The EMBO Journal Vol.16 No.22 pp.6835–6848, 1997

Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA

Huatao Guo[1], Steven Zimmerly[2], Philip S.Perlman[3] and Alan M.Lambowitz[1,4]

Departments of Molecular Genetics, Biochemistry, and Medical Biochemistry, The Ohio State University, 484 West Twelfth Avenue, Columbus, OH 43210-1292 and [3]Department of Molecular Biology and Oncology, University of Texas, Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-9148, USA

[1]Present address: Institute of Cellular and Molecular Biology, Departments of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712, USA
[2]Present address: Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4

[4]Corresponding author
e-mail: lambowitz.1@osu.edu

Group II introns use intron-encoded reverse transcriptase, maturase and DNA endonuclease activities for site-specific insertion into DNA. Remarkably, the endonucleases are ribonucleoprotein complexes in which the excised intron RNA cleaves the sense strand of the recipient DNA by reverse splicing, while the intron-encoded protein cleaves the antisense strand. Here, studies with the yeast group II intron aI2 indicate that both the RNA and protein components of the endonuclease contribute to recognition of an ~30 bp DNA target site. Our results lead to a model in which the protein component first recognizes specific nucleotides in the most distal 5′ exon region of the DNA target site (E2–21 to –11). Binding of the protein then leads to DNA unwinding, enabling the intron RNA to base pair to a 13 nucleotide DNA sequence (E2–12 to E3+1) for reverse splicing. Antisense-strand cleavage requires additional interactions of the protein with the 3′ exon DNA (E3+1 to +10). Our results show how enzymes can use RNA and protein subunits cooperatively to recognize specific sequences in double-stranded DNA.

Keywords: DNA endonuclease/gene therapy/recombination/reverse transcriptase/ribozyme

Introduction

Group II introns are of interest both as catalytic RNAs that may be the progenitors of spliceosomal introns and as mobile genetic elements that use novel RNA-based mechanisms for site-specific DNA cleavage and insertion (Lambowitz and Belfort, 1993; Michel and Ferat, 1995; Curcio and Belfort, 1996). Yeast mtDNA contains two mobile group II introns, aI1 and aI2, which are found in the *COX1* gene encoding subunit I of cytochrome oxidase. During crosses between strains that contain and lack the introns, aI1 and aI2 insert site-specifically into the intronless *COX1* genes, a process referred to as 'homing' (Lazowska *et al.*, 1994; Moran *et al.*, 1995). Homing occurs at high frequency by a duplicative transposition mechanism, so that virtually all of the progeny alleles contain the intron. aI1 has also been shown to transpose to ectopic sites at low frequency (Mueller *et al.*, 1993; Schmidt *et al.*, 1994).

The mobility of group II introns is mediated by multifunctional reverse transcriptases (RTs) encoded within the introns. The group II intron RTs are related to those encoded by the non-LTR family of retroelements, a diverse group that includes the R2 elements of insects and the abundant human LINE elements (Eickbush, 1994). The RT proteins encoded by aI1 and aI2 are translated in-frame with the upstream *COX1* exon and appear to be proteolytically processed to a mature form (68 and 62 kDa for the aI1 and aI2 proteins, respectively; Carignani *et al.*, 1983; Moran *et al.*, 1994). In addition to the RT activity, both the aI1 and aI2 proteins function as maturases to promote splicing of the intron in which they are encoded, presumably by helping the intron RNA fold into the catalytically active structure (Carignani *et al.*, 1983; Moran *et al.*, 1994). Both proteins also contain a C-terminal Zn^{2+}-finger-like region (Zn domain), which contributes to a DNA endonuclease activity involved in intron mobility (Zimmerly *et al.*, 1995a,b; Yang *et al.*, 1996).

Recent studies have shown that the homing of aI1 and aI2 occurs by target DNA-primed reverse transcription (Zimmerly *et al.*, 1995a; Eskes *et al.*, 1997). The process is initiated by the intron-encoded DNA endonuclease, which site-specifically cleaves at the intron insertion site in the recipient DNA. The 3′ end of the cleaved antisense strand is then used as a primer for reverse transcription of the intron RNA. Analogous target DNA-primed reverse transcription mechanisms are used by nuclear non-LTR retrotransposons, which encode phylogenetically related RTs (Luan *et al.*, 1993; Feng *et al.*, 1996). However, while the non-LTR retrotransposons rely on conventional protein endonucleases with various degrees of specificity to cleave the target DNA, the group II intron endonucleases are ribonucleoprotein (RNP) complexes in which both the intron-encoded protein and the excised intron RNA function catalytically in DNA cleavage (Zimmerly *et al.*, 1995b). For both aI1 and aI2, the intron RNA cleaves the sense strand of the recipient DNA at the intron insertion site by a partial or complete reverse splicing reaction, while the intron-encoded protein cleaves the antisense strand at position +10 of the 3′ exon (Zimmerly *et al.*, 1995a,b; Yang *et al.*, 1996). Following DNA cleavage, the template for reverse transcription is either the unspliced pre-mRNA containing the intron or the intron RNA that had reverse-spliced into the DNA target site. Mobility may be completed by insertion of a complete cDNA copy of the intron or by using a partial or complete cDNA copy of the intron to initiate double-stranded break repair on

an intron-containing *COX1* allele (Eskes *et al.*, 1997; R.Eskes, S.Zimmerly, C.Mazurek, A.M.Lambowitz and P.S.Perlman, in preparation).

Initial characterization of the aI2 endonuclease raised the possibility that both its RNA and protein components contribute to recognition of the DNA target site (Zimmerly *et al.*, 1995b). The DNA target site for aI2 was found to include the 5′ exon sequences IBS1 and IBS2 (intron binding sites 1 and 2), which are complementary to intron sequences EBS1 and EBS2 (exon binding sites 1 and 2) located in different regions of domain I of the intron RNA. These EBS–IBS interactions are required during RNA splicing and also for reverse splicing of group II introns into RNA substrates (Michel and Ferat, 1995). An additional interaction, δ-δ′, between the nucleotide preceding EBS1 in the intron and position +1 of the 3′ exon, appears to play a role in the second step of splicing (Michel and Ferat, 1995). In principle, these base-pairing interactions could also play a role in the reverse splicing of group II introns into DNA target sites. Consistent with this possibility, a single nucleotide change in the IBS1 sequence of the 5′ exon inhibited both aI1 reverse splicing and intron mobility, and those defects were reversed by a compensatory change in the EBS1 sequence of the intron RNA (Eskes *et al.*, 1997).

The ability to change the DNA target site raised the possibility that it might be feasible to systematically modify group II intron endonucleases to target DNA cleavage or intron insertion to preselected sites in DNA genomes, with potentially widespread applications in genetic engineering and gene therapy. Here, we tested this idea by detailed characterization of the DNA target site for the aI2 endonuclease. We show that the aI2 target site for reverse splicing extends from position E2–21 to E3+1, with additional sequences between E3+1 and +10 required for antisense-strand cleavage. A continuous 13 nt stretch of the DNA target site (E2–12 to E3+1), which encompasses the IBS1, IBS2 and δ′ sequences, is recognized primarily by base pairing with the intron RNA, providing a basis for modifying target specificity. The regions of the DNA target site flanking the IBS sequences are presumably recognized by the intron-encoded protein, which appears to play a role in DNA unwinding enabling base pairing of the intron RNA. Our results show how enzymes can use RNA and protein subunits cooperatively to target specific DNA sequences for site-specific cleavage and insertion.

Results

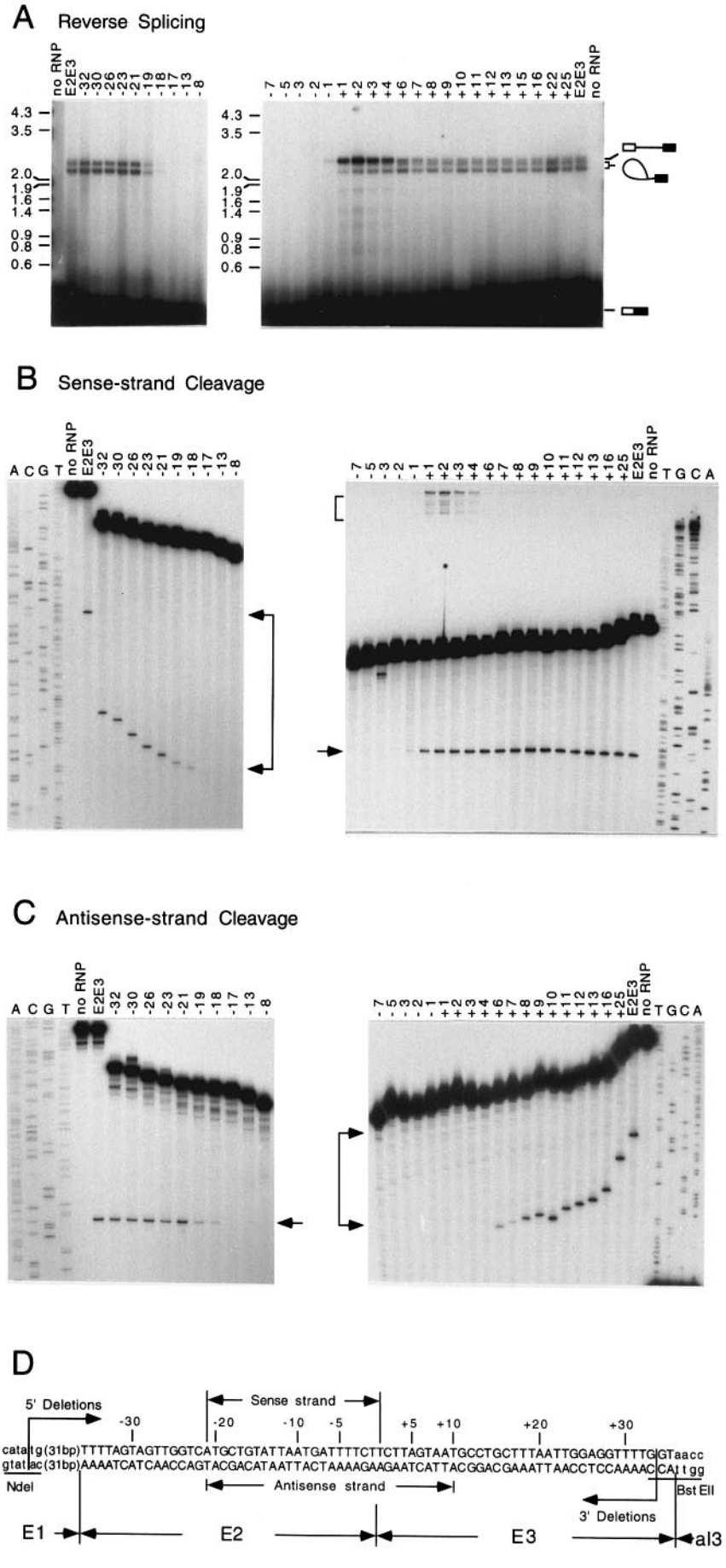
5′- and 3′-boundaries of the aI2 DNA target site

To determine the boundaries of the aI2 target site in the *Saccharomyces cerevisiae COX1* gene, we constructed nested 5′ and 3′ deletions in the target site region by *Bal*31 digestion. The DNA substrates were assayed for reverse splicing and DNA endonuclease activities using mtRNP particles from yeast strain $1°2^{+t}$, which contains wild-type aI2 (Figure 1). For reverse splicing assays, small ^{32}P-labeled DNA substrates containing the target site region were incubated with the mtRNP particles, and the products were analyzed in agarose gels (see Materials and methods). The incubations were carried out for times determined to be in the linear range of the reactions.

Figure 1A shows that reverse splicing with a DNA substrate consisting of *COX1* exons 2 and 3 yielded three product bands. The two lower bands correspond to previously described partially reverse-spliced products containing different forms of aI2 lariat RNA linked to the 3′ exon DNA, and the upper band corresponds to the fully reverse-spliced product containing the intron RNA inserted between the two DNA exons (cf. Zimmerly *et al.*, 1995b; Yang *et al.*, 1996). The identity of the fully reverse-spliced product, which comprised ~7% of the total, was confirmed by additional experiments using 5′ sense-strand labeled DNA substrates (not shown). This product was better resolved than in previous experiments for aI2 because of the longer 5′ exon sequences in the DNA substrate used here (cf. Zimmerly *et al.*, 1995b). The results for the deleted DNA substrates show that the region required for maximal reverse splicing extends from positions E2–20 or –21 to E3+1. The requirement for E2–21 cannot be determined from these constructs due to the replacement of the deleted base pair with the same base pair from the flanking sequence. However, experiments below show that E2–21 contributes to target site recognition. Interestingly, the 3′ deletions E3+1 to +4 showed a marked increase in the amount of fully reverse-spliced product.

To assay DNA cleavage, double-stranded DNA substrates labeled at the 5′ end of the sense or antisense strands were incubated with mtRNP particles, and the cleavage products were analyzed in denaturing 6% polyacrylamide gels. Figure 1B shows that the sense-strand cleavage parallels reverse splicing, as expected, with the required sequences extending from positions E2–21/–20 to E3+1. The proportion of fully reverse-spliced product,

Fig. 1. *Bal*31 deletion analysis of the aI2 DNA target site. (**A**) DNA target site for reverse splicing. Reverse splicing reactions were carried out by incubating $1°2^{+t}$ mtRNP particles with internally labeled wild-type E2E3 or deleted DNA substrates. The DNA substrates were synthesized from pHG3 (wild-type) or *Bal*31-deletion plasmids by PCR in the presence of [α-^{32}P]dTTP, using primers E1-JY and aI3-62. The products of reverse splicing were glyoxalated and analyzed in a 1% agarose gel. Left and right panels show results for 5′ and 3′ deletions, respectively. Schematics to the right show the structures of fully and partially reverse-spliced products and the DNA substrate. Numbers to the left indicate size markers (kb). (**B**) and (**C**) DNA target sites for sense- and antisense-strand cleavage. Wild-type E2E3 or deleted DNA substrates labeled at the 5′ end of the sense or antisense strand were incubated with $1°2^{+t}$ mtRNP particles as in (A), and cleavage products were analyzed in a denaturing 6% polyacrylamide gel, alongside sequencing ladders generated from pHG3 using the same 5′ end-labeled primers. For (B) and the left panel of (C), DNA substrates were synthesized by PCR using the primers E1-JY and aI3-62; for the right panel of (C), the DNA primers were E2(161) and aI3-40. Cleavage products are indicated by arrows. The full reverse splicing products in (B) are indicated by brackets near the top of the gel; the multiple bands presumably reflect degradation of the inserted intron RNA. The band migrating just below the substrate in lane –3 of (B) is present with substrate alone (not shown) and is presumably an artifact of substrate preparation. The antisense-strand cleavage product for the E3+10 substrate in (C) migrates faster than those around it because the construct has a deletion of the GT of the flanking sequence to avoid replacing E3+11 with the same nucleotide. For all panels, lane headings indicate the position to which the deletion is made. Control lanes (no RNP) show the E2E3 DNA substrate incubated in the absence of mtRNP particles. (**D**) Diagram of *Bal*31 deletion constructs. 5′ deletions begin at the *Nde*I site in E1 and extend downstream; 3′ deletions begin at the *Bst*EII site in E3 and extend upstream. Upper-case letters indicate *COX1* exon 2 and 3 sequences, and lower-case letters indicate exon 1 and aI3 sequences. The DNA target site for maximal reverse splicing and sense-strand cleavage extends from E2–21 to E3+1, and the DNA target site for maximal antisense-strand cleavage extends from E2–21 to E3+10 (see text).


A Reverse Splicing
no RNP
E2E3
-32
-30
-26
-23
-21
-19
-18
-17
-13
-8
-7
-5
-3
-2
-1
+1
+2
+3
+4
+6
+7
+8
+9
+10
+11
+12
+13
+15
+16
+22
+25
4.3
3.5
2.0
1.9
1.6
1.4
0.9
0.8
0.6
B Sense-strand Cleavage
A C G T
T G C A
C Antisense-strand Cleavage
D
5' Deletions
Sense strand
Antisense strand
3' Deletions
-30
-20
-10
-5
+5
+10
+20
+30
catatg(31bp)TTTTAGTAGTTGGTCATGCTGTATTAATGATTTTCTTCTTAGTAATGCCTGCTTTAATTGGAGGTTTTGTGTaacc
gtatac(31bp)AAAATCATCAACCAGTACGACATAATTACTAAAAGAAGAATCATTACGGACGAAATTAACCTCCAAAACACAttgg
NdeI
BstEII
E1
E2
E3
al3

which migrates above the DNA substrate in polyacrylamide gels (bracket in Figure 1B, right), was again seen to increase for the 3′ deletions E3+1 to +4.

For antisense-strand cleavage, the 5′ boundary was the same as for sense-strand cleavage, but additional 3′ exon sequences were required. The 3′ boundary for maximal antisense cleavage was E3+10. The E3+9 and +8 substrates showed small decreases confirmed below by single nucleotide substitutions. The amount of cleavage dropped substantially for the E3+7 substrate, but increased for the E3+6 substrate, where the sequence GTAAT at E3+6 to +10 was replaced by GGTAA, before becoming undetectable for the E3+4 deletion. Notably, antisense-strand cleavage remained at position +10, even for the E3+6 to +10 substrates, where the normal cleavage-site sequence was replaced with other sequences. Because of the different target site requirements for sense- and antisense-strand cleavage, the 3′ deletions E3+1 to +4 underwent efficient reverse splicing, but no antisense-strand cleavage.

Reconstitution of the aI2 endonuclease with in vitro-synthesized intron RNA

The sense strand of the DNA target site contains the exon sequences IBS1, IBS2 and δ′, which can potentially base pair with the intron sequences EBS1, EBS2 and δ (see Introduction). In the case of aI2, the IBS and δ′ sequences could extend from E2–13 to E3+1 (Figure 2C). To investigate whether these base-pairing interactions are required for recognition of the DNA target site, we developed a procedure for reconstituting the aI2 endonuclease with *in vitro*-synthesized intron RNA. This method allows much more rapid testing of intron modifications than mtDNA transformation used previously (Eskes *et al.*, 1997). The source of aI2 protein for reconstitution was the yeast mutant $1°2^{\Delta D5}$, which has a deletion of intron domain V and is unable to splice aI2 (Moran *et al.*, 1995). This mutant overproduces aI2 protein from unspliced precursor RNA, and the protein appears to be more loosely associated with mtRNP particles than in the wild-type strain (S.Zimmerly, J.V.Moran, P.S.Perlman and A.M.Lambowitz, in preparation). Reconstitution was carried out by mixing $1°2^{\Delta D5}$ mtRNP particles with aI2 RNA obtained by self-splicing of an *in vitro*-transcribed precursor RNA (see Materials and methods). The $1°2^{\Delta D5}$ RNA remains in the preparations but is inactive. Figure 2A shows that the reconstituted RNP particles had levels of reverse splicing activity comparable with those of wild-type $1°2^{+t}$ mtRNP particles (lanes 2 and 5), whereas the aI2 RNA or $1°2^{\Delta D5}$ mtRNP particles alone had no detectable reverse splicing activity (lanes 3 and 4). The reconstituted RNPs also had wild-type levels of sense- and antisense-strand cleavage in DNA endonuclease assays (not shown).

Reverse splicing requires base pairing between the IBS1 sequence in the DNA target site and the EBS1 sequence in the intron RNA

To test the extent to which the IBS sequences in the DNA target are recognized by base pairing with the EBS sequences in the intron RNA, we used the reconstitution system to make RNP particles containing derivatives of aI2 that have the EBS1 and/or EBS2 sequences ordinarily found in aI1. Figure 2B shows that wild-type aI2 endonuclease reverse spliced into the unmodified DNA target site, but not into altered target sites that contain mismatched IBS1 and/or IBS2 sequences (lanes 1–4). By contrast, a modified endonuclease in which aI2 contains the aI1 EBS1 sequence no longer recognized the normal DNA target site, but did recognize a modified target site that contains the complementary IBS1 sequence of aI1 (lanes 5–8). These findings show that the IBS1(DNA) sequence is in fact recognized by base pairing with the EBS1 sequence in the intron RNA. In other experiments testing different substitutions, we were able to change all six base pairs of the EBS1/IBS1 pairing to a different sequence without decreasing the efficiency of the reverse splicing reaction. However, some combinations of nucleotide changes were suboptimal (C.Beall and A.M. Lambowitz, unpublished data).

The IBS2 sequence in the DNA target site is recognized primarily by base pairing with the intron RNA

The situation for the IBS2/EBS2 interaction was more complicated. Figure 2C (top, right) shows that the IBS2/EBS2 interaction for aI2 could potentially comprise as many as seven base pairs. An aI2 construct in which both the EBS1–IBS1 and EBS2–IBS2 pairings were changed to those of aI1 self-spliced at the wild-type rate under conditions favoring the RNA-only reaction (not shown; see Materials and methods for self-splicing conditions). However, reconstitution mixtures containing this RNA did not reverse splice into any DNA target site, including one that contained the complementary aI1 IBS1 and IBS2 sequences (Figure 2B, lanes 9–12). Likewise, an aI2 derivative in which only EBS2 was changed to that of aI1 also self-spliced at the wild-type rate, but gave no detectable reverse splicing into any of the DNA target sites (not shown). The inability of the modified introns to reverse splice into DNA substrates with complementary IBS2 sequences could reflect that the aI2 IBS2(DNA) sequence is recognized in part by the aI2 protein.

To identify which regions of IBS2 are important for the reaction, we tested derivatives of the aI2 DNA target site in which each nucleotide in IBS2 was changed to its complement. All of those changes strongly inhibited reverse splicing with wild-type $1°2^{+t}$ mtRNP particles (<20% of wild-type), except for the change at position E2–7, which had little effect on the level of reverse splicing (Figure 3A). We were somewhat surprised by the low level of reverse splicing with the E2–8C substrate (6% of wild-type), since a strain (GII-O) with a mismatched A at this position (E2–8A) had been found to support high levels of homing *in vivo* and was used as a standard recipient in crosses (Moran *et al.*, 1995). An additional experiment showed that reverse splicing into the GII-O (E2–8A) target site occurred at 29% of the wild-type level (not shown). Assuming this is also true *in vivo*, this level of reverse splicing activity must not be limiting for efficient homing. Further experiments examining the E2–7 position showed that any of the four nucleotide residues could be tolerated almost equally well at this position for reverse splicing (88–112% of the wild-type level; not shown). All of the IBS2 changes that inhibited reverse splicing also inhibited antisense-strand cleavage (not shown).

Next, each of the IBS2 positions at which nucleotide

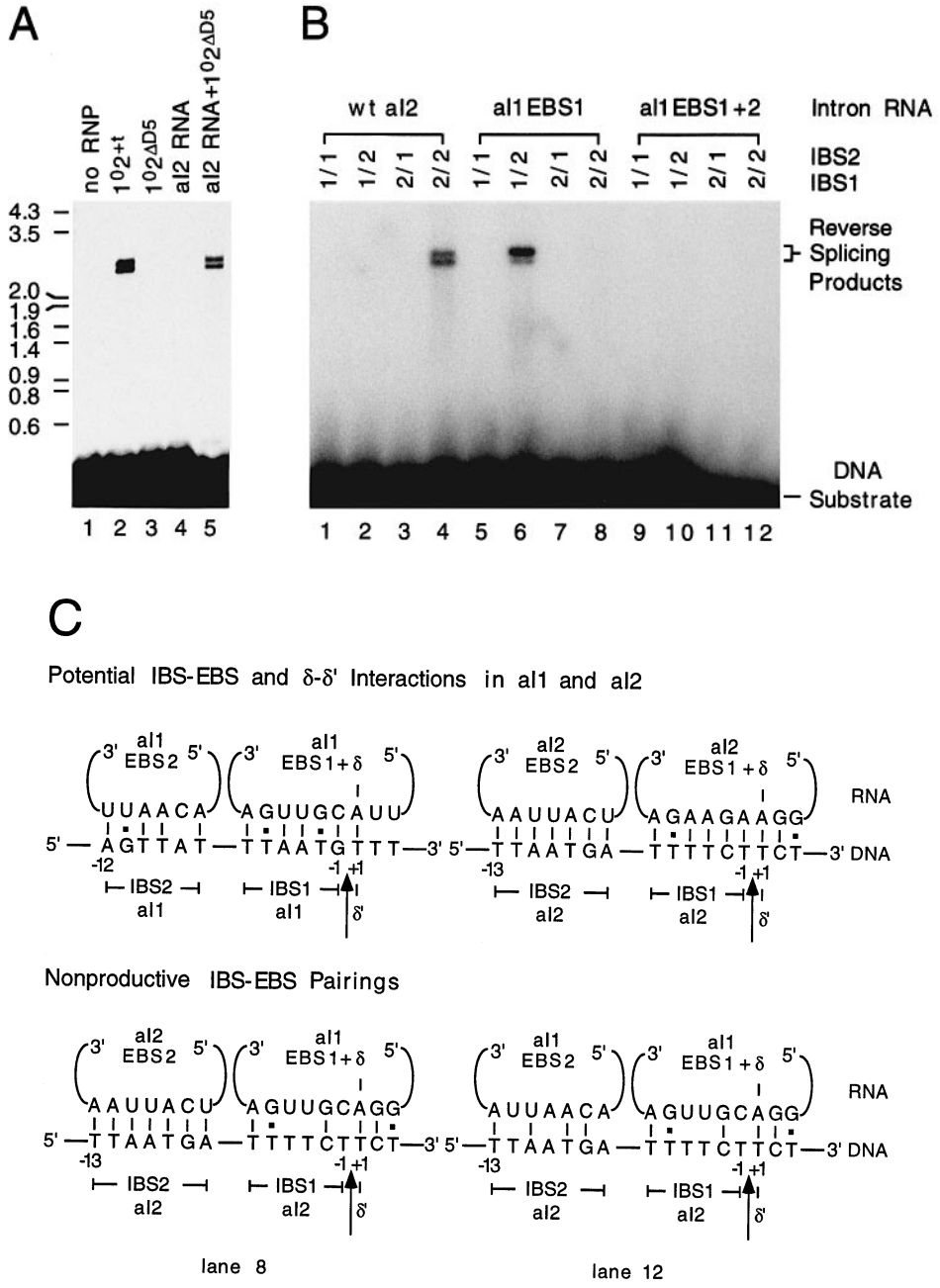


Fig. 2. Reconstitution of the aI2 endonuclease and recognition of IBS1 in the DNA target site by base pairing with EBS1 in the intron RNA. (**A**) Reconstitution of the aI2 endonuclease. The aI2 endonuclease was reconstituted by mixing $1°2^{\Delta D5}$ RNP particles with aI2 RNA obtained by self-splicing of the *in vitro* transcript of pJVM4 cleaved with *Bst*EII. Reverse splicing with $1°2^{+t}$ mtRNP particles (lane 2) or reconstituted RNP particles (lane 5) was assayed using internally labeled E2E3 DNA substrate synthesized from pE2E3(161) by PCR using the primers SK and KS. Control lanes show the DNA substrate incubated in the absence of RNP particles (lane 1) or with $1°2^{\Delta D5}$ mtRNP particles or aI2 RNA by themselves (lanes 3 and 4). Reverse splicing products were denatured with glyoxal and analyzed in a 1% agarose gel. Size markers (kb) are indicated to the left. (**B**) The EBS1–IBS1 pairing is required for reverse splicing into the DNA target site. aI2 endonucleases were reconstituted with wild-type aI2 or derivatives containing different combinations of aI1 and aI2 EBS1 and EBS2 sequences. The endonucleases were then assayed for reverse splicing with internally labeled E2E3 DNA substrate or modified DNA substrates containing different combinations of aI1 and aI2 IBS1 and IBS2 sequences. The DNA substrates were synthesized by PCR with primers E2(161) and HG1 from synthetic oligonucleotides having the wild-type E2E3 sequence [E2E3(161)] or modified oligonucleotides having aI1 sequences substituted for IBS1 (E2–6 to –1), IBS2 (E2–12 to –7) or both (E2–12 to –1). The products were analyzed as above. (**C**) Schematic showing the normal aI1 and aI2 EBS–IBS and δ–δ′ pairings and two nonproductive combinations corresponding to lanes 8 and 12 of (B). The arrow indicates the junction between exons 2 and 3.

substitutions were found to be inhibitory was tested for reverse splicing with RNP particles reconstituted with a modified intron RNA that had a complementary change at the corresponding position of EBS2 (note: EBS positions are numbered according to the IBS position with which they pair). E2–9 and –10 were tested together, since we had preliminary evidence that these positions are recognized by base pairing. Figure 3B shows that reverse splicing was restored to wild-type levels by the compensatory changes at EBS2 positions –8 and –9/–10 (lanes 8 and 9 compared with lane 1). The compensatory changes at –11 and –12 increased reverse splicing, but not to wild-

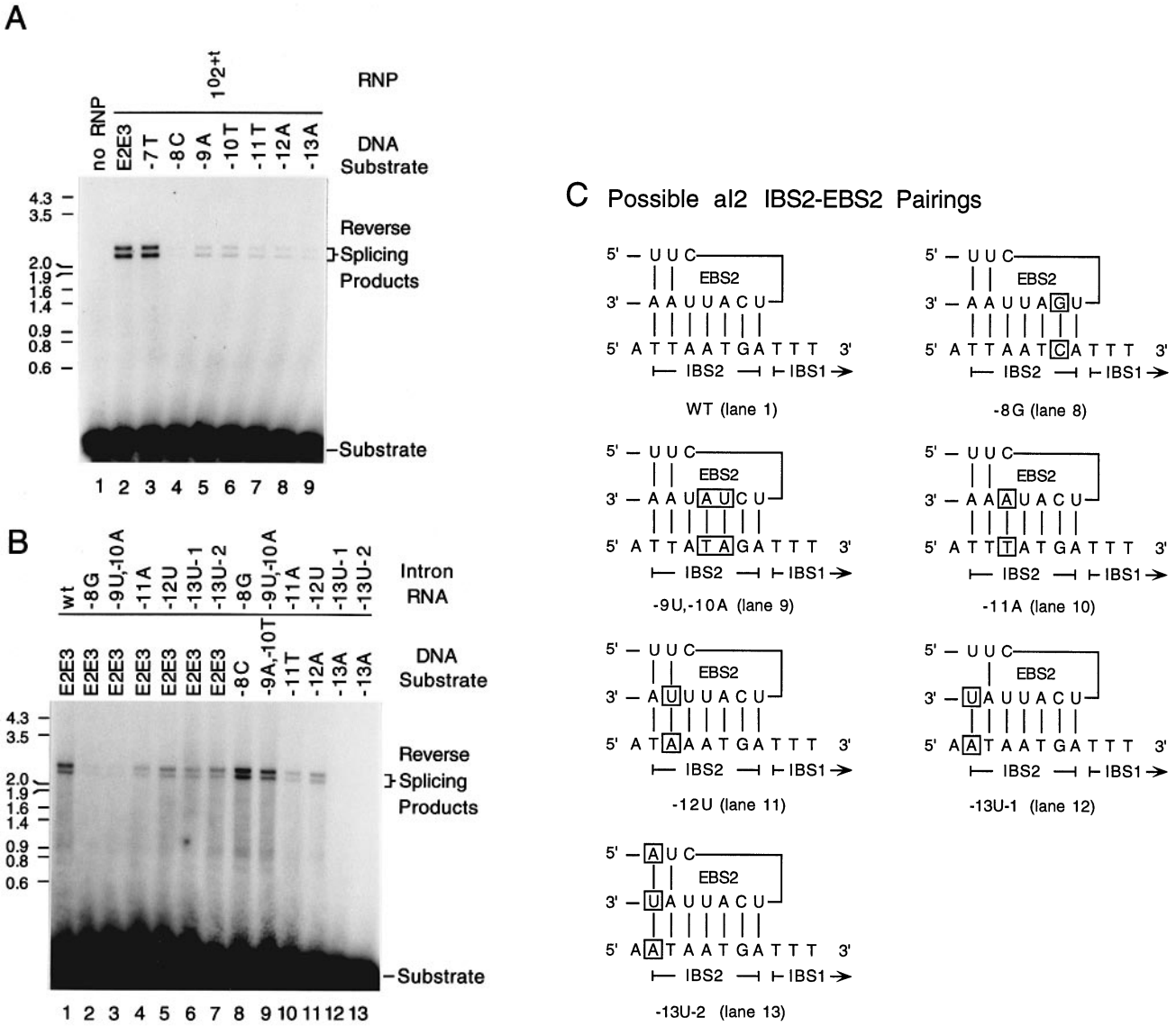


Fig. 3. IBS2 in the DNA target site is recognized primarily by base pairing with EBS2 in the intron RNA. (**A**) $1°2^{+t}$ mtRNP particles were assayed for reverse splicing with internally labeled E2E3 DNA substrate or with modified DNA substrates having the indicated single nucleotide changes at E2–7 to –13. The DNA substrates were synthesized by PCR with the primers E2(161) and HG1 from synthetic oligonucleotides E2E3(161), which has the wild-type E2E3 sequence, or from modified E2E3(161) oligonucleotides having the indicated nucleotide substitutions. Lane 1 (no RNP) shows the wild-type E2E3 DNA substrate incubated in the absence of RNP particles. Reverse splicing products were denatured with glyoxal and analyzed in a 1% agarose gel. Size markers (kb) are indicated to the left. (**B**) aI2 endonucleases were reconstituted with wild-type aI2 or modified introns having single or double nucleotide changes at EBS2 positions that potentially base pair with E2–8 to –13 of the DNA target site. The endonucleases were assayed for reverse splicing with internally labeled wild-type E2E3 or modified DNA substrates having the indicated changes at E2–8 to –13 (see above). Lane 1, reverse splicing of endonuclease containing wild-type aI2 RNA into the wild-type E2E3 DNA substrate; lanes 2–7, reverse splicing of endonucleases containing modified aI2 RNAs into the E2E3 DNA substrate; lanes 8–13, reverse splicing of endonucleases containing aI2 RNAs that restore base-pairing with the modified DNA substrates. The reverse-spliced products were analyzed as above. (**C**) Schematics showing IBS2–EBS2 pairings for the indicated lanes of (B). Mutated nucleotides are boxed.

type levels (lanes 10 and 11), while the compensatory change at –13 gave no detectable reaction (lane 12). We noticed that introducing the compensatory nucleotide change at EBS2–13 could disrupt a potential base pair within the intron RNA (Figure 3C). However, an additional construct, which retained the compensatory change at EBS2–13 while restoring that base pair (–13U–2), also failed to reverse splice into the DNA target site modified at position –13 (Figure 3B, lane 13). In all of the above cases where complete restoration of reverse splicing was not observed, the modified introns were confirmed to self-splice at rates equal to or somewhat higher than wild-type aI2, indicating that these EBS2 changes did not impair the catalytic activity of the intron RNA, at least under self-splicing conditions.

Together, the above findings suggest that while most of IBS2 is recognized by base pairing with the intron RNA, E2–13 may be recognized by the aI2 protein and E2–11 and –12 may be recognized in part by the aI2 protein (see Discussion). Consistent with this inference, the intron mutants aI2-EBS2–13U–1 and –13U–2 remained active with the wild-type DNA target site, despite the mispairing at the E2–13 position (53 and 65% activity; Figure 3B, lanes 6 and 7). The mutant introns with the changes at

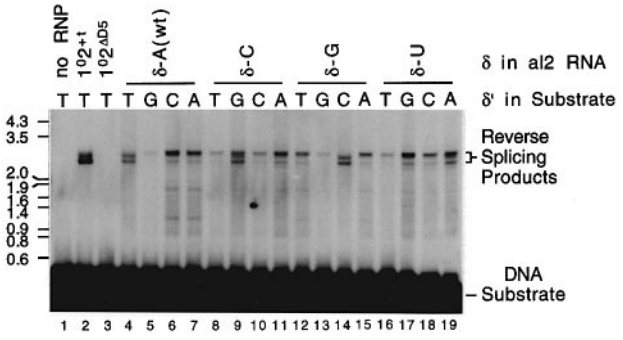


Fig. 4. The δ–δ′ interaction contributes to DNA target site recognition. aI2 endonucleases were reconstituted with wild-type aI2 (where δ is A), or modified intron RNAs having single nucleotide changes (T, C or G) at the δ position. The endonucleases were assayed for reverse splicing with the internally labeled wild-type E2E3 DNA substrate (where δ′ is T), or with modified DNA substrates in which δ′ was changed to G, C or A. DNA substrates were synthesized from plasmids pHG3 (wild-type), pHG3-δ′A, pHG3-δ′C, or pHG3-δ′G by PCR with primers E1-JY and aI3-62. Control lanes show the wild-type E2E3 DNA substrate incubated in the absence of RNP particles (lane 1) or with $1°2^{+t}$ or $1°2^{\Delta D5}$ mtRNP particles (lanes 2 and 3). The reverse-spliced products were denatured with glyoxal and analyzed in a 1% agarose gel. Sizes markers (kb) are indicated to the left.

EBS2 positions –11 and –12 also remained at least somewhat active with the wild-type DNA target sequence, despite the mispairings at these positions (40 and 73% activity, respectively; lanes 4 and 5). The inability to switch the aI1 and aI2 EBS2 and IBS2 sequences in the experiment of Figure 2 may reflect in part a requirement for aI2 protein recognition at E2–11 and –12, since E2–13 did not change in that experiment.

The δ–δ′ interaction contributes to reverse splicing into DNA

The requirement for the δ–δ′ interaction, which involves the intron nucleotide preceding EBS1 and the nucleotide at E3+1 in the sense strand of the DNA target site (see Figure 2C), was investigated by testing all combinations of nucleotide changes at these positions. Figure 4 shows that almost all the combinations of δ and δ′ allowed some reverse splicing, but the ratio of fully to partially reverse spliced products appeared skewed for many combinations. A clear effect of base pairing was evident, since for each mutation of the δ nucleotide of the intron, the closest match to wild-type activity was obtained for a DNA substrate that restored a Watson–Crick or wobble base pair at δ′ (lanes 4, 9, 14, 17 and 19). However, other effects must be operating since G–C and A–T/U pairs gave different results depending on which nucleotide residue of the pair was in the intron RNA (cf., lanes 9 and 14; 4 and 19). Most combinations that could not form conventional base pairs appeared to increase the proportion of fully reverse spliced products. Notably, the strongest inhibition of reverse splicing occurred for the combinations involving a G at position E3+1 and a purine at the δ position in the intron (~20% of wild-type; lanes 5 and 13). The inhibitory effect of G at E3+1 explains the requirement for this position found with the 3′ *Bal*31-deletion construct E3–1, where a G from the flanking sequence was brought into the +1 position (Figure 1). It also explains why an intron-containing DNA is not a good substrate for reverse splicing (Zimmerly *et al.*, 1995b), since the first nucleotide of the intron is G.

Sequence requirements in regions flanking the IBS and δ′ sequences

The regions of the DNA target site flanking the IBS and δ′ sequences are putatively recognized by the protein component of the endonuclease. The experiments above already suggested this was the case for position E2–13. To test the sequence requirements for positions E2–21 to –14, reverse splicing of wild-type $1°2^{+t}$ mtRNP particles was tested with DNA substrates in which each of these positions was changed to a mixture of the three other bases (Figure 5A). This procedure enabled us to rapidly sample all possible substitutions at a given position. The strongest effect was observed for substitutions at positions E2–15 and –18, where reverse splicing was reduced to 9 and 3% of the wild-type level, respectively. The substitutions at E2–16 and –19 had moderate effects (23 and 31% wild-type activity, respectively), while substitutions at positions E2–21, –20, –17 or –14 had little or no effect (67–115% wild-type activity). All these mutant DNA substrates showed parallel effects on antisense-strand cleavage (not shown), consistent with the finding from the *Bal*31 deletion constructs that the 5′ boundary of the target site for antisense-strand cleavage extends to E2–21 (Figure 1). In accord with the results for single nucleotide substitutions, an additional experiment (not shown), in which substrates capable of undergoing full reverse splicing were selected from a pool of DNA substrates doped 50% at positions E2–33 to –14, showed that the most critical positions are E2–15 and –18 (0 mutations in 63 products analyzed), followed by positions E2–16 and –19 (3/63 and 7/63 mutations, respectively). Other positions showed less conservation (14 to 26 mutations in 63 products analyzed).

The sequence requirements for positions E3+1 to +10 for antisense-strand cleavage were also analyzed by using a series of DNA substrates in which each position was individually changed to a mixture of the three other bases. The greatest inhibition was observed for changes at E3+1, +4 and +6 (38, 36 and 30% wild-type activity, respectively), whereas changes at other positions had less effect (45 to 139% wild-type; Figure 5B). The inhibition caused by nucleotide substitutions at E3+1 could reflect disruption of the δ-δ′ interaction and/or interaction of the protein perhaps with the opposite strand of the DNA target site at this position. The modest effect of the nucleotide substitutions at E3+10 (33% inhibition) was observed in three independent experiments and is consistent with the results for the *Bal*31-deletion construct E3+9 (see Figure 1C). None of these 3′ exon mutations inhibited reverse splicing, although the +1 and +2 mutations appeared to increase the proportion of fully reverse spliced product (not shown). The finding that 3′ *Bal*31 deletions to position E3+4 completely inhibited antisense-strand cleavage presumably reflects the cumulative effect of multiple mutations between E3+6 and +10.

Interaction of the protein component with the distal 5′ exon region of the DNA target site is required for DNA unwinding

Base pairing of the intron RNA to sequences in double-stranded DNA substrates necessarily requires DNA

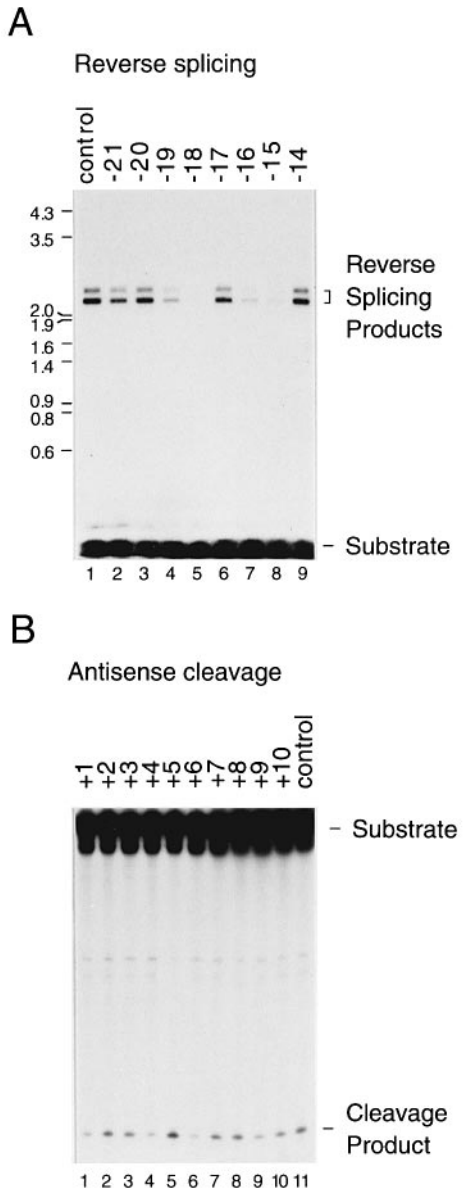


Fig. 5. Nucleotides in the aI2 target site putatively recognized by the protein component of the endonuclease. (**A**) Effect of mutations at E2–21 to –14 on reverse splicing. $1°2^{+t}$ mtRNP particles were assayed for reverse splicing with internally labeled wild-type E2E3 DNA substrate (control; lane 1) or mutant DNA substrates having a mixture of the three non-wild-type base pairs substituted at the indicated positions (lanes 2–9). The reverse-spliced products were glyoxalated and analyzed in a 1% agarose gel. Size markers (kb) are indicated to the left. (**B**) Effect of mutations at E3+1 to +10 on antisense-strand cleavage. $1°2^{+t}$ mtRNP particles were assayed for antisense-strand cleavage with 5′ end-labeled wild-type E2E3 DNA substrate (control; lane 11) or mutant DNA substrates having a mixture of the three non-wild-type base pairs at the indicated positions (lanes 1–10). Cleavage products were analyzed in a denaturing 10% polyacrylamide gel.

unwinding. To identify regions of the DNA target site required for DNA unwinding, we compared reverse splicing into single- and double-stranded DNA substrates in which different portions of the aI2 target site were placed in an unrelated sequence context in the *Neurospora crassa cyt-18* gene. The assumption was that regions required for DNA unwinding would be dispensable for reverse splicing into single-stranded, but not double-stranded DNA. The *cyt-18* sequence by itself gave no detectable reverse splicing as either single- or double-stranded DNA (Figure 6A, lane 3 and 6B, lane 4).

Figure 6A shows that insertion of E2–21 to E3+1 into the double-stranded *cyt-18* substrate was sufficient for maximal reverse splicing, confirming the minimal DNA target site for reverse splicing defined by the *Bal*31 deletions (Figure 1). In fact, the levels of reverse splicing with the E2–21 to E3+1 or +3 insertions were higher than that with the normal E2E3 substrate (cf. lanes 2, 4 and 5), consistent with the 3′ *Bal*31 deletions (Figure 1A), which showed that replacement of the normal 3′ exon sequences leads to elevated full reverse splicing.

Reverse splicing into single-stranded *cyt-18*-derived substrates was compared with both single-stranded E2E3 substrate and with a smaller substrate (SZ4) that is closer in size to the *cyt-18* substrates (see Figure 6D). Figure 6B shows that the insertion of E2–21 to E3+1 into the single-stranded *cyt-18* substrate gave some reverse splicing (lane 12), but additional sequence to E3+3 was required for maximal reverse splicing (lane 13). We interpret these findings to indicate that interaction of the protein with the missing DNA strand is required for efficient reverse splicing into the E2–21 to E3+1 sequence, but that the absence of this interaction in single-stranded DNA can be compensated for by a more stable δ–δ′ interaction, which in the case of aI2 could extend for the first three base pairs of E3 (see Figure 2C).

The most significant difference between the single- and double-stranded *cyt-18* substrates was the dispensability of positions E2–21 to –13 for reverse splicing into single-stranded DNA. Thus, substrates containing only E2–13 to E3+3 or E2–12 to E3+3, the region recognized primarily by the intron RNA, gave substantial reverse splicing as single-stranded DNA (82–86% of SZ4 DNA; Figure 6B, lanes 11 and 15), but no detectable reverse splicing as double-stranded DNA (Figure 6A; lanes 7 and 9). We confirmed by RT–PCR and sequencing that reverse splicing into the single-stranded E2–12 to E3+3 substrate occurred at the expected position (not shown). Additional single-stranded DNA substrates containing only IBS2 and IBS1 (–13 to –1 or –12 to–1), IBS2, IBS1 and E3+1 (–13 to +1), or IBS1 to E3+3 (–6 to +3) gave much lower levels of reverse splicing (9–23% of SZ4 DNA; Figure 6B, lanes 7, 9, 10 and 14), and those containing only IBS1 (–6 to –1), IBS1 and E3+1 (–6 to +1), or IBS2 (–13 to –7) gave no detectable reverse splicing (<2.5% of SZ4 DNA; Figure 6B, lanes 5, 6 and 8). A control construct, AntiIBS1+2, in which positions E2–13 to –1 of SZ4 were changed to their complements, also gave no detectable reverse splicing (Figure 6B, lane 3). The very low levels of reverse splicing with single-stranded DNAs containing only subsets of IBS1, IBS2 and δ′ indicate that these elements contribute cooperatively to reverse splicing. The finding that positions E2–21 to –13 are dispensable for reverse splicing into single-stranded, but not double-stranded DNA suggests that the interaction of the endonuclease with this region is required for DNA unwinding to permit base pairing of the intron RNA (see Figure 7).

The Zn domain of the aI2 protein is required for DNA unwinding

Comparison of reverse splicing into single- and double-stranded DNA substrates was also employed to identify

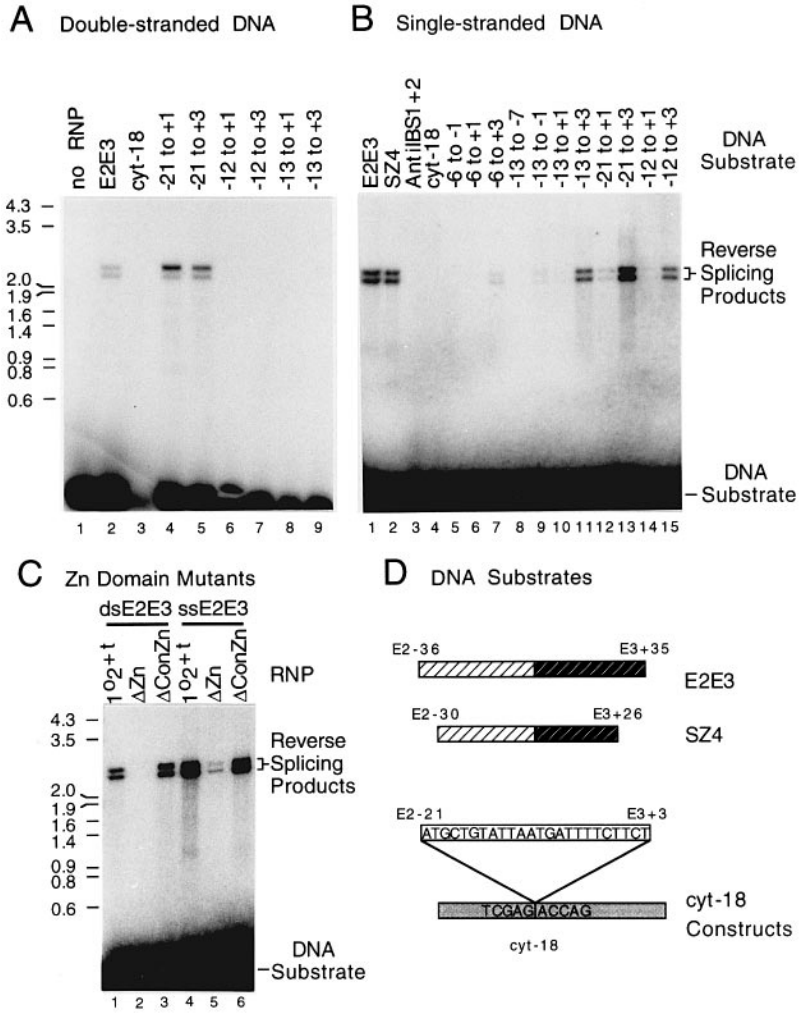


Fig. 6. Different sequence requirements for reverse splicing into single-stranded and double-stranded DNA substrates. Portions of the E2E3 target site were inserted into a segment of the *N.crassa cyt-18* gene to identify sequences minimally required for reverse splicing. (**A**) Reverse splicing with double-stranded DNA substrates. DNA substrates labeled internally with [α-^{32}P]dCTP were generated by PCR from synthetic DNA oligonucleotides E2E3(161), cyt-18 or cyt-18 with the indicated insertions of aI2 target site sequences. Primers E2(161) and HGl were used for the E2E3 substrate, and primers cyt-18-5′ and cyt-18-3′ were used for the *cyt-18* substrates. Reverse splicing reactions with the different DNA substrates were carried out with 1°2^{+t} mtRNP particles, and the products were glyoxalated and analyzed in a 1% agarose gel. Lane 1, wild-type E2E3 substrate incubated in the absence of mtRNP particles. Lanes 2–9, reverse splicing reactions: lane 2, wild-type E2E3 substrate; lane 3, *cyt-18* substrate with no insertions; lanes 4–9, *cyt-18* substrates with the indicated insertions of aI2 target site sequences. (**B**) Reverse splicing reactions with single-stranded DNA substrates. Reverse splicing reactions were carried out with 1°2^{+t} mtRNP particles, and the products were analyzed as above. The single-stranded DNA substrates were gel-purified synthetic oligonucleotides, which had been 3′ end-labeled with [α-^{32}P]dCTP using terminal deoxyribonucleotidyl transferase. Lane 1, full-length E2E3 substrate; lane 2; SZ4, a derivative of E2E3, which has 6 and 9 nt deleted from its 5′ and 3′ ends, respectively (see D); lane 3, AntiIBS1+2, a derivative of SZ4 in which positions E2–13 to –1 were changed to their antisense sequences; lane 4, *cyt-18* substrate with no insertions; lanes 5–15, *cyt-18* substrates with the indicated insertions of aI2 target site sequences. (**C**) Effect of the 1°2$^{\Delta Zn}$ and 1°2$^{\Delta ConZn}$ mutations on reverse splicing into double- and single-stranded DNA substrates. ^{32}P-labeled double- and single-stranded E2E3 substrates were synthesized as described above, and assayed for reverse splicing with mtRNP particles from yeast strains 1°2^{+t}, 1°2$^{\Delta Zn}$ and 1°2$^{\Delta ConZn}$. Lanes 1–3 show results for double-stranded DNAs and lanes 4–6 show results for single-stranded DNAs. (**D**) Schematic diagram of DNA substrates. The E2E3 and SZ4 substrates contain the indicated *COX1* E2 and E3 sequences encompassing the aI2 target site. *cyt-18* substrates contain different portions of the aI2 target site inserted at the indicated position in a segment of the *N.crassa cyt-18* gene (Cherniack *et al.*, 1990). Open hatched box, 5′ exon; filled hatched box, 3′ exon; gray, *cyt-18* sequence.

regions of the aI2 protein required for DNA unwinding. We analyzed previously two C-terminal truncation mutants of the aI2 protein: 1°2$^{\Delta ConZn}$, which has a 67 aa truncation that removes just the region of the Zn domain containing the conserved Zn^{2+}-finger-like and endonuclease motifs, and 1°2$^{\Delta Zn}$, which has a larger truncation (110 aa) that also eliminates the N-terminal region of the Zn domain (Zimmerly *et al.*, 1995b). Both mutants remained active in RNA splicing *in vivo* (Zimmerly *et al.*, 1995b; R.Eskes and P.S.Perlman, unpublished data). The 1°2$^{\Delta Zn}$ mutant was totally blocked for reverse splicing and DNA endonuclease activity for both strands, while the 1°2$^{\Delta ConZn}$ mutant remained active in reverse splicing and sense-strand cleavage, but was totally blocked in antisense-strand cleavage, which is presumably catalyzed by the conserved region of the Zn domain (Zimmerly *et al.*, 1995b; see Figure 7).

Figure 6C shows that mtRNP particles from the 1°2$^{\Delta ConZn}$ mutant reverse-spliced efficiently into both single- and double-stranded E2E3 DNA substrates, as expected (lanes 3 and 6). By contrast, mtRNP particles from the 1°2$^{\Delta Zn}$ mutant showed no detectable reverse

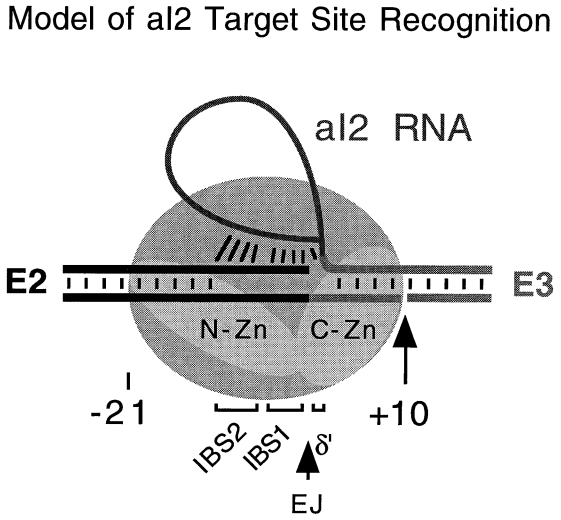


Fig. 7. Model of target site recognition by the aI2 endonuclease. A 13 nt region of the DNA target site containing the IBS2, IBS1 and δ′ sequences is recognized primarily by base pairing with the intron RNA. The N-terminal region of the Zn domain is proposed to interact with nucleotide residues between E2–21 and –11, leading to DNA unwinding. The C-terminal region of the Zn domain, which includes conserved endonuclease and Zn^{2+}-finger-like motifs, interacts with nucleotide residues between E3+1 and +10 for antisense-strand cleavage. EJ (exon junction) indicates the boundary between exons 2 and 3.

splicing into the double-stranded DNA substrate (lane 2), but did show significant, albeit reduced, reverse splicing into the single-stranded DNA substrate (~10% of wild-type; lane 5). Reverse splicing of the $1°2^{\Delta Zn}$ mtRNP particles into double-stranded DNA remained undetectable in darker exposures of Figure 6C (not shown). By contrast, the intron RNA by itself showed no detectable reverse splicing into the single-stranded DNA substrate under these conditions (Zimmerly *et al.*, 1995b and not shown). Together, these findings suggest that the N-terminal region of the Zn domain plays a role in DNA unwinding, since it is required for reverse splicing into double-stranded but not single-stranded DNA. The finding that the ΔZn mutation reduced reverse splicing even with single-stranded DNA may reflect that the aI2 protein contributes to binding both single- and double-stranded DNA substrates in addition to its putative role in DNA unwinding (cf. Zimmerly *et al.*, 1995b).

Discussion

Considered together, our results lead to the model of target site recognition by the aI2 endonuclease shown in Figure 7. The DNA target site for reverse splicing and sense-strand cleavage extends from position E2–21 to E3+1, with additional sequences between E3+1 and +10 required for antisense-strand cleavage. A 13 nt stretch of the DNA target site from E2–12 to E3+1 is recognized primarily by base pairing with the intron RNA, while the flanking regions of the DNA target site are recognized by the intron-encoded protein. The very long DNA target site presumably limits promiscuous insertion of group II introns into other sites in the mtDNA. Because a 13 nt region of the DNA target site is recognized primarily by base pairing with the intron RNA, the specificity of cleavage and insertion can in principle be changed predictably by modifying the intron sequences.

Target site recognition by the RNA and protein components of the endonuclease

To delineate the roles of the RNA and protein components of the endonuclease, we developed a procedure for reconstituting the aI2 endonuclease with *in vitro*-synthesized intron RNA. The reconstitution system enabled us to rapidly introduce modifications into the intron RNA and assess their effects on DNA endonuclease activity. By using this approach, we showed that regions of the DNA target site recognized by the intron RNA include the IBS1, IBS2 and δ′ sequences, which base pair with complementary sequences EBS1, EBS2 and δ in the intron RNA. Nucleotide substitutions in each of these DNA target site sequences were found to affect reverse splicing, and the activity could be restored by compensatory changes in the corresponding element of the intron RNA. The pattern of rescue by compensatory base changes establishes that these interactions primarily involve conventional base pairing and not triplex formation, which obeys different recognition rules (Frank-Kamenetskii and Mirkin, 1995).

The finding that compensatory changes in EBS1 and δ and at some positions in EBS2 restore full activity, while compensatory changes at EBS2–11 and –12 restore only partial activity likely reflects that the latter positions are also recognized by the intron-encoded protein. The protein may interact at these positions with the strand opposite that which base pairs with the intron RNA, or it may interact on the same strand or with the RNA–DNA base pair. Alternative possibilities are as yet uncharacterized interactions with the intron RNA or structural constraints apart from base pairing. The procedures developed for *in vitro* reconstitution of the aI2 endonuclease are also applicable to development of protein-dependent *in vitro* splicing systems, and we have in fact been able to demonstrate protein-dependent *in vitro* splicing under low salt conditions for both aI2 (S.Zimmerly, H.Guo and A.M.Lambowitz, unpublished data) and the *Lactococcus lactis* Ll.ltrB intron (Matsuura *et al.*, 1997).

The regions of the DNA target site flanking the IBS1, IBS2 and δ′ sequences are presumably recognized primarily by the protein component of the endonuclease. The binding of the protein to these regions appears to involve recognition of specific nucleotide residues and could also involve interactions with the phosphodiester backbone. Although unlikely, some recognition in these regions could also involve previously uncharacterized base pairing or triplex interactions with the intron RNA. Analysis of nucleotide substitutions in the region from E2–21 to –13 shows a hierarchy of importance for different positions, with nucleotide substitutions at only five positions (E2–19, –18, –16, –15 and –13) inhibiting reverse splicing by >50%. Similar analysis of E3+1 to +10, the region required for antisense-strand cleavage, also shows only a small number of key positions, with the most important being E3+1, +4 and +6. Single nucleotide changes in the downstream region have only moderate effects on antisense-strand cleavage (30–139% wild-type activity), but there appear to be cumulative effects of multiple mutations. Since the E2–21 to –13 and E3+1 to +10 regions are not strongly conserved in most other group II

intron DNA target sites, it is likely that protein recognition requirements differ, and this has been confirmed by analysis of DNA target sites for two additional group II introns (the yeast aI1 intron and the *L.lactis* L1.ltrB intron; G.Mohr and A.M.Lambowitz, unpublished data).

Role of the intron-encoded protein in DNA unwinding

Base pairing of the intron RNA to the double-stranded DNA target site requires DNA unwinding, which is expected to be promoted by the intron-encoded protein. Experiments comparing reverse splicing into single- and double-stranded DNA substrates show that the upstream flanking region (E2–21 to –13) is not needed for reverse splicing into single-stranded DNA, identifying this as one region where interaction with the protein may contribute to DNA unwinding. The binding of the endonuclease to this region may directly promote DNA unwinding, or it may be a prerequisite for other interactions that lead to DNA unwinding. As discussed above, the protein may also interact with some positions in IBS2 on the strand opposite that which base pairs with the intron RNA, and such interactions could also contribute to DNA unwinding. DNA unwinding by the aI2 endonuclease is not dependent on ATP and presumably uses binding energy for strand displacement. The Zn^{2+}-finger domains of the human transcription factor Sp1 and related proteins have been shown to unwind DNA upon binding (Shi and Berg, 1996).

Analysis of C-terminal truncations of the aI2 protein indicates that the Zn domain can be divided into two functionally distinct regions. The N-terminal region is required for reverse splicing into double-stranded DNA and DNA unwinding, while the C-terminal region, which contains the conserved Zn^{2+}-finger-like and endonuclease motifs, is required only for antisense-strand cleavage (Figure 6C and Zimmerly *et al.*, 1995b). This functional division suggests that the N-terminal region of the Zn domain may interact with the upstream part of the DNA target site (E2–21 to –11), which is also required for reverse splicing and DNA unwinding, while the C-terminal region may interact with the E3 sequences, which are required only for antisense-strand cleavage (see Figure 7). These interactions may occur simultaneously, perhaps with DNA bending, or sequentially by repositioning of the endonuclease after sense-strand cleavage. We note that mtRNP particles from the $1°2^{\Delta ConZn}$ mutant, which lacks the conserved region of the Zn domain containing the Zn^{2+}-finger-like motif, still reverse splice efficiently into double-stranded DNA, indicating that the Zn^{2+}-finger-like motif itself is not required for DNA unwinding.

Sequence of events in target site recognition

Our results suggest a temporal order for the interactions involved in target site recognition. Since the DNA substrate is initially double-stranded, it seems likely that recognition of the target site is initiated by the protein component of the endonuclease, probably by the recognition of specific nucleotide residues in the E2–21 to –11 region. Binding of the protein then leads to DNA unwinding, enabling base pairing of the intron's EBS1, EBS2 and δ sequences to the complementary sequences IBS1, IBS2 and δ′ in the DNA target site. These base pairing interactions are required to position the intron RNA for reverse splicing. Antisense-strand cleavage appears to occur after sense-strand cleavage (Zimmerly *et al.*, 1995b) and is dependent on additional interactions between the protein and E3 sequences that are not required for reverse splicing.

All of the mutations analyzed in the E2–21 to E3+1 region that inhibited sense-strand cleavage also inhibited antisense-strand cleavage (see also Zimmerly *et al.*, 1995b; Eskes *et al.*, 1997). Thus, either the same interactions that position the endonuclease for sense-strand cleavage also position it for antisense-strand cleavage, or sense-strand cleavage is a prerequisite for antisense-strand cleavage. Notably, the interaction of the protein with the E3 region appears to influence the ratio of partially to fully reverse-spliced products, with the 3′ deletion constructs E3+1 to +4 showing a greatly increased proportion of fully reverse-spliced products. A possible explanation is that, after sense-strand cleavage by partial reverse splicing, antisense-strand cleavage and the second step of reverse splicing are competing reactions, with the latter becoming more difficult after antisense-strand cleavage has occurred. In that case, weakened interactions with the downstream region may favor the second step of reverse splicing by decreasing the rate of antisense-strand cleavage.

Practical considerations

In principle, group II intron endonucleases could be modified to target specific DNA sequences for site-specific cleavage and insertion by introducing complementary nucleotides in the target-site recognition regions of the intron RNA. In practice, we have been able to change simultaneously all six base pairs of the EBS1–IBS1 and δ–δ′ interactions without significantly decreasing the reverse splicing into the DNA target site, although some combinations of EBS1–IBS1 base pairings are suboptimal (C.Beall and A.M.Lambowitz, unpublished data). We have also had difficulty introducing multiple changes into EBS2. These difficulties could reflect that certain combinations of EBS nucleotides are not allowed either because they adversely affect the structure of the intron RNA or because there are as yet undetermined rules for the EBS–IBS interactions (e.g., a required range of thermodynamic stabilities). In addition, the aI2 protein may contribute to the recognition of E2–11 and –12 in the IBS2 region and thus further limit the nucleotide residues that can be changed. The use of group II introns for targeted DNA cleavage and insertion will require further analysis of these different constraints.

Evolutionary considerations

We find that reverse splicing of group II introns into double-stranded DNA substrates occurs by essentially the same mechanism as reverse splicing into RNA, except that the intron-encoded protein appears to be required for DNA binding and strand displacement to permit base pairing of the intron RNA. Griffin *et al.* (1995) have shown that the yeast group II intron aI5γ cleaves DNA and RNA substrates by reverse splicing with similar catalytic efficiency. Thus, mobility via reverse splicing into DNA may be intrinsic to the catalytic RNA and could have predated acquisition of the RT ORF. In the absence of the intron-encoded RT, the inserted intron RNA might still be incorporated into DNA by cellular DNA replication and repair enzymes. Alternative scenarios in which group

II ribozyme activity evolved from retrotransposons that already encoded RTs are also possible (see Curcio and Belfort, 1996).

Finally, our results show how an enzyme can evolve to use RNA and protein subunits cooperatively to recognize specific sequences in double-stranded DNA. The use of RNA subunits for recognition of specific DNA sequences may have been widely employed at the time of transition from an RNA to a DNA world and then gradually replaced by the use of proteins that recognize the same sequences. Among present-day enzymes, telomerase, which contains an endogenous RNA template that base pairs with the telomere repeat at the 3′ end of chromosomal DNA (Greider and Blackburn, 1987), is the best candidate to use an analogous mode of target site recognition. In addition, mammalian DNA demethylation activity has an RNA component that could in principle contribute to substrate recognition (Weiss *et al.*, 1996). Practically, the mechanism used by group II introns is analogous to recent strategies that attempt to target DNA cleavage by using DNA triplexes or peptide nucleic acids conjugated to or in conjunction with nucleases (e.g. Demidov *et al.*, 1993; Norton *et al.*, 1995). The advantage of group II introns is that they have had millions of years of evolutionary selection to refine this process.

Materials and methods

Yeast strains and isolation of mtRNP particles

Yeast strains $1°2^{+t}$ (containing the wild-type aI2 intron), C1036 1°2°, $1°2^{\Delta D5}$, $1°2^{\Delta Zn}$ and $1°2^{\Delta ConZn}$ were described in Moran *et al.* (1995) and Zimmerly *et al.* (1995b). A superscript '+' indicates the presence of wild-type aI1 or aI2, a superscript 'o' indicates the absence of an intron, and other superscripts refer to specific alleles. All strains were derived from wild-type ID41-6/161 MAT*a ade1 lys1* (denoted 161), which contains both aI1 and aI2. The $1°2^{\Delta Zn}$ and $1°2^{\Delta ConZn}$ strains were made by substituting TAA stop codons for aI2 ORF codons 676–678 and 719–721, respectively. The preparation of mitochondria and mtRNP particles were as described (Kennell *et al.*, 1993; Zimmerly *et al.*, 1995a).

Oligonucleotides

Oligonucleotides used in this study were: aI3, 5′-CCATCTGCATCTGTCATACCAGCTAATC; aI3-40, 5′-CTTGATTATTATTATTATTACTTTC; aI3-62, 5′-TTATATTCTTTATTTTCTATTACTTG; AsE1-BclINdeI, 5′-CATATGATCAGGTGCAGCTAATTCTAATCTAA; cyt-18, 5′-GCTGGGATTGGTTCGAGACCAGCAGGGAGTTCAGATGCAGATT; modified cyt-18 oligonucleotides containing different segments of the aI2 target sequence inserted between positions 17G and 18A were used to generate DNA substrates in Figure 6A and B; cyt-18-3′, 5′-AATCTGCATCTGAACTCCCTG; cyt-18-5′, 5′-GCTGGGATTGGTTCGAG; D273-10B gap, 5′-CCGGCCCGCCCCCGCGGGACCCCTT; E1-BclINdeI, 5′-CACCTGATCATATGTATTTACATGGTAATTCA; E1-JY, 5′-TAATCATTAGATTAGAATTAGCTGCACCTG; E2(161), 5′-TTTTAGTAGTTGGTCATGCTG; E2-S, 5′-TTTTAGTAGTTGGTC; E2E3(161), 5′-TTTTAGTAGTTGGTCATGCTGTATTAATGATTTTCTTCTTAGTAATGCCTGCTTTAATTGGAGGTTTTGGT; modified E2E3(161) oligonucleotides containing mutations in the aI2 target site were used to generate mutant DNA substrates specified in the figure legends; HG1, 5′-ACCAAAACCTCCAATTAAAGCAGGC; KS, 5′-TCGAGGTCGACGGTATC; SK, 5′-CGCTCTAGAACTAGTGGATC; SZ4, 5′-TAGTTGGTCATGCTGTATTAATGATTTTCTTCTTAGTAATGCCTGCTTTAATTGGA; SZ4-AntiIBS1+2, 5′-TAGTTGGTCATGCTGTAAATTACTAAAAGATCTTAGTAATGCCTGCTTTAATTGGA.

Recombinant plasmids used for the preparation of DNA target sites

pHG3 contains a 983 bp segment of the *COX1* gene of C1036 1°2°, extending from 225 bp upstream of the ATG of E1 to 518 bp into aI3, with sequence changes that introduce *Bcl*I and *Nde*I sites at E1+129 and +133, respectively. The insert was generated in two pieces by PCR of C1036 1°2° mtDNA using the primer pairs D273-10B gap plus AsE1-BclINdeI and E1-BclINdeI plus aI3. The PCR products were cleaved at the overlapping *Bcl*I sites at E1+129 and ligated together into the *Sma*I site of pBluescript II KS(+)(Stratagene, La Jolla, CA).

pE2E3(161) contains a 71 bp insert corresponding to *COX1* E2/E3 of wild-type 161. The insert was generated from pHG3 by PCR with primers E2(161) and HG1 and cloned into the *Sma*I site of pBluescript II KS(+) in the minus orientation.

To construct 5′ deletions of the aI2 target site region, pHG3 was linearized at the *Nde*I site in E1 and digested with *Bal*31 nuclease (Boehringer Mannheim, Indianapolis, IN) for different times (Sambrook *et al.*, 1989). After fill-in with phage T4 DNA polymerase, the *Bal*31-digested plasmids were cleaved at the *Bst*EII site in E3. Inserts containing different-sized deletions were purified in a 1.5% agarose gel and cloned between the *Nde*I site (blunt-ended with phage T4 DNA polymerase) and *Bst*EII site of pHG3. 3′ deletions were made similarly by *Bal*31 digestion of pHG3 linearized at the *Bst*EII site in E3. In this case, the *Bal*31-digested plasmids were cut at the *Nde*I site in E1, and the inserts were recloned in pHG3 between the *Bst*EII site (blunt ended with phage T4 DNA polymerase) and *Nde*I site. The cloned inserts were sequenced by the dideoxy method using primers EI-JY or aI3-62 to identify the deletion end points.

pHG3-δ′A, pHG3-δ′C and pHG3-δ′G contain aI2 DNA target sites in which E3+1 was replaced by A, C or G, respectively. The plasmids were derived from pHG3 by PCR mutagenesis with appropriate primers. The cloned PCR products were sequenced to verify that the correct mutation had been introduced with no adventitious mutations.

aI2 intron constructs

pJVM4 contains aI2 and flanking exon sequences from wild-type 161 cloned downstream of the phage T3 promoter in pBluescript II KS(+) (Kennell *et al.*, 1993). Plasmids containing modified introns were derived from pJVM4 by PCR mutagenesis with appropriate primers. The modified region was sequenced to verify the correct mutation and the absence of adventitious mutations.

pJVM4-aI1EBS1, pJVM4-aI1EBS2 and pJVM4-aI1EBS1+2 contain aI2 derivatives in which the EBS1 and/or EBS2 sequences were replaced with those of aI1. In each case, appropriate changes were also introduced into IBS1 and IBS2 to permit *in vitro* splicing. pJVM4-aI1EBS1 has EBS1 positions 2985–2990 (numbered according to Bonitz *et al.*, 1980) changed from 5′ AGAAGA to 5′ CGTTGA; pJVM4-aI1EBS2 has EBS2 positions 2935–2940 changed from 5′ TCATTA to 5′ ACAATT; and pJVM4-aI1EBS1+2 has EBS1 and EBS2 positions 2935–2940 and 2985–2990 changed from 5′ TCATTA to 5′ ACAATT and 5′ AGAAGA to 5′ CGTTGA, respectively. For pJVM4-aI1EBS1 and pJVM4-aI1EBS1+2, the 5′ portion of the *COX1* insert consisting of aI1 and E2 sequences (positions 2257–2653) was replaced with the last 24 bp of E1. For pJVM4-aI1EBS2, positions 2257–2647 were replaced with 5′-ATGGTAATTCACAATTAT, leaving the aI2 IBS1 sequence unchanged. For all three constructs, the sequence downstream of the intron was replaced with the first 15 bp of E2.

pJVM4-EBS2–8G, pJVM4-EBS2–9U,–10A, pJVM4-EBS2–11A, pJVM4-EBS2–12U and pJVM4-EBS2–13U–1 are derivatives of pJVM4 in which the indicated changes were introduced at different positions in EBS2 by PCR mutagenesis with appropriate primers. pJVM4-EBS2–13U–2 is identical to pJVM4-EBS2–13U–1, except that it contains a second mutation, T to A, at intron position 2932.

pJVM4-δ-C, pJVM4-δ-G and pJVM4-δ-T are derivatives of pJVM4 in which the δ nucleotide (position 2984) was changed to C, G or T, respectively, with a compensatory nucleotide substituted at the δ′ position of E3 (position 5169) for *in vitro* splicing.

Synthesis of DNA substrates

^{32}P-labeled DNA substrates for reverse splicing and endonuclease assays were synthesized from recombinant plasmid or synthetic oligonucleotide templates by PCR (Zimmerly *et al.*, 1995a,b). For internally labeled DNA substrates, the PCR was carried out in 25 μl of reaction medium containing 200 μM each of dATP, dCTP and dGTP, 30 μM dTTP, 50 μCi [α-^{32}P]dTTP (3000 Ci/mmol; DuPont, New England Nuclear, Boston, MA), 125–150 ng primers and 1.0–2.5 ng plasmid or 0.5 ng oligonucleotide template and 2.5 units *Taq* DNA polymerase (Life Technologies, Gaithersburg, MD). The internally labeled DNA substrates in Figure 6 were synthesized in a modified reaction medium using [α-^{32}P]dCTP (3000 Ci/mmol; DuPont, New England Nuclear, Boston, MA) instead of [α-^{32}P]dTTP. Unless otherwise indicated, the labeled DNA substrates were purified in a 1.5% agarose gel. For 5′ end-labeled DNA substrates, the PCR used 75 ng 5′ ^{32}P-labeled primer (specific activity 180 000–

330 000 c.p.m./ng), 150 ng unlabeled primer and ~4 ng DNA template, as described (Zimmerly *et al.*, 1995b), and the DNA substrates were purified in a nondenaturing 6% polyacrylamide gel.

3′ end-labeled single-stranded DNA substrates were synthesized by labeling synthetic oligonucleotides with [α-^{32}P]dCTP using recombinant terminal deoxyribonucleotidyl transferase (rTdT; Life Technologies, Gaithersburg, MD), according to the *Promega Protocols and Applications Guide*, 2nd Edn (1991). 1.5 pmol of gel-purified oligonucleotide was labeled in 10 µl of reaction medium containing 15 µCi [α-^{32}P]dCTP and 7.5 units of rTdT. The labeled oligonucleotide was ethanol-precipitated twice in the presence of 2.3 M ammonium acetate and 5 µg *Escherichia coli* tRNA (Sigma, St Louis, MO) carrier, dissolved in TE (10 mM Tris–HCl, pH 8.0, 1 mM EDTA) to a concentration of 75 nM, and then adjusted to a final specific activity of 2000 c.p.m./fmol by adding 75 nM unlabeled oligonucleotide.

For the experiment shown in Figure 5, DNA substrates having mixtures of the three non-wild-type nucleotides at specific positions were made by PCR using complementary primers with a 10 nt overlap. The PCR was carried out for 10 cycles with an annealing temperature of 35°C. For mutations at E2–21 through –14 (Figure 5A), the upstream oligonucleotide extended from E2–36 to –4, and the downstream oligonucleotide extended from E3+35 to E2–13. For mutations at E3+1 to +10 (Figure 5B), the upstream oligonucleotide extended from E2–36 to E2–1, and the downstream oligonucleotide extended from E3+35 to E2–10. Internally labeled substrates were synthesized by PCR with [α-^{32}P]dTTP and purified in a 2% agarose gel. For 5′ end-labeled substrates, gel-purified non-radiolabeled substrates were prepared first and used as templates for PCR with primers E2-S and 5′ end-labeled HG1. The 5′ end-labeled substrates were purified in a non-denaturing 6% polyacrylamide gel.

Reconstitution of aI2 endonuclease using in vitro-synthesized intron RNA

aI2 endonuclease was reconstituted by mixing *in vitro*-synthesized aI2 RNA with aI2 protein from 1°2$^{\Delta D5}$ mtRNP particles. *In vitro* transcripts containing aI2 were synthesized with phage T3 RNA polymerase from plasmid pJVM4 or its derivatives linearized with either *Bgl*I (pJVM4-aI1EBS1, pJVM4-aI1EBS2 and pJVM4-aI1EBS1+2) or *Bst*EII (all other plasmids). Self-splicing of the *in vitro* transcripts was as described (cf. Hebbar *et al.*, 1992; Zimmerly *et al.*, 1995b), except that the concentration of NH_4Cl was increased to 2 M. The products were ethanol precipitated twice in the presence of 0.3 M NaOAc, pH 7.9, and dissolved in TE to a final concentration of 1.0 µg/µl.

For reconstitution, 1 µl of the *in vitro* splicing products containing aI2 lariat RNA, which had been heated to 90°C for 2 min and cooled on ice, was mixed with 2 µl (0.025 OD_{260} units) of 1°2$^{\Delta D5}$ mtRNP particles. The mixture was incubated on ice for 1–10 min, and then used directly for reverse splicing or endonuclease reactions. The length of preincubation had little effect on the efficiency of reconstitution. Optimal conditions for the reverse splicing and endonuclease reactions with the reconstituted RNP particles were the same as those with wild-type mtRNP particles (Zimmerly *et al.*, 1995b). With 1 µg of aI2 RNA, the reconstitution was limited by the amount of aI2 protein in the 1°2$^{\Delta D5}$ mtRNP particles. The extent of reaction increased linearly with increasing amounts of the 1°2$^{\Delta D5}$ mtRNP particles up to the highest amount tested (0.075 OD_{260} units).

Reverse splicing and DNA endonuclease assays

Reverse splicing and DNA endonuclease assays were carried out as described (Zimmerly *et al.*, 1995a,b) by incubating labeled DNA substrates with 0.025 OD_{260} units of yeast mtRNP particles or the equivalent amount of reconstituted RNP particles. The reactions were carried out for 20 min, which is within the linear range for reverse splicing and sense- and antisense-strand cleavage with 1°2^{+t} mtRNP particles. For reverse splicing assays, the products were denatured with glyoxal and analyzed in a 1% agarose gel, using glyoxalated *Eco*RI–*Hin*dIII fragments of phage λ DNA as size markers. For DNA endonuclease assays, the DNA cleavage products were analyzed in a denaturing 6% or 10% polyacrylamide gel, alongside DNA sequencing ladders. The gels were dried and autoradiographed or quantitated with a Molecular Dynamics Phosphorimager 445 SI.

Acknowledgements

We thank Mr James Williams and Ms Michelle Simons for technical assistance. This work was supported by NIH grant GM37949 to A.M.L. and GM31480 to P.S.P.

References

Bonitz,S.G., Coruzzi,G., Thalenfield,B.E., Tzagoloff,A. and Macino,G. (1980) Assembly of the mitochondrial membrane system: structure and nucleotide sequence of the gene coding for subunit I of cytochrome oxidase. *J. Biol. Chem.*, **255**, 11927–11941.

Carignani,G., Groudinsky,O., Frezza,D., Schiavon,E., Bergantino,E. and Slonimski,P.P. (1983) An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in *S. cerevisiae*. *Cell*, **35**, 733–742.

Cherniack,A.D., Garriga,G., Kittle,J.D.,Jr, Akins,R.A. and Lambowitz,A.M. (1990) Function of Neurospora mitochondrial tyrosyl-tRNA synthetase in RNA splicing requires an idiosyncratic domain not found in other synthetases. *Cell*, **62**, 745–755.

Curcio,M.J. and Belfort,M. (1996) Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. *Cell*, **84**, 9–12.

Demidov,V., Frank-Kamenetskii,M.D., Egholm,M., Buchardt,O. and Nielsen,P.E. (1993) Sequence selective double strand DNA cleavage by Peptide Nucleic Acid (PNA) targeting nuclease S1. *Nucleic Acids Res.*, **21**, 2103–2107.

Eickbush,T.H. (1994) Origin and evolutionary relationships of retroelements. In Morse,S.S. (ed.), *The Evolutionary Biology of Viruses*. Raven Press, Ltd, New York, pp. 121–157.

Eskes,R., Yang,J., Lambowitz,A.M. and Perlman,P.S. (1997) Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell*, **88**, 865–874.

Feng,Q., Moran,J.V., Kazazian,H.H.,Jr and Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.

Frank-Kamenetskii,M.D. and Mirkin,S.M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.*, **64**, 65–95.

Greider,C.W. and Blackburn,E.H. (1987) The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell*, **51**, 887–898.

Griffin,E.A.,Jr, Qin,Z., Michels,W.J.,Jr and Pyle,A.M. (1995) Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2′-hydroxyl groups. *Chem. Biol.*, **2**, 761–770.

Hebbar,S.K., Belcher,S.M. and Perlman,P.S. (1992) A maturase-encoding group IIA intron of yeast mitochondria self-splices *in vitro*. *Nucleic Acids Res.*, **20**, 1747–1754.

Kennell,J.C., Moran,J.V., Perlman,P.S., Butow,R.A. and Lambowitz,A.M. (1993) Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell*, **73**, 133–146.

Lambowitz,A.M. and Belfort,M. (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem.*, **62**, 587–622.

Lazowska,J., Meunier,B. and Macadre,C. (1994) Homing of a group II intron in yeast mitochondrial DNA is accompanied by unidirectional co-conversion of upstream-located markers. *EMBO J.*, **13**, 4963–4972.

Luan,D.D., Korman,M.H., Jakubczak,J.L. and Eickbush,T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, **72**, 595–605.

Matsuura,M. *et al.* (1997) A bacterial group II intron encoding reverse transcriptase, maturase and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.*, in press.

Michel,F. and Ferat,J.L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem.*, **64**, 435–461.

Moran,J.V., Mecklenburg,K.L., Sass,P., Belcher,S,M., Mahnke,D., Lewin,A. and Perlman, P. (1994) Splicing defective mutants of the *COXI* gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron AI2. *Nucleic Acids Res.*, **22**, 2057- 2064.

Moran,J.M., Zimmerly,S., Eskes,R., Kennell,J.C., Lambowitz,A.M., Butow,R.A. and Perlman,P.S. (1995) Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol. Cell. Biol.*, **15**, 2828–2838.

Mