsen et al. 2003). The *Anabaena* conditions were expanded to include either a higher reaction temperature (42°C) or a higher ionic strength obtained by increasing the concentration of monovalent or divalent cations. Such modifications have the potential to help the splicing of structurally compromised introns (Joyce et al. 1989). Finally, a condition that enables processing of bacterial tRNA by RNase P was added (Baer et al. 1990; Warnecke et al. 2000). CYT-18 protein-facilitated splicing was tested under conditions used for splicing of the *Neurospora crassa* group IA1 intron from the mitochondrial large ribosomal subunit rRNA (LSU) gene (Guo et al. 1991) and at a lower monovalent salt concentration with two different ratios of monovalent to divalent ion concentration.

Three-fourths of the introns surveyed are self-splicing

Precursor RNAs were annealed to facilitate proper folding and then incubated for 1 h or 24 h under the six different reaction conditions described (Table 2) with an [α-^{32}P]GTP cofactor. Self-splicing was thus monitored by the covalent attachment of the labeled GTP to the 5′-end of either the excised intron or the splicing intermediate (intron + 3′ exon) (Fig. 1), as assessed by denaturing gel electrophoresis of the reaction products (Fig. 2). Hydrolysis reactions (Inoue et al. 1986) and subsequent circularizations (Zaug et al. 1983) were not monitored by our approach, as the exogenous G is not incorporated into these products. Other rare events such as GTP attack at the 3′ splice site (Inoue et al. 1986) would give smaller products than could be observed by our methods. In this study, strong self-splicing is defined as an activity no less than 75% of that obtained for the highly active *Anabaena* intron, after 1 h of incubation (as judged by yield of the products of the two steps of the self-splicing reaction). Overall, five introns were strongly self-splicing: the three tRNA introns (A.s., S.h., A.t.) self-spliced with a similar efficiency, while both the tmRNA intron (C.b.) and LSU intron (S.d.) self-spliced twice as efficiently as the tRNA introns (intron abbreviations are described in Table 1). Four introns were poorly self-splicing, being only 1%–30% as active as the A.s. intron after 24 h incubation under the best condition. Finally, three introns showed no detectable self-splicing (<<1%) under any condition tested (Table 1; Fig. 2).

The folding of each RNA into a homogeneous tertiary structure was assessed by native gel electrophoresis (Fig. 3). A single sharp band was taken to indicate a homogeneously folded molecule (e.g., C.n.), while broad and/or diffuse bands were taken as indicators of nonhomogeneous folding (e.g., D.p., P.b.). Based of these criteria, nonhomogeneous folding may account for the failure of some introns (e.g., D.p., P.b.), but not all (e.g., N.a., C.n.) to self-splice. Other reasons such as misfolding into a predominant non-native stable structure (as for the *Tetrahymena* rRNA intron [Emerick and Woodson 1993; Zarrinkar and Williamson 1994; Downs and Cech 1996]) or the requirement for a protein facilitator for final folding may explain the lack of efficient self-splicing observed in vitro.

Self-splicing correlates with a GC content higher than 35%

In this limited sample of 12 introns, GC content showed a better correlation with self-splicing activity than did intron length. For example, the 393 nt-long and GC-rich (68%) S.d. intron underwent self-splicing in five of six conditions, whereas the 421 nt-long and GC-poor (30%) P.b. intron as well as the 183 nt-long and GC-poor (32%) N.a. intron did not splice efficiently. For the 12 introns tested here, a minimum GC content of about 35% was correlated with

TABLE 1. Introns analyzed in this study

Organism name (gene location, gene type,[a] position[b])	Abbreviation used	Classification Kingdom/Phylum/Class	Subgroup[c]	RNA-only conditions[d] A	B	C	D	E	F	Diagnosis[e]	CYT-18 conditions[d] a	b	c	Diagnosis[f]	GC content (%)	Sizes (nt) 5′E–I–3′E	G–I–3′E	G–I	5′E–3′E
Anabaena PCC7120 (nucleus, *trnL* (UAA), pos. 34)	A.s.	Bacteria/Cyanobacteria/ Hormogoneae	IC3	++	++	++	++	++	++	Self-splicing (100)	+	++	++	Unaffected (0)	43	334	295	250	85
Scytalidium dimidiatum (nucleus, *rns*, pos. 1199*)[g]	S.d.	Fungi/Ascomycota/ Ascomycetes	IE3	++	+	++	+	++	–	Self-splicing (218)	++	++	+	Unaffected (0)	68	467	439	394	75
Clostridium botulinum (nucleus, *tmrna*, pos. 338)	C.b.	Bacteria/Firmicutes/ Clostridia	IA3	++	++	+	++	++	++	Self-splicing (200)	++	++	+	Unaffected (0)	36	649	312	288	362
Scytonema hofmanii (nucleus, *trnM*, pos. 34)	S.h.	Bacteria/Cyanobacteria/ Hormogoneae	IC3	++	++	++	++	++	–	Self-splicing (117)	i++	i++	i++	Inhibited (27)	45	318	285	245	74
Agrobacterium tumefaciens (nucleus, *trnR* (CCU), pos. 36)	A.t.	Bacteria/Proteobacteria/ Alphaproteobacteria	IC3	++	++	++	++	++	++	Self-splicing (94)	–	i+	i++	Inhibited (11)	58	314	278	238	77
Amoebidium parasiticum (mitochondrion, *rnl*, pos. 2500)	A.p.LSU	Protozoa/Choanozoa/ Mesomycetozoea	IB1	+	–	++	–	–	–	Poorly self-splicing (30)	–	f++	f+	Facilitated (25)	41	318	289	244	75
Chlamydomonas nivalis (chloroplast, *rnl*, 2593)	C.n.	Plantae/Chlorophyta/ Chlorophyceae	IA3	+	–	+	+	–	–	Poorly self-splicing (24)	–	–	–	Unaffected (0)	31	375	346	301	75
Amoebidium parasiticum (mitochondrion, *rns*, pos. 1403)	A.p.SSU	Protozoa/Choanozoa/ Mesomycetozoea	IB1	+	–	+	–	+	–	Poorly self-splicing (10)	–	–	–	Unaffected (0)	37	347	318	273	75
Pediastrum biradiatum (chloroplast, *rnl*, pos. 2593)	P.b.	Plantae/Chlorophyta/ Chlorophyceae	IA3	+	–	–	–	–	–	Poorly self-splicing (2)	–	–	–	Unaffected (0)	30	496	467	422	75
Neurospora crassa (mitochondrion, *rnl*, pos. 2449)	N.c.	Fungi/Ascomycota/ Ascomycetes	IA1	–	–	–	–	–	–	Not self-splicing (0)	f++	f++	f+	Facilitated (100)	33	503	439	389	115
Dunaliella parva (chloroplast, *rnl*, pos. 1931)	D.p.	Plantae/Prasinophyta/ Prasinophyceae	IB4	–	–	–	–	–	–	Not self-splicing (0)	–	f+	f+	Facilitated (0.3)	34	393	364	319	75
Neochloris aquatica (chloroplast, *rnl*, pos. 1931)	N.a.	Bacteria/Cyanobacteria/ Gdom	IB4	–	–	–	–	–	–	Not self-splicing (0)	–	–	–	Unaffected (0)	32	258	229	184	75

[a](*rnl*) large subunit rRNA; (*rns*) small subunit rRNA; (*tmrna*) transfer-messenger RNA; [*trnL* (anticodon)] $tRNA^{Leu}$; (*trnM*) $tRNA^{fMet}$; [*trnR* (anticodon)] $tRNA^{Arg}$.
[b]Nucleotide number in the corresponding *Escherichia coli* small or large subunit rRNA molecule or in the host tRNA/tmRNA, except otherwise stated.
[c]Based on the classification by Michel and Westhof (1990), expanded by Suh et al. (1999) and Li and Zhang (2005).
[d](++) Strong self-splicing; (+) weak self-splicing; (–) no self-splicing; (f) facilitated; (i) inhibited.
[e](Self-splicing) Self-splicing activity >75% of that observed for the A.s. intron; (poorly self-splicing) self-splicing comprised within 1%–30% of that observed for the A.s. intron; (not self-splicing) no self-splicing observed (activity << 1%). In parenthesis, percentage of activity (under the best condition) relative to that of the A.s. intron (condition A after 1 h of incubation).
[f](Unaffected) No protein-dependent enhancement greater than twofold of the self-splicing activity observed in absence of any protein; (facilitated) protein-dependent self-splicing >0.1% of that observed for the N.c. intron; (inhibited) decrease of self-splicing activity upon CYT-18 addition. In parenthesis, percentage of activity (under the best condition) relative to that of the N.c. intron (condition a after 1 h of incubation).
[g](*) *Scytalidum dimidiatum* small subunit rRNA numbering.

TABLE 2. Conditions used to screen for self-splicing and protein-facilitated splicing

	RNA-only conditions						CYT-18 conditions		
	A	B	C	D	E	F	a	b	c
Buffer composition									
Monovalent salt	25 mM NaCl	25 mM NaCl	1.0 M NaCl	25 mM NaCl	1.0 M NH_4OAc	0.5 M KCl	100 mM KCl	50 mM KCl	50 mM KCl
Divalent salt	15 mM $MgCl_2$	15 mM $MgCl_2$	15 mM $MgCl_2$	200 mM $MgCl_2$	100 mM $Mg(OAc)_2$	15 mM $MgCl_2$	5 mM $MgCl_2$	5 mM $MgCl_2$	30 mM $MgCl_2$
Buffer	25 mM HEPES (pH 7.5)	25 mM HEPES (pH 7.5)	25 mM HEPES (pH 7.5)	25 mM HEPES (pH 7.5)	25 mM HEPES (pH 7.5)	40 mM Tris-HCl (pH 7.5)	20 mM Tris-HCl (pH 7.5)	20 mM Tris-HCl (pH 7.5)	20 mM Tris-HCl (pH 7.5)
Additive	—	—	—	—	—	5 mM DTT; 2 mM spermine	5 mM DTT; 0.1 mg/mL BSA; 10% glycerol	5 mM DTT; 0.1 mg/mL BSA; 10% glycerol	5 mM DTT; 0.1 mg/mL BSA; 10% glycerol
Incubation temperature (°C)	32	42	32	32	37	50	37	37	37
Reference	Zaug et al. 1993	—	Jaeger et al. 1990; Zaug et al. 1993	Joyce et al. 1989; Zaug et al. 1993	Baer et al. 1990; Warnecke et al. 2000	Johansen and Vogt 1994; Nielsen et al. 2003	Guo et al. 1991	—	—

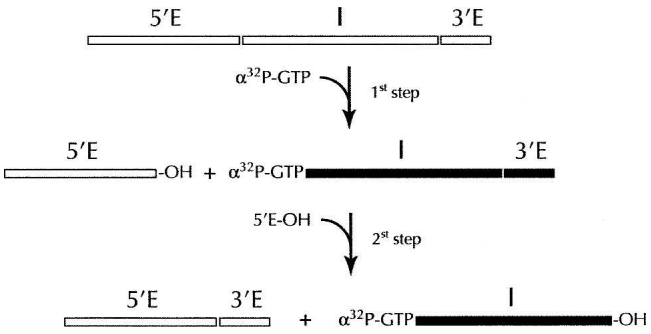

FIGURE 1. Mechanism of group I intron splicing. The intron-containing products that are labeled upon addition of [α-^{32}P]GTP to the reaction mix, and hence visible on polyacrylamide gels, are shown in black, while the unlabeled precursor RNA and the released exons are shown in white.

efficient self-splicing (Table 1). Data for 26 other group I introns described in the literature generally support this criterion and revealed a correlation to the difference between the GC content of an intron and that of its host genome (Supplemental material), as follows. All introns that were >35% GC rich and more GC-rich than the host genome were self-splicing (Fig. 4). However this double criterion misses 29% of the self-splicing introns, therefore it may not be as useful as the simple 35% GC rule. It would be worth analyzing more extensively the group I introns in the database for which self-splicing activity has been assayed in order to further refine this simple 35% GC criterion.

Introns that contain optimal long-range tertiary interactions are self-splicing

Long-range interactions that have been shown previously to be crucial for folding and activity of group I introns were modeled on the predicted secondary structure diagrams of the 12 introns surveyed here (Fig. 5; Supplemental material). These included tertiary contacts within the core, for example, involving the P4, P6, and P1 helices and various joining regions (Michel and Westhof 1990; Michel et al. 1990; Strobel and Cech 1995; Strobel et al. 1998). Tertiary interactions were also proposed at the periphery, for example, involving the tetraloop L9 and either two consecutive base pairs or an 11-nt receptor in P5 (an L9–P5 interaction is almost universally found in group I introns; Michel and Westhof 1990; Costa and Michel 1995; Hammann and Westhof 2007) or involving L2 and P8 in a similar interaction, which plays a critical role in properly positioning the P1–P2 substrate domain in the catalytic site (Michel and Westhof 1990; Costa and Michel 1995; Jaeger et al. 1996). Subgroup-specific interactions, such as the triple-helical interaction involving P6, P2.1, and P3 in subgroup IE introns (Li and Zhang 2005), were compiled as well. Additionally, the sarcin/ricin motif (Leontis and Westhof 1998; Leontis et al. 2002) was modeled in peripheral regions when the consensus sequence was detected (e.g., in P7.1 of C.b. and P6b of A.s.).

As anticipated from previous analyses (Costa and Michel 1995, 1997; Brion et al. 1999), self-splicing ability correlates with the type and the number of tertiary interactions present, as well as their presumptive stability. A type of L9–P5 long-range interaction (involving the L9 tetraloop and either adjacent base pairs or an 11-nt receptor) was predicted for eight introns (as well as for the remaining four, although with less confidence because of ambiguous pairings in P5). A related L2–P8 interaction was predicted for seven introns (Supplemental material). Remarkably, all introns that contain at least one tetraloop-receptor interaction (A.p.LSU, A.s., A.t.) or tetraloop-base pair interactions (C.b., S.h., C.n., A.p.SSU, P.b.) were self-splicing while the three introns that lack a P2 extension suitable for contacting P8 (N.c., D.p., and N.a.) were not (Fig. 5; Supplemental material). Moreover, as expected from earlier studies (Costa and Michel 1995, 1997), the more closely the sequence of these tetraloop-receptor motifs resembled the consensus sequence (e.g., GAAA–[CCUAAG/A...U/CAUGG] for the tetraloop–11 nt receptor interaction), the higher the level of self-splicing activity detected. For example, robustly self-splicing introns contained both consensus L9–P5 and L2–P8 interactions (A.t.), or just the consensus L9–P5 interaction (A.s.), while a poorly self-splicing intron contained sequences that diverge from the consensus at these positions (A.p.LSU) (Supplemental material). Finally, as expected from the classification of introns into different structural subgroups (Michel and Westhof 1990), this survey further illustrates how subgroup-specific interactions may also support catalysis. For example, the S.d. intron (which possesses a L9–P5 interaction but no L2–P8 interaction) is poised for catalysis by subgroup IE-specific interactions involving the P3, P2.1, and P6 helices, which all adopt the consensus sequence observed for these helices (Li and Zhang 2005). Overall, these results further support the notion that long range interactions between peripheral elements are key determinants of the ability of group I introns to self-splice. Such comparative analyses are beginning to lead to the emergence of patterns by which the autocatalytic activity of a group I intron can be predicted (see Discussion).

One-third of the introns surveyed are positively or negatively affected by the CYT-18 protein

The CYT-18 protein binds to and facilitates the folding of several *Neurospora crassa* mitochondrial group I introns, thereby allowing the RNA to carry out the splicing reaction (Collins and Lambowitz 1985; Akins and Lambowitz 1987; Wallweber et al. 1997). The CYT-18 protein can also stimulate the splicing of some group I introns from organisms that do not express a CYT-18 homolog (Chen

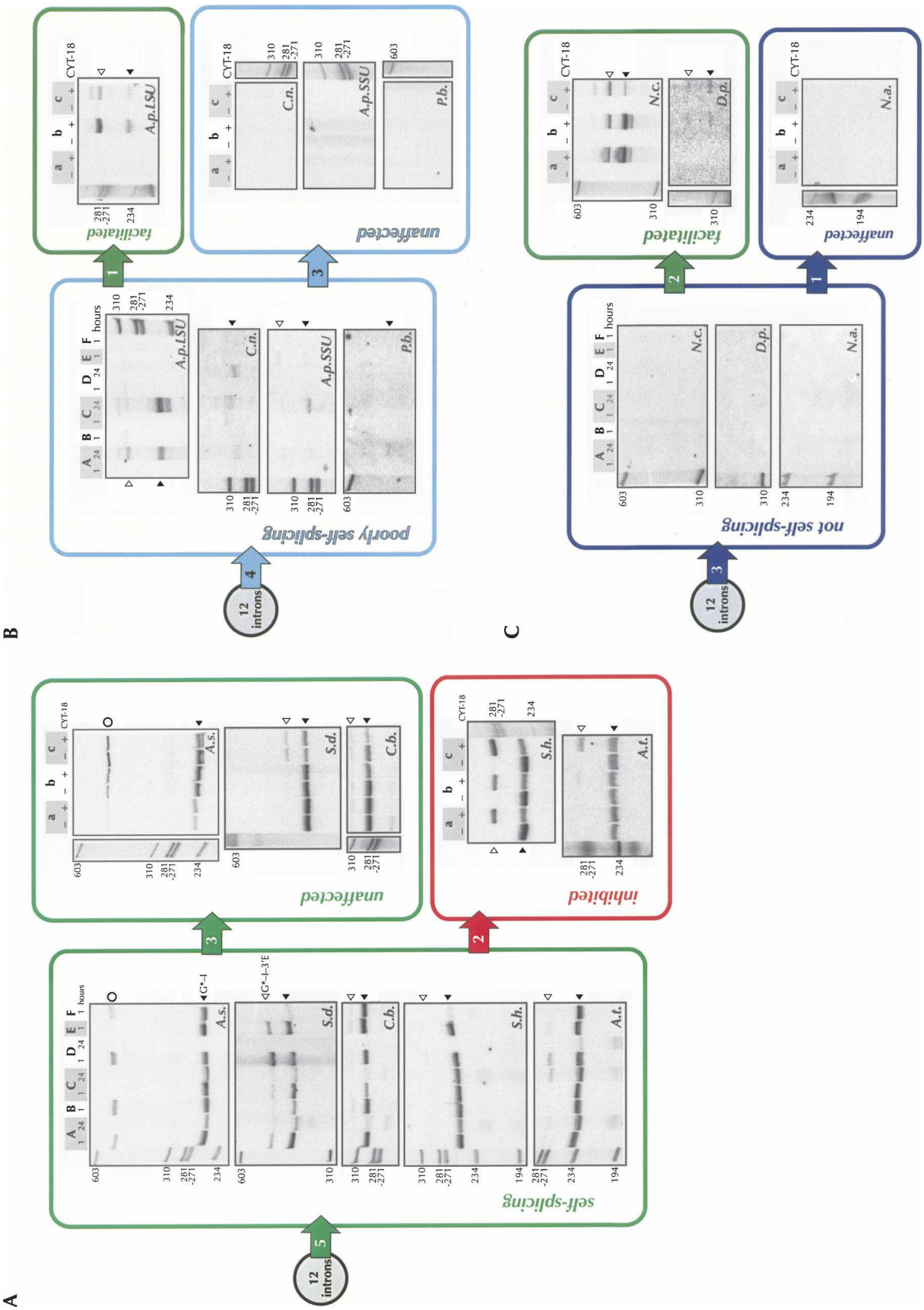

FIGURE 2. Self-splicing and protein-facilitated splicing of the 12 group I introns studied. (*A*) Results of the activity assays of the self-splicing introns (green). Introns that are inhibited upon CYT-18 binding are circled in red. (*B*) Results of the activity assays of the poorly self-splicing introns (cyan). (*C*) Results of the activity assays of the introns that are not self-splicing (dark blue). Intron abbreviations are defined in Table 1. Times of incubation are 1 h and 24 h. The numbers in the arrows indicate the total number of introns that followed each scenario. The previously uncharacterized circular product from the self-splicing reaction of the A.s. intron (indicated by an open circle) will be described elsewhere.

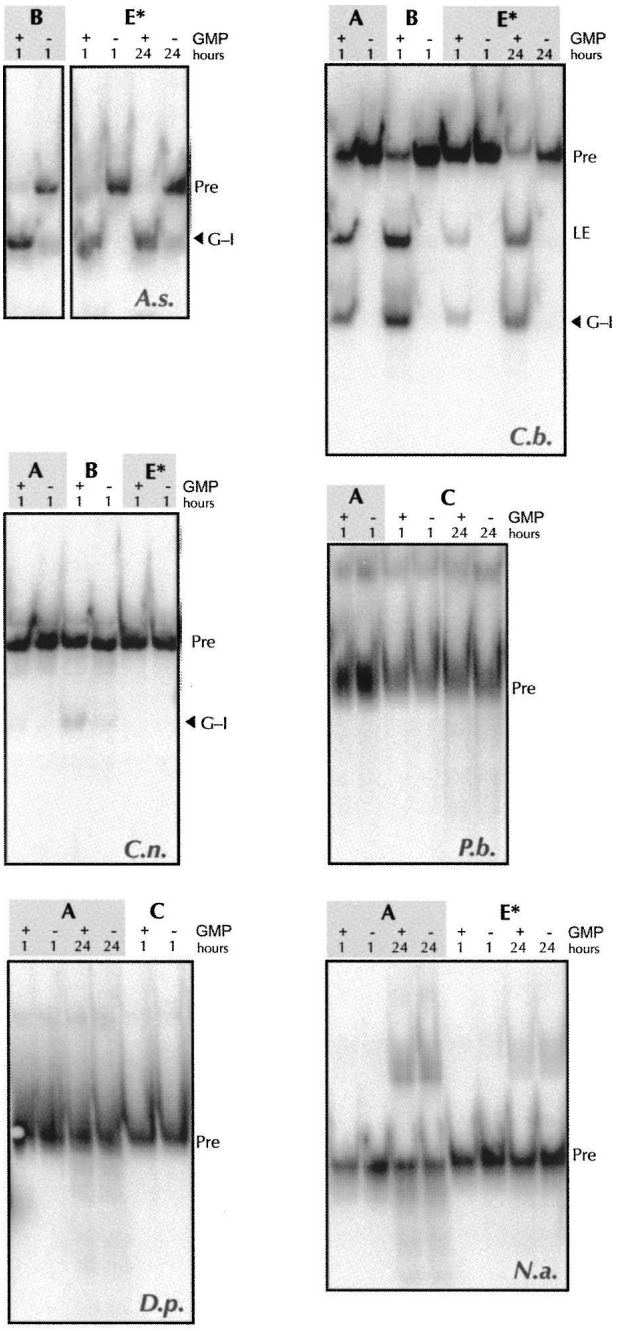

FIGURE 3. Native gel electrophoresis for six introns. Pre, precursor; G-I, excised intron with GMP at its 5′ end; LE, ligated exons. The asterisk denotes the following modification of condition E: 0.4 M NH_4OAc, 10 mM $Mg(OAc)_2$. GMP was used in place of GTP in this particular series of experiments.

et al. 2004), provided that such introns do not possess a large P5abc extension (Mohr et al. 1994). Upon addition of the CYT-18 protein, we observed stimulated splicing for two introns (A.p.LSU and D.p.) that were either poorly or not self-splicing in the absence of the protein (Fig. 2C). This protein-dependent splicing was observed only under conditions of lower KCl concentration (50 mM instead of 100 mM) and/or higher $MgCl_2$ concentration (30 mM instead of 5 mM). Gel shift and filter-binding experiments showed that CYT-18 binds to A.p.LSU under these conditions (Fig. 6A–C). In 1 h of incubation, the CYT-18 protein was able to promote reaction of the A.p.LSU intron to the level previously detected after 24 h of incubation in RNA-only conditions (Fig. 2B). In this case, however, CYT-18 mainly facilitated the first step of splicing, not the complete excision of the intron and ligation of the flanking exons. Finally, the splicing facilitation of the D.p. intron was weak but detectable compared to the absence of activity in RNA-only conditions (Fig. 2C).

Two of the self-splicing introns, S.h. and A.t., showed strong inhibition of splicing upon the addition of CYT-18 (Fig. 2A). CYT-18 bound to the S.h. intron (Fig. 6A–C; binding to the A.t. intron was not tested), up to a concentration of 0.7 M KCl (Fig. 6B,C) indicating an unusually strong binding that could be related to the inhibition. A time course revealed that increasing concentrations of CYT-18 slowed the reaction to the point where the ratio of spliced intron/intermediate product after 1 h in the presence of CYT-18 had still not reached the level seen after 5 min in the absence of CYT-18 (Fig. 7A). Because the first step of the self-splicing reaction was still rapid and efficient in the presence of the CYT-18 protein, CYT-18 likely affects either the conformational change that brings the omega G into the G-binding site for the second step of splicing, or the second step itself. The inhibition by CYT-18 was specific for these introns, as three other self-splicing introns did not show this behavior (Fig. 2A).

The C-terminal domain of CYT-18 is responsible for the inhibition of self-splicing

The N-terminal and the C-terminal domains of the CYT-18 protein work in concert to promote the self-splicing of the N.c. mitochondrial LSU intron (Mohr et al. 2001). The N-terminal nucleotide-binding domain of CYT-18 helps the intron fold into a catalytically active structure by binding to the P4–P6 region in a manner similar to that of the P5abc extension characteristic of the IC1 introns (Mohr et al. 1994; Caprara et al. 1996; Paukstelis et al. 2008). It also stabilizes tertiary interactions between the P4–P6 and P3–P9 domains (Mohr et al. 1992; Myers et al. 1996; Saldanha et al. 1996). The C-terminal domain of the protein appears to bind in proximity to the J4/5 junction and the P1 helix, which contains the 5′ exon, and in proximity to P3 and P8 (Myers et al. 2002; Paukstelis et al. 2005).

We analyzed splicing of the S.h. and A.p.LSU introns in the presence of a previously constructed CYT-18 variant (CYT-18/Δ424-669) that lacks the C-terminal domain but can still stimulate the splicing of a number of different group I introns (Mohr et al. 2001). Gel shift assays and filter binding experiments confirmed that the C-terminally

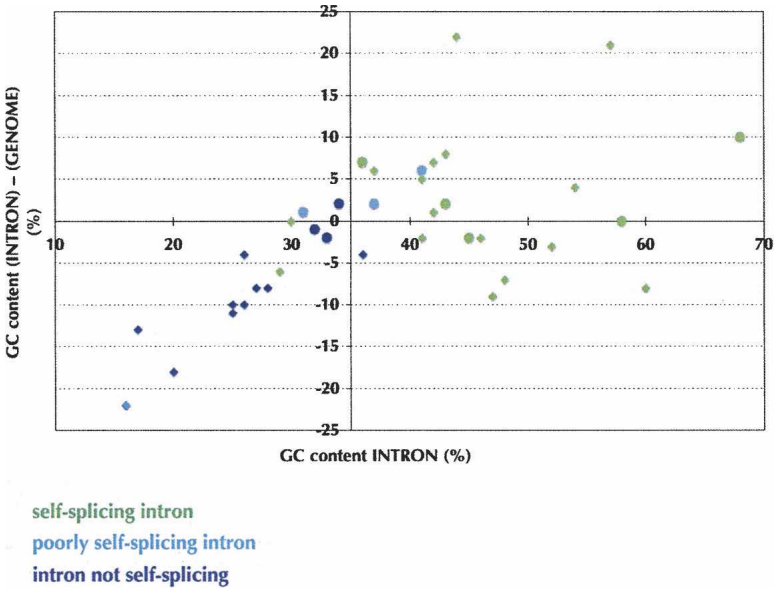

FIGURE 4. Correlation between self-splicing activity and the GC contents of 38 introns and of their host genomes. Introns from both this work (circles) and the literature (diamonds) are shown (Supplemental material). Same color code as in Figure 2.

truncated CYT-18 variant bound to the S.h. and A.p.LSU introns (Fig. 6A–C) and that the protein was active, based on its ability to bind to and activate splicing of the *N. crassa* ND1m intron (Figs. 6, 7; Wallweber et al. 1997). However, the truncated variant did not inhibit self-splicing of the S.h. intron (cf. lanes 7 and 9 with lanes 8 and 10 in Fig. 7B; Fig. 7C; self-splicing activity of the A.t. intron in the presence of the truncated protein was not tested). In contrast, in the case of the A.p.LSU intron, the C-terminal domain appears to be required for facilitation of splicing, rather than causing an inefficient second step of splicing (cf. lanes 11, 12, and 14 with lanes 13 and 15 in Fig. 7B; Fig. 7C). Overall, these observations suggest that depending on the intron, the C-terminal domain of CYT-18 can either contribute to or inhibit splicing.

DISCUSSION

Most but not all conditions are capable of revealing self-splicing activity

The survey of six RNA-only reaction conditions that have been found previously to promote ribozyme activity in different contexts supports the conclusion that introns that are robustly self-splicing in one condition are often similarly active under most or all of the other conditions (Table 1; Fig. 2). However, a diagnosis of self-splicing based on only one particular condition is not always accurate. For example, when conditions A (*Anabaena*) and F (*Didymium*) of our screen are independently considered, the first one detected self-splicing for five introns, while the second one identified only three of these (Table 1). This result also illustrates that higher salt concentrations (in condition F) do not necessarily improve self-splicing for all introns. Sampling different types and concentrations of salts may prove useful in identifying introns that have particular salt requirements for activity (Davila-Aponte et al. 1991). Additionally, we found that a 24 h-time point, especially at high salt concentration (1.0 M NaCl or 200 mM $MgCl_2$), was successful at distinguishing between low activity (1%–30% as active as the A.s. intron) and no activity (<<1% as active as the A.s. intron; compare, for example, the two A.p. introns with the N.a. and D.p. introns; Fig. 2).

Similarly, the effect of CYT-18 on splicing was in some cases dependent on salt concentration. For example, even though a twofold difference in the KCl concentration had no effect on the splicing of the N.c. intron, the A.p.LSU and the A.t. introns were respectively promoted and inhibited by CYT-18 only at the lower KCl concentration tested (Table 1, condition b; Fig. 2). The opposite was observed upon increasing the magnesium ion concentration: the N.c. activity was hampered while the inhibition of the A.t. splicing was more pronounced (Table 1, condition c; Fig. 2). Therefore, in general it is important to test a number of conditions before concluding that a particular group I intron is or is not self-splicing, and whether or not splicing is affected by protein factors. Other factors that may limit the extrapolation of our results include the small sample size, the generic annealing conditions of the precursor (2 min at 60°C, followed by slow cooling to 37°C), and the artificial exon length chosen for some constructs, which could perhaps give exon context-dependent misfolding (Jackson et al. 2006).

How these self-splicing and protein-facilitated behaviors observed in vitro would extrapolate to in vivo situations is not certain. However, it is reasonable to propose that introns that are active under all or most of the protein-free conditions surveyed would be self-splicing in vivo. Whereas efficient self-splicing is an obvious advantage in gene expression mechanisms, what about an inefficient one? The two previously untested introns (N.a. and D.p.) that show no activity in our screen (except for low activity after addition of CYT-18 for D.p.) belong to the IB4 subgroup. Introns from this subgroup are found mostly in large rRNA genes (Cannone et al. 2002) and are among the least-studied (Hur et al. 1997). Yet,

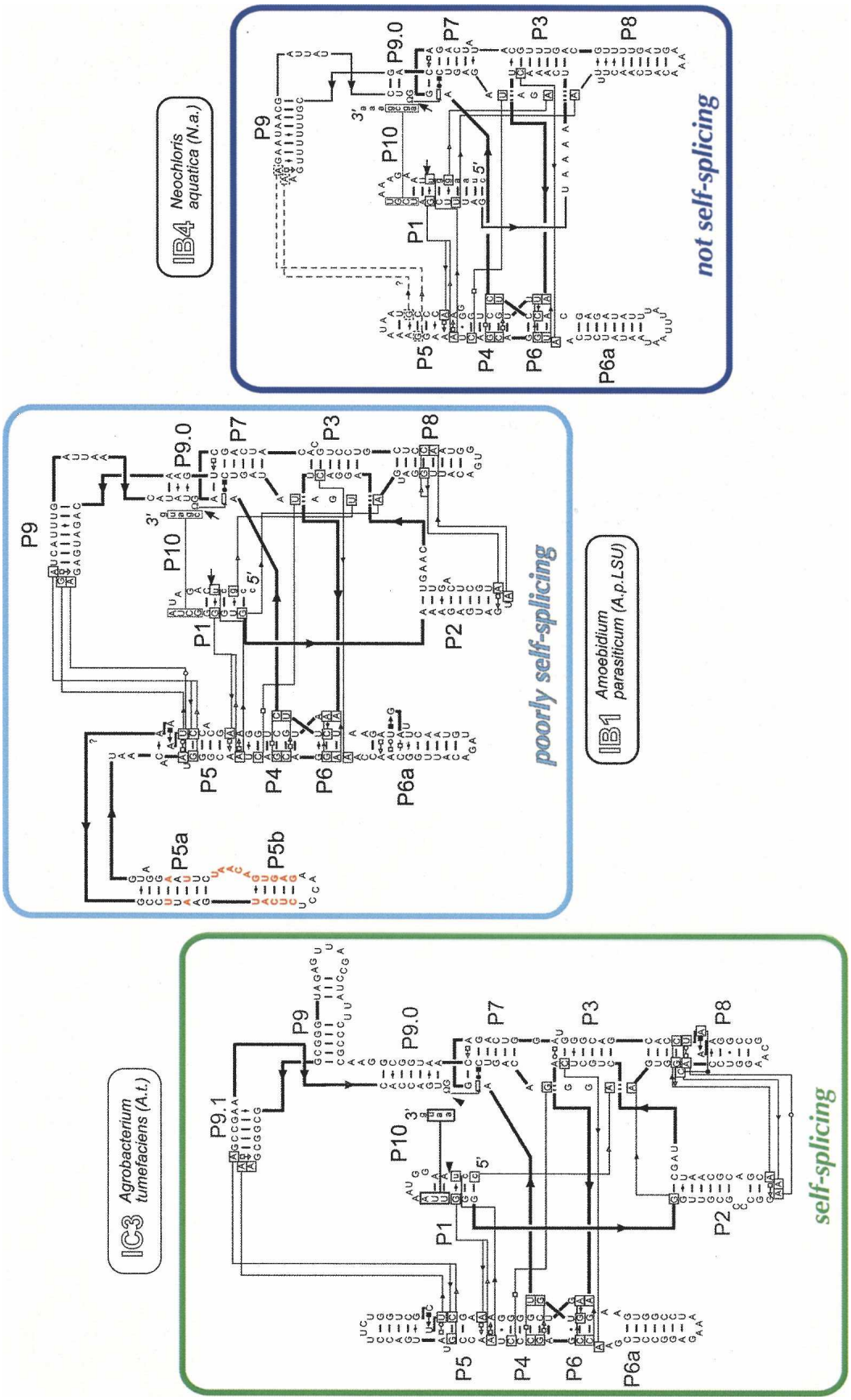

FIGURE 5. Interaction network diagrams for three representative group I introns. Base pairing within long range interactions and structural motifs is shown using the nomenclature of Leontis and Westhof (2001). Circles denote interactions involving the Watson-Crick face, squares the Hoogsteen face, and triangles the sugar edge. A question mark on the A.p.LSU diagram indicates the ambiguous topology of the J5/5a junction (residues 100% conserved in the P5ab extensions of the A.p.LSU and A.p.SSU introns are shown in red). The self-splicing behavior is indicated in the same color code as in Figure 2.

the general observation has been that they are not self-splicing in vitro (Cote and Turmel 1995; Hur et al. 1997), and may not splice even in vivo (Everett et al. 1999). Hence, the lack of splicing in our screen may also be a relevant in vivo property of these introns and in some cases may suggest protein facilitation by a protein different from CYT-18.

How close are we to predicting self-splicing activity?

The self-splicing activity of group I introns appears to result from a complex interplay between conserved regions and structural domains, some of which may favor the binding of one or more essential protein partners. Hence, our ability to predict self-splicing activity from the sequence of a group I intron is clouded (Hammann and Westhof 2007; Lang et al. 2007). The results presented herein contribute to fill in these gaps of knowledge in order to offer gateways to a broader understanding of group I intron biology, chemistry, and evolutionary history.

Even though the elements necessary for exon or guanosine binding (e.g., the G·U pair at the 5′ splice site, the P7 helix, and the J8/7 and J4/5 junctions; Cech 1988; Michel and Westhof 1990; Strobel and Cech 1995; Strobel et al. 1998) are highly conserved among group I introns, the extent to which a sequence difference in one of these regions will inhibit self-splicing is difficult to predict (Couture et al. 1990; Hammann and Westhof 2007). Additionally, recurrent structural motifs (e.g., the triple helix involving the P4 and P6 helices [Green et al. 1990; Michel et al. 1990], and the L9–P5 and L2–P8 interactions [Michel and Westhof 1990; Costa and Michel 1995]) need to be present, although different introns may utilize distinct combinations of these stabilizing elements to function (e.g., the triple helix is dispensable in the *td* intron from the bacteriophage T4 [Ikawa et al. 2000], but not in the *Tetrahymena* rRNA intron [Doudna and Cech 1995]).

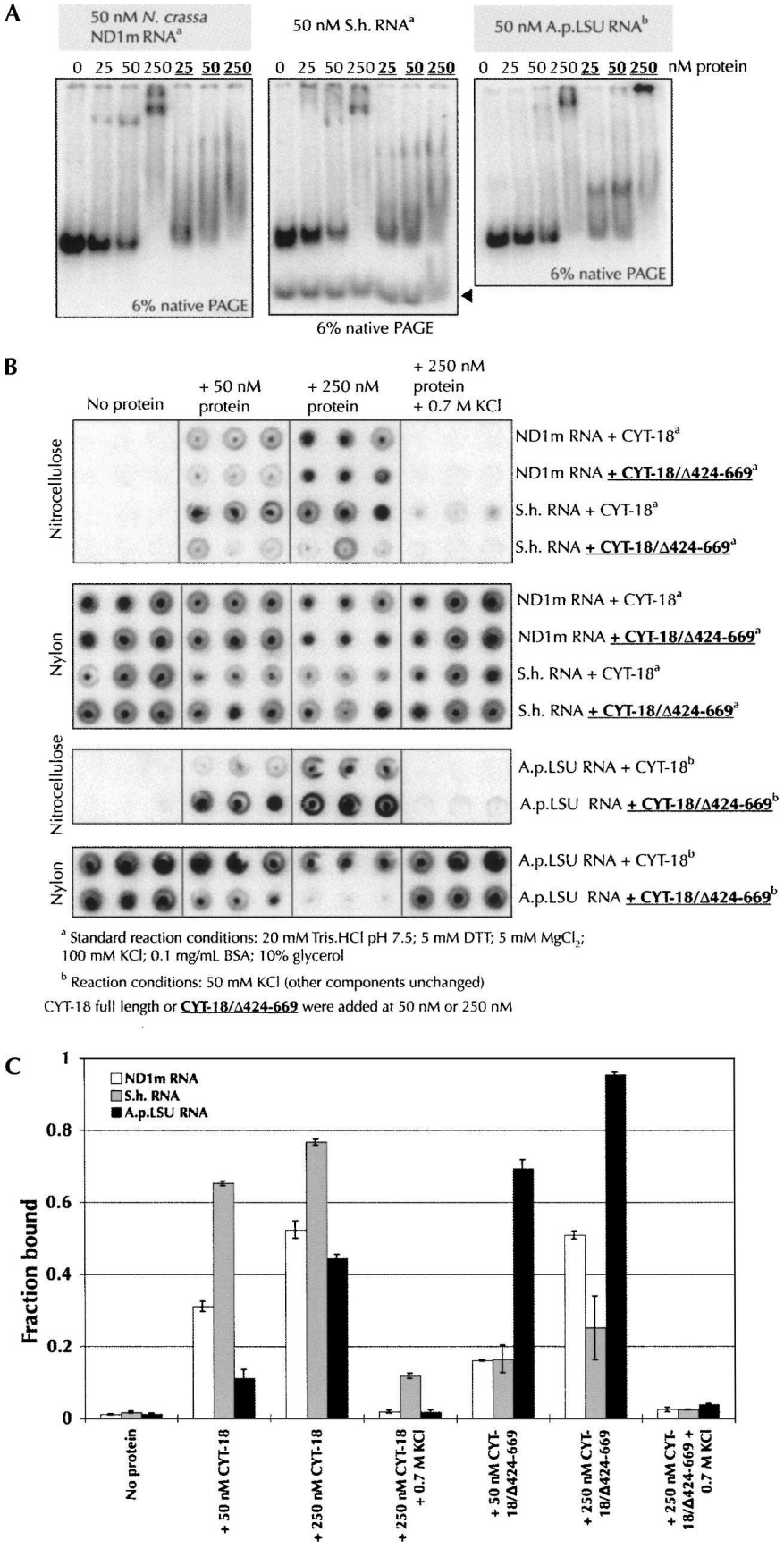

FIGURE 6. (Legend on next page)

Finally, self-splicing introns typically possess helical domains, junctions and loops that have optimal lengths, as first inferred from modeling studies (e.g., 12 base pairs are preferred between the G·U pair in P1 and the loop L2) (Michel and Westhof 1990). Introns with longer helices and loops may be more prone to misfolding and hence more likely to require a protein for activity (Akins and Lambowitz 1987; Weeks and Cech 1995; Goddard and Burt 1999; Webb et al. 2001; Longo et al. 2005), yet the thresholds determining such optimal lengths are not known. As a consequence, many basic questions concerning the self-splicing ability of group I introns, the optimal conditions that support catalysis, and the roles of protein facilitators remain barely addressed. More advanced queries pertaining for example to the relationship between secondary or tertiary structure and self-splicing ability (Brion et al. 1999; Hammann and Westhof 2007), or to the evolutionary pathways which lead to the acquisition of novel functions from pre-existing scaffolds (Beckert et al. 2008) have only recently been tackled. It is worth noting that such investigations are complicated by the poor sequence conservation exhibited by group I introns (Cech 1988; Michel and Westhof 1990).

While further supporting these criteria, the results of the present study expand our appreciation of the intricate combination of factors that determine self-splicing activity. We observed a few basic criteria that are predictive of self-splicing. The 35% GC correlation with self-splicing for the introns tested here is generally consistent with self-splicing tendencies reported in the literature (Fig. 4; Supplemental material). For example, the self-splicing group I introns from *Tetrahymena* (Kruger et al. 1982), *Azoarcus* (Reinhold-Hurek and Shub 1992), and the bacteriophages Twort (Landthaler and Shub 1999) and T4 (Chu et al. 1986; Ehrenman et al. 1986) have GC contents of 44%, 60%, 37%, and 43%, respectively. In contrast, the nonself-splicing group I introns from *Neurospora crassa* (Wallweber et al. 1997) are only about 25% GC rich (Supplemental material). Higher GC contents would likely strengthen secondary and tertiary interactions, as observed in RNase P RNAs (Baird et al. 2006). Such differences could account for the 200-fold difference in activity between a low-GC content intron (P.b.) and a high GC content intron (C.b.) that otherwise display the same set of long-range tertiary interactions.

An example where the 35% GC criterion is not sufficient occurs with the group I introns found at position 788 in the nuclear small ribosomal subunit gene of various fungi (Haugen et al. 2004). In that study, self-splicing and nonself-splicing introns had similar average GC contents of 57.6% and 57.9%, respectively (Supplemental material). Conversely, some yeast mitochondrial introns can be only 16% GC rich but are nonetheless self-splicing (e.g., bI5; Supplemental material). One limitation of this GC-content criterion is that structural features can improve thermostability without increasing the GC content (Juneau and Cech 1999; Guo and Cech 2002), which could account for the self-splicing ability of lower GC content introns. Another caveat to this analysis is that proving that an intron is not self-splicing is more difficult than proving that it is, as it requires a more extended survey of conditions. In conclusion, prediction of self-splicing activity likely requires an integrative analysis of multiple parameters, including but not limited to GC content, length of helices and single-stranded regions, number and type of long range interactions, particular sequences at the core, and exon context (Supplemental material).

Unexpected activities of the CYT-18 protein are revealed

Currently, our ability to predict the requirement of a protein partner to facilitate a splicing reaction is even less advanced than our ability to predict self-splicing. First, the sample size of introns dependent on protein(s) for splicing that can be studied is smaller than that for the self-splicing introns. Second, protein facilitators are diverse, adopt different strategies to facilitate splicing (compare, for example, CYT-18 [Mohr et al. 1994], CBP2 [Bokinsky et al. 2006], and StpA [Waldsich et al. 2002]), and are specific to particular organisms (Bonnefond et al. 2007; Paukstelis and Lambowitz 2008). Finally, the absence of self-splicing can neither imply nor rule out the necessity of a protein facilitator.

An unexpected result of this survey was the CYT-18 facilitated splicing of the A.p.LSU intron but not the A.p.SSU intron. Both introns share 50% of their residues overall. Additionally, out of 38 core residues shared by CYT-18 dependent introns (Wallweber et al. 1997), 29 are possessed by the A.p.SSU and A.p.LSU introns. However, they also possess a P5ab (A.p.LSU) or P5abd (A.p.SSU) extension (Supplemental material) that should be large enough (43 nt, A.p.LSU; 67 nt, A.p. SSU) to preclude CYT-18 binding (the smallest P5ab domain sufficient to

FIGURE 6. Binding assays of the full-length and C-terminally truncated CYT-18 proteins to the ND1m, S.h., and A.p.LSU introns. (*A*) Gel shift assays with 50 nM RNA and increasing concentrations (0, 25 nM, 50 nM, and 250 nM) of CYT-18 full length or truncated (bold underlined). In this and subsequent panels, CYT-18 binding and protein-facilitation assays were carried out in standard conditions for the ND1m and S.h. introns (20 mM Tris-HCl at pH 7.5, 5 mM DTT, 5 mM $MgCl_2$, 100 mM KCl, 0.1 mg/mL BSA, 10% glycerol) and at 50 mM KCl for the A.p.LSU intron (other components unchanged). The black arrow indicates the linearized S.h. intron. (*B*) Filter binding assays upon addition of increasing concentrations (0, 50 nM, and 250 nM) of CYT-18 full length or tuncated (bold underlined). Experiments were performed in triplicate. (*C*) Quantitation of the filter binding experiments shown in panel *B*. The error bars indicate the standard deviation calculated from the triplicate experiments shown in *B*.

impair CYT-18 binding is 42-nt long; Mohr et al. 1994; Wallweber et al. 1997). Clearly, this is a case where examination of RNA secondary structure is insufficient to predict whether or not the position of a peripheral element in three-dimensional space will sterically inhibit the binding of CYT-18. The protein-induced stabilization necessary for catalysis of the A.p.LSU intron presumably results from CYT-18 binding to regions of the P4–P6 domain that are not blocked by the P5ab extension, or to other regions. In general, the inability of CYT-18 to stimulate the splicing of some introns could be explained by the following possibilities: (1) a CYT-18 binding site is not accessible (i.e., obstructed by peripheral elements); (2) the core sequences and structures do not allow for CYT-18 binding (Wallweber et al. 1997; and see discussion above for the two A.p. introns); (3) the rate of self-splicing for some introns cannot be increased under the reaction conditions tested, because the rate-limiting step is not intron folding (Paukstelis et al. 2008).

Another challenge in understanding the roles of proteins in splicing of group I introns was brought by the unexpected ability of CYT-18 to inhibit self-splicing of two introns. As these results were obtained from in vitro experiments of heterologous complexes, it is difficult to extrapolate them in vivo. In particular, since this inhibition is observed for two of the tRNA-embedded introns, it could be linked to the ability of CYT-18 to bind either tRNAs or group I introns (Akins and Lambowitz 1987). Nonetheless, this previously unreported activity may prove useful in studying the conformational rearrangement that occurs between the two steps of splicing (Hougland et al. 2006). Specifically, if CYT-18 could trap a group I intron at this particular stage, structural analyses could give a clear view of that splicing intermediate.

Future surveys of additional introns are needed to test the criteria for self-splicing activity presented herein. The ability to predict the self-splicing and the protein-facilitated splicing activity

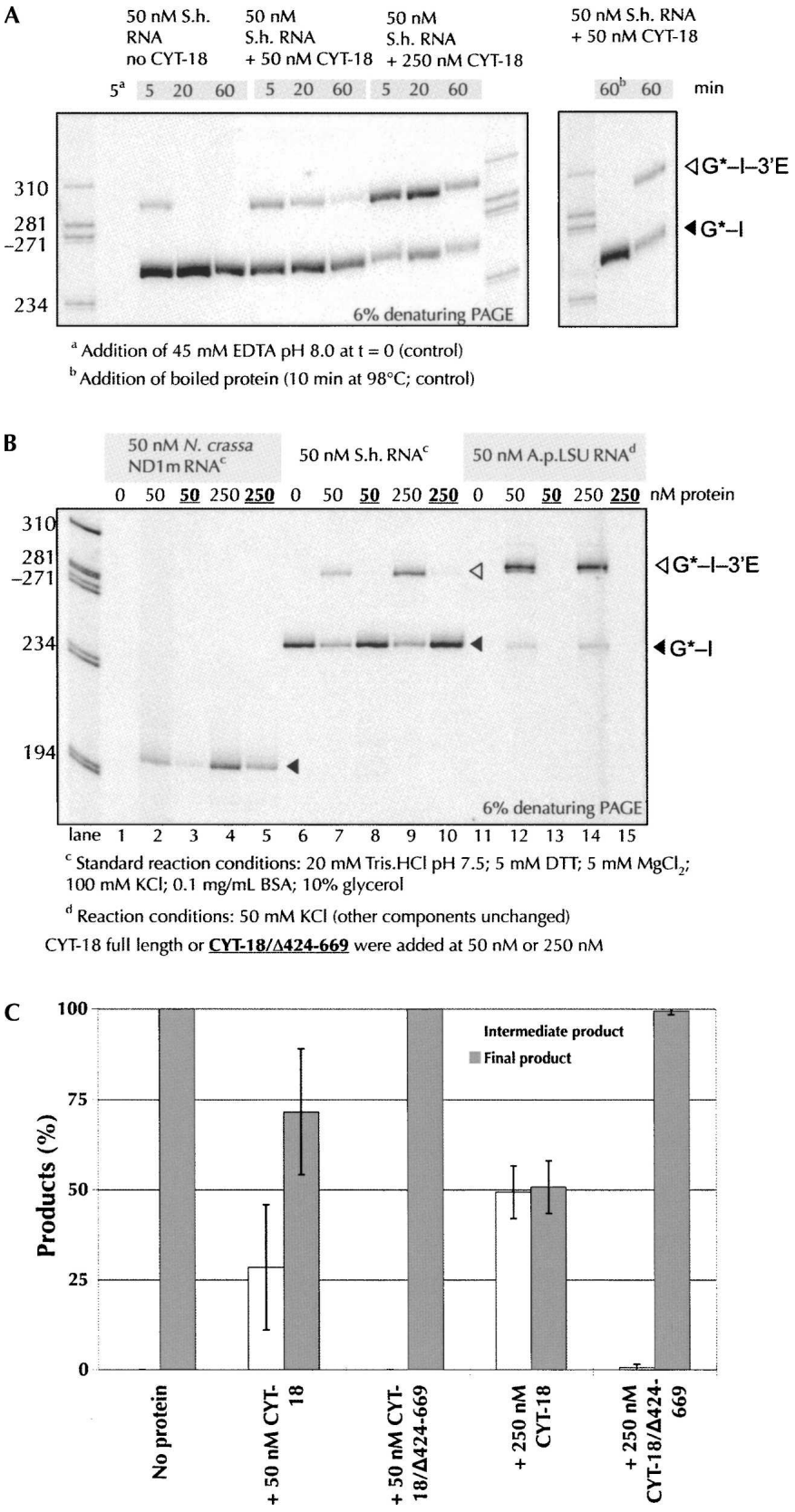

FIGURE 7. (Legend on next page)

of any group I intron as well as further developments of comprehensive sets of conditions to routinely test for such RNA catalytic functions would contribute to the efforts of the RNA Ontology Consortium in systematizing and organizing our knowledge of RNA molecules and their activities (Leontis et al. 2006).

MATERIALS AND METHODS

Intron selection

Forty requests for group I introns of various eucaryotes and bacteria, sampling the 14 IA1-IE3 subclasses, containing 22%–72% GC, and ranging in size from 183 to 496 nt, were sent to several laboratories throughout the world. Fifteen introns were obtained from those requests, plus three introns that had not been requested. Out of these 18 introns, only half were suitable for PCR reactions and could be PCR-amplified (see below) without any major difficulty. The N.c., S.h., and A.t. introns were added afterward to make the 12 introns that are studied herein.

Construct synthesis and preparation

PCR amplification

Introns were either obtained as plasmids that could be readily used for transcription (N.c., C.b., A.s.), or were amplified by PCR in 50 μL-reaction mixes containing: 1.0 μL of the DNA template containing the intron (or 1 μL of a 1/10 dilution, depending on the conditioning of each sample); 1.0 μM forward primer (containing six random nucleotides, a restriction site for cloning at the 5′ end, the T7 polymerase promoter 5′-TAATACGACT CACTATA-3′, three guanosine residues, 30 nucleotides corresponding to the 5′ exon, and 5–20 residues corresponding to the 5′ end of the intron); 1.0 μM reverse primer (containing 5–20 residues of the 3′ end of the intron, 45 nucleotides from the 3′ exon, a restriction site for linearization, a restriction site for cloning at the 3′ end, and six random nucleotides); 51.3 μM each dNTP; 1.0 mM $MgCl_2$; 2.5 U Herculase DNA polymerase (Stratagene); and 1× reaction buffer supplied by the manufacturer. PCRs were performed in a Robocycler Gradient 96 apparatus (Stratagene) using the following protocol: 4 min at 94°C (1 cycle); 55 sec at 94°C, 55 sec at 60°C, 4 min at 72°C (35 cycles); 10 min at 72°C (1 cycle).

Cloning procedure and transformation

The PCR amplified products were purified using the QIAquick PCR purification kit (Qiagen), digested using the appropriate restriction enzymes (EcoRI, New England Biolabs [NEB]; BamHI, NEB), purified by agarose gel electrophoresis, and eluted in a final volume of 30 μL using the QIAquick gel extraction kit (Qiagen). Ligation occurred for 35 min at room temperature (~23°C–25°C) in a solution containing: 0.2 μg/μL pUC19 vector (NEB); 50–100 ng/μL PCR product; 400 U T4 DNA ligase (NEB); and 1× reaction buffer supplied by the manufacturer. The pUC19 had been previously digested using the EcoRI and BamHI restriction enzymes and dephosphorylated using 1.0 U calf intestine alkaline phosphatase (Roche) for 30 min at 37°C followed by 20 min at 70°C. 1.0 μL of the ligation reaction mix was used to transform 10 μL of Solopack Gold Supercompetent cells (Stratagene). Transformation was performed using the following protocol: 20 min at 4°C; 1 min at 54°C; addition of LB (−Amp) broth; incubation on LB (+Amp) plates for 16–20 h at 37°C. Plasmids possessing the expected sequence were purified using the QIAprep spin miniprep kit (Qiagen) from 3–5 mL LB cultures grown for 16–20 h at 37°C. The S.h. and A.t. introns were independently subcloned into pUC19 in order to include the complete precursor tRNAs downstream of a T7 polymerase promoter (A.R. Gooding and T.R. Cech, unpubl.).

In vitro transcription of unlabeled RNAs

Plasmids were linearized using a suitable restriction enzyme (Supplemental Table S1; BanI, BsaI, EarI, FokI [NEB]). The linearized plasmids of all introns except N.c. were transcribed for 1 h at 37°C in 25-μL reaction mixes containing: 1.0–4.0 μg/μL linearized plasmid, 4.0–5.0 mM each rNTP, 5.0–16.0 mM $MgCl_2$ (the concentration ratio of rNTP/$MgCl_2$ was adjusted on a construct per construct basis so that occasional cotranscriptional self-splicing would be inhibited; Zaug et al. 1993), 2.0 mM spermidine-HCl, 10 mM DTT, 0.1% Triton X-100, 4.0 ng/μL inorganic pyrophosphatase (Sigma), 40 U of RNasin Plus RNase inhibitor (Promega), 2 μL of T7 RNA polymerase (prepared in house by Anne Gooding), 50 mM Tris-HCl (pH 7.5). The linearized plasmid containing the N.c. intron was transcribed for 1 h at 37°C in a 25-μL reaction mix containing: 1.0 μg/μL linearized plasmid, 4.0 mM each rNTP, 8.0 mM $MgCl_2$, 25 mM NaCl, 2.0 mM spermidine-HCl, 10 mM DTT, 40 U RNasin Plus RNase inhibitor, 20 U T3 RNA polymerase (Ambion), 50 mM Tris-HCl (pH 8.0). Prior to being used in splicing assays, all transcription products were purified using the MegaClear kit (Ambion), precipitated for 18 h at −20°C in 100% EtOH, 10% (v/v) 3.0 M Na acetate (pH 5.3), and resuspended in water.

In vitro transcription of body-labeled RNAs

For native gel electrophoreses, the linearized plasmids containing the P.b., C.n., C.b., N.a., D.p., and A.s. introns were transcribed in a similar reaction solution, but in the presence of 70 μCi [α-^{32}P]ATP (PerkinElmer). These transcription products were purified by denaturing gel electrophoresis (5% acrylamide/bisacrylamide [19:1], 8.0 M urea, 1× TBE). Gel slices were crushed and soaked for 12 h at 4°C in a buffer containing: 1.0 mM EDTA, 250 mM NaCl, 10 mM Tris-HCl (pH 7.5). RNAs were precipitated in 100% EtOH, 10% (v/v) 3.0 M Na acetate (pH 5.3) and resuspended in water. The final radioactivity of the samples

FIGURE 7. Facilitation or inhibition of splicing upon addition of the CYT-18 protein. (*A*) Time course self-splicing experiment (5, 20, and 60 min time points) of the S.h. intron in the presence of increasing concentrations of the CYT-18 protein (onefold and fivefold of the RNA concentration). (*B*) Comparison of the effects of the addition of full length CYT-18 or of C-terminally truncated CYT-18 on the self-splicing activity of the N.c., S.h., and A.p.LSU introns. (*C*) Product ratios obtained in the presence of full length or C-terminally truncated CYT-18 protein. The error bars indicate the standard deviation calculated from three independent experiments.

was measured by the number of counts per minute (cpm) for 1.0 μL of the RNA solution diluted into 3.0 mL ScintiSafe Econo 1 solution (Fisher Scientific), obtained using a scintillation counter (Beckman).

Splicing assays

RNA-only conditions

RNA constructs were diluted to 500 nM in a buffer containing: 15 mM $MgCl_2$, 25 mM NaCl, 25 mM HEPES (pH 7.5). RNAs were subsequently incubated for 2 min at 60°C in a heating block, then slow cooled for ~1 h to 32°C or 42°C. For the native gel analyses, body-labeled RNAs were incubated for 3 min at 50°C in 25 mM HEPES (pH 7.5), and then for 10 min at 32°C in buffer A, B, C, or E*. E* was a modification of condition E and contained 0.4 M NH_4OAc, 10 mM $Mg(OAc)_2$. Splicing behavior of the RNAs was independent of the chosen annealing protocol (data not shown). Self-splicing reactions proceeded for 1 h or 24 h at various temperatures (Table 2) in 10-μL reaction mixes containing: 50 nM RNA; 0.2 mM GTP; 10 μCi [α-^{32}P]GTP (PerkinElmer); salts and buffer as described in Table 2. Reactions were stopped by the addition of a solution containing: 500 mM EDTA, 7.0 M urea, 0.02% bromophenol blue, 0.02% xylene cyanol, 1× TBE (100 mM Tris-base, 83 mM boric acid, 1.0 mM EDTA). For the native gel analyses, body-labeled RNAs were incubated for 1 min, 1 h, or 23 h at various temperatures (Table 2) in 6.0-μL reaction mixes containing: 20,000 cpm body-labeled RNA, 0.1 mM GMP, salts, and buffer as described in Table 2. Reactions were stopped by placing the samples at 4°C and by adding 10% (v/v) 50% glycerol prior to loading on native polyacrylamide gels.

CYT-18 conditions

RNA constructs were diluted to 500 nM in a buffer containing: 5.0 mM DTT, 0.1 mg/mL bovine serum albumine (NEB), 10% glycerol, salts and buffer as described in Table 2. RNAs diluted under conditions a, b, or c (Table 2) were incubated for 2 min at 60°C in a heating block, then slow cooled for ~50 min to 37°C. Full-length and C-terminally truncated CYT-18 proteins were expressed and purified as previously described (Paukstelis et al. 2008). Full-length and truncated proteins were stored at −80°C at 7.5 μM and 10 μM, respectively (concentration of the dimer), in buffers containing: 500 mM KCl, 10% glycerol, 1.0 mM DTT, 25 mM Tris-HCl (pH 7.5) (full length); or 100 mM KCl, 50% glycerol, 25 mM Tris-HCl (pH 7.5) (truncated). Immediately prior to the splicing assays, both proteins were diluted to 500 nM or 5.0 μM in the buffer used to anneal the RNA. Protein-facilitated reactions proceeded for 1 h at 37°C in 10-μL reaction mixes containing: 50 nM RNA, 50 nM protein (unless otherwise stated), 0.2 mM GTP, 10 μCi [α-^{32}P]GTP, salts and buffer as described in Table 2. Reactions were stopped by the addition of 50 mM EDTA (pH 8.0), phenol-chloroform extracted, precipitated for 16 h at −20°C in 100% EtOH, 10% (v/v) 3.0 M Na acetate (pH 5.3), 0.5 mg/mL glycogen (Roche), and resuspended in a solution containing: 30 mM EDTA, 8.0 M urea, 0.02% bromophenol blue, 0.02% xylene cyanol, 1× TBE.

Denaturing and native polyacrylamide gel electrophoreses

Products were separated on either denaturing sequencing gels (6% acrylamide/bisacrylamide [19:1], 7.0 M urea, 1× TBE) or native gels (6% acrylamide/bisacrylamide [19:1], 15 mM $MgCl_2$, 25 mM NaCl, 0.5× TBE, run at 4°C). The gels were dried under vacuum and placed for 18–30 h in PhosphorImager screens (Molecular Dynamics). The screens were scanned using a PhosphorImager (Amersham Bioscience/GE) and the gel images were analyzed using ImageQuant TL v. 2005 (Amersham Bioscience/GE).

Gel shift and filter binding assays

Body labeled RNAs for gel shift and filter binding assays were transcribed from linearized templates using a Megascript transcription kit (Ambion) with 3.3 mM [α-^{32}P]UTP (0.445 Ci $mmol^{-1}$; PerkinElmer). Full-length transcripts were separated from unincorporated nucleotides by sequential 1 mL Sephadex G-50 columns and quantified based on incorporation of the radiolabel. 20 μL reactions were assembled as described for the CYT-18-dependent splicing reactions with GTP omitted. Reactions were incubated for 1 h at 37°C and were directly loaded on 1× TBE 6% native polyacrylamide [19:1] gels with 10 mM $MgCl_2$ for gel shift assays. For filter binding, 5 μL of each reaction was spotted on to layered nitrocellulose (Trans-Blot; BioRad) and nylon (Hybond-N; Amersham) membranes that were presoaked in the appropriate reaction buffer (BSA omitted) in a 96-well vacuum manifold (Schleicher & Schuell). Membranes were washed three times with 50 μL of the appropriate buffer. Gels and membranes were imaged on a phosphoimager, and membranes were quantitated with ImageQuant software (Amersham Bioscience/GE).

Computational analysis of the GC contents

Self-splicing activities, gene locations, intron positions, subgroups, lengths, and GenBank accession numbers of the 38 introns shown in Figure 4 and in Supplemental material were retrieved from the literature or from the Comparative RNA website and project (http://www.rna.ccbb.utexas.edu; Cannone et al. 2002). Sequences were retrieved from GenBank (http://www.ncbi.nlm.nih.gov). GC contents of the introns were calculated using the Oligonucleotide Properties Calculator (http://www.basic.northwestern.edu/biotools/oligocalc.html). GC contents of the host genomes were obtained from the literature or retrieved from the Codon Usage Database at http://www.kazusa.or.jp/codon/. Coding sequences containing fewer than 500 codons were not used in the analysis. Differences in GC contents were calculated and graphed in Microsoft Excel.

Secondary structure diagrams and prediction of tertiary interactions

Diagrams were adapted from publications (N.c., C.b., A.s., A.t., S.d.; Table 1), or predicted based on pre-existing alignments of similar introns (Michel and Westhof 1990; Lehnert et al. 1996) and altered from M-FOLD outputs (Zuker 2003) for peripheral elements (P.b., C.n., A.p.LSU, A.p.SSU, N.a., D.p., S.h.). Two-dimensional diagrams of introns possessing a P2 element were modified to satisfy the 12-base pair rule (between the G·U pair at the 5′ splice site and the tetraloop L2) (Michel and Westhof 1990) and to incorporate consensus sequences for the sarcin/ricin motif (Leontis and Westhof 1998; Leontis et al. 2002). Eight tertiary interactions that are universal among group I introns and

subgroup-specific interactions were modeled based on available crystal structures (Adams et al. 2004; Guo et al. 2004; Golden et al. 2005) and additional available references (complete list in Supplemental material).

SUPPLEMENTAL DATA

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We thank all of our intron providers (Supplemental material) for their generous gifts of genomic DNA and plasmids, Anne Gooding for the clones of the S.h. and A.t. introns, Rob Knight for insights on the computational analysis of GC contents, anonymous reviewers for their comments, and the Cech laboratory members for helpful suggestions on experiments and on the manuscript. Q.V. is an Associate of the Howard Hughes Medical Institute. P.J.P. and A.M.L. acknowledge the support of NIH grant GM037951.

Received February 2, 2008; accepted July 8, 2008.

REFERENCES

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

Akins, R.A. and Lambowitz, A.M. 1987. A protein required for splicing group I introns in *Neurospora* mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. *Cell* **50:** 331–345.

Baer, M.F., Arnez, J.G., Guerrier-Takada, C., Vioque, A., and Altman, S. 1990. Preparation and characterization of RNase P from *Escherichia coli*. *Methods Enzymol.* **181:** 569–582.

Baird, N.J., Srividya, N., Krasilnikov, A.S., Mondragon, A., Sosnick, T.R., and Pan, T. 2006. Structural basis for altering the stability of homologous RNAs from a mesophilic and a thermophilic bacterium. *RNA* **12:** 598–606.

Beckert, B., Nielsen, H., Einvik, C., Johansen, S.D., Westhof, E., and Masquida, B. 2008. Molecular modelling of the GIR1 branching ribozyme gives new insight into evolution of structurally related ribozymes. *EMBO J.* **27:** 667–678.

Bokinsky, G., Nivon, L.G., Liu, S., Chai, G., Hong, M., Weeks, K.M., and Zhuang, X. 2006. Two distinct binding modes of a protein cofactor with its target RNA. *J. Mol. Biol.* **361:** 771–784.

Bonnefond, L., Frugier, M., Touze, E., Lorber, B., Florentz, C., Giege, R., Sauter, C., and Rudinger-Thirion, J. 2007. Crystal structure of human mitochondrial tyrosyl-tRNA synthetase reveals common and idiosyncratic features. *Structure* **15:** 1505–1516.

Brion, P., Schroeder, R., Michel, F., and Westhof, E. 1999. Influence of specific mutations on the thermal stability of the *td* group I intron in vitro and on its splicing efficiency in vivo: A comparative study. *RNA* **5:** 947–958.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., et al. 2002. The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3:** 2. doi: 10.1186/1471-2105-3-2.

Caprara, M.G., Lehnert, V., Lambowitz, A.M., and Westhof, E. 1996. A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core. *Cell* **87:** 1135–1145.

Cech, T.R. 1988. Conserved sequences and structures of group I introns: Building an active site for RNA catalysis—A review. *Gene* **73:** 259–271.

Chen, X., Mohr, G., and Lambowitz, A.M. 2004. The *Neurospora* crassa CYT-18 protein C-terminal RNA-binding domain helps stabilize interdomain tertiary interactions in group I introns. *RNA* **10:** 634–644.

Chu, F.K., Maley, G.F., West, D.K., Belfort, M., and Maley, F. 1986. Characterization of the intron in the phage T4 thymidylate synthase gene and evidence for its self-excision from the primary transcript. *Cell* **45:** 157–166.

Collins, R.A. and Lambowitz, A.M. 1985. RNA splicing in *Neurospora* mitochondria. Defective splicing of mitochondrial mRNA precursors in the nuclear mutant cyt18-1. *J. Mol. Biol.* **184:** 413–428.

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

Costa, M. and 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.

Cote, M.J. and Turmel, M. 1995. In vitro self-splicing reactions of chloroplast and mitochondrial group-I introns in *Chlamydomonas eugametos* and *Chlamydomonas moewusii*. *Curr. Genet.* **27:** 177–183.

Couture, S., Ellington, A.D., Gerber, A.S., Cherry, J.M., Doudna, J.A., Green, R., Hanna, M., Pace, U., Rajagopal, J., and Szostak, J.W. 1990. Mutational analysis of conserved nucleotides in a self-splicing group I intron. *J. Mol. Biol.* **215:** 345–358.

Davila-Aponte, J.A., Huss, V.A., Sogin, M.L., and Cech, T.R. 1991. A self-splicing group I intron in the nuclear pre-rRNA of the green alga, *Ankistrodesmus stipitatus*. *Nucleic Acids Res.* **19:** 4429–4436.

Doudna, J.A. and Cech, T.R. 1995. Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA* **1:** 36–45.

Downs, W.D. and Cech, T.R. 1996. Kinetic pathway for folding of the *Tetrahymena* ribozyme revealed by three UV-inducible crosslinks. *RNA* **2:** 718–732.

Ehrenman, K., Pedersen-Lane, J., West, D., Herman, R., Maley, F., and Belfort, M. 1986. Processing of phage T4 td-encoded RNA is analogous to the eukaryotic group I splicing pathway. *Proc. Natl. Acad. Sci.* **83:** 5875–5879.

Emerick, V.L. and Woodson, S.A. 1993. Self-splicing of the *Tetrahymena* pre-rRNA is decreased by misfolding during transcription. *Biochemistry* **32:** 14062–14067.

Everett, K.D., Kahane, S., Bush, R.M., and Friedman, M.G. 1999. An unspliced group I intron in 23S rRNA links Chlamydiales, chloroplasts, and mitochondria. *J. Bacteriol.* **181:** 4734–4740.

Goddard, M.R. and Burt, A. 1999. Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci.* **96:** 13880–13885.

Golden, B.L., Gooding, A.R., Podell, E.R., and Cech, T.R. 1998. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* **282:** 259–264.

Golden, B.L., Kim, H., and 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., and Szostak, J.W. 1990. In vitro genetic analysis of the Tetrahymena self-splicing intron. *Nature* **347:** 406–408.

Guo, F. and Cech, T.R. 2002. Evolution of *Tetrahymena* ribozyme mutants with increased structural stability. *Nat. Struct. Biol.* **9:** 855–861.

Guo, Q.B., Akins, R.A., Garriga, G., and Lambowitz, A.M. 1991. Structural analysis of the *Neurospora* mitochondrial large rRNA intron and construction of a mini-intron that shows protein-dependent splicing. *J. Biol. Chem.* **266:** 1809–1819.

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

Hammann, C. and Westhof, E. 2007. Searching genomes for ribozymes and riboswitches. *Genome Biol.* **8:** 210.

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

Hougland, J.L., Piccirilli, J.A., Forconi, M., Lee, J., and Herschlag, D. 2006. How the group I intron works: A case study of RNA structure and function. In *The RNA world*, 3d ed. (eds. R.F. Gesteland et al.), pp. 133–205. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hur, M., Geese, W.J., and Waring, R.B. 1997. Self-splicing activity of the mitochondrial group-I introns from *Aspergillus nidulans* and related introns from other species. *Curr. Genet.* **32:** 399–407.

Ikawa, Y., Shiraishi, H., and Inoue, T. 2000. Minimal catalytic domain of a group I self-splicing intron RNA. *Nat. Struct. Biol.* **7:** 1032–1035.

Inoue, T., Sullivan, F.X., and Cech, T.R. 1986. New reactions of the ribosomal RNA precursor of *Tetrahymena* and the mechanism of self-splicing. *J. Mol. Biol.* **189:** 143–165.

Jackson, S.A., Koduvayur, S., and Woodson, S.A. 2006. Self-splicing of a group I intron reveals partitioning of native and misfolded RNA populations in yeast. *RNA* **12:** 2149–2159.

Jaeger, J.A., Zuker, M., and Turner, D.H. 1990. Melting and chemical modification of a cyclized self-splicing group I intron: Similarity of structures in 1 M Na^+, in 10 mM Mg^{2+}, and in the presence of substrate. *Biochemistry* **29:** 10147–10158.

Jaeger, L., Westhof, E., and Michel, F. 1991. Function of P11, a tertiary base pairing in self-splicing introns of subgroup IA. *J. Mol. Biol.* **221:** 1153–1164.

Jaeger, L., Michel, F., and Westhof, E. 1996. The structure of group I ribozymes. In *Nucleic acids and molecular biology* (eds. F. Eckstein and D.M.J. Lilley), pp. 33–51. Springer-Verlag, Berlin.

Johansen, S. and Vogt, V.M. 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.

Joyce, G.F., van der Horst, G., and Inoue, T. 1989. Catalytic activity is retained in the *Tetrahymena* group I intron despite removal of the large extension of element P5. *Nucleic Acids Res.* **17:** 7879–7889.

Juneau, K. and Cech, T.R. 1999. In vitro selection of RNAs with increased tertiary structure stability. *RNA* **5:** 1119–1129.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. 1982. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31:** 147–157.

Lambowitz, A.M. and Perlman, P.S. 1990. Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem. Sci.* **15:** 440–444.

Landthaler, M. and Shub, D.A. 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: Introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. *Proc. Natl. Acad. Sci.* **96:** 7005–7010.

Lang, B.F., Laforest, M.J., and Burger, G. 2007. Mitochondrial introns: A critical view. *Trends Genet.* **23:** 119–125.

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

Leontis, N.B. and 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. and Westhof, E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7:** 499–512.

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

Leontis, N.B., Altman, R.B., Berman, H.M., Brenner, S.E., Brown, J.W., Engelke, D.R., Harvey, S.C., Holbrook, S.R., Jossinet, F., Lewis, S.E., et al. 2006. The RNA Ontology Consortium: An open invitation to the RNA community. *RNA* **12:** 533–541.

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

Longo, A., Leonard, C.W., Bassi, G.S., Berndt, D., Krahn, J.M., Hall, T.M., and Weeks, K.M. 2005. Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nat. Struct. Mol. Biol.* **12:** 779–787.

Michel, F. and 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.

Michel, F., Ellington, A.D., Couture, S., and Szostak, J.W. 1990. Phylogenetic and genetic evidence for base-triples in the catalytic domain of group I introns. *Nature* **347:** 578–580.

Mohr, G., Zhang, A., Gianelos, J.A., Belfort, M., and Lambowitz, A.M. 1992. The neurospora CYT-18 protein suppresses defects in the phage T4 td intron by stabilizing the catalytically active structure of the intron core. *Cell* **69:** 483–494.

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

Mohr, G., Rennard, R., Cherniack, A.D., Stryker, J., and 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.

Murphy, F.L. and Cech, T.R. 1993. An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry* **32:** 5291–5300.

Myers, C.A., Wallweber, G.J., Rennard, R., Kemel, Y., Caprara, M.G., Mohr, G., and Lambowitz, A.M. 1996. A tyrosyl-tRNA synthetase suppresses structural defects in the two major helical domains of the group I intron catalytic core. *J. Mol. Biol.* **262:** 87–104.

Myers, C.A., Kuhla, B., Cusack, S., and Lambowitz, A.M. 2002. tRNA-like recognition of group I introns by a tyrosyl-tRNA synthetase. *Proc. Natl. Acad. Sci.* **99:** 2630–2635.

Nielsen, H., Fiskaa, T., Birgisdottir, A.B., Haugen, P., Einvik, C., and Johansen, S. 2003. The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA* **9:** 1464–1475.

Paukstelis, P.J. and Lambowitz, A.M. 2008. Identification and evolution of fungal mitochondrial tyrosyl-tRNA synthetases with group I intron splicing activity. *Proc. Natl. Acad. Sci.* **105:** 6010–6015.

Paukstelis, P.J., Coon, R., Madabusi, L., Nowakowski, J., Monzingo, A., Robertus, J., and Lambowitz, A.M. 2005. A tyrosyl-tRNA synthetase adapted to function in group I intron splicing by acquiring a new RNA binding surface. *Mol. Cell* **17:** 417–428.

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

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

Saldanha, R., Ellington, A., and Lambowitz, A.M. 1996. Analysis of the CYT-18 protein binding site at the junction of stacked helices in a group I intron RNA by quantitative binding assays and in vitro selection. *J. Mol. Biol.* **261:** 23–42.

Strobel, S.A. and 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., and Moncoeur, E. 1998. Complementary sets of non-canonical base pairs mediate RNA helix packing in the group I intron active site. *Nat. Struct. Biol.* **5:** 60–65.

Suh, S.O., Jones, K.G., and Blackwell, M. 1999. A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetous fungus: Evidence for a new major class of Group I introns. *J. Mol. Evol.* **48:** 493–500.

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

Waldsich, C., Grossberger, R., and Schroeder, R. 2002. RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron in vivo. *Genes & Dev.* **16:** 2300–2312.

Wallweber, G.J., Mohr, S., Rennard, R., Caprara, M.G., and Lambowitz, A.M. 1997. Characterization of *Neurospora* mitochondrial group I introns reveals different CYT-18 dependent and independent splicing strategies and an alternative 3′ splice site for an intron ORF. *RNA* **3:** 114–131.

Warnecke, J.M., Sontheimer, E.J., Piccirilli, J.A., and Hartmann, R.K. 2000. Active site constraints in the hydrolysis reaction catalyzed by bacterial RNase P: Analysis of precursor tRNAs with a single 3′-S-phosphorothiolate internucleotide linkage. *Nucleic Acids Res.* **28:** 720–727.

Webb, A.E., Rose, M.A., Westhof, E., and Weeks, K.M. 2001. Protein-dependent transition states for ribonucleoprotein assembly. *J. Mol. Biol.* **309:** 1087–1100.

Weeks, K.M. and Cech, T.R. 1995. Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5′ splice site domain. *Cell* **82:** 221–230.

Weeks, K.M. and Cech, T.R. 1996. Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science* **271:** 345–348.

Zarrinkar, P.P. and Williamson, J.R. 1994. Kinetic intermediates in RNA folding. *Science* **265:** 918–924.

Zaug, A.J., Grabowski, P.J., and Cech, T.R. 1983. Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. *Nature* **301:** 578–583.

Zaug, A.J., McEvoy, M.M., and 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.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31:** 3406–3415.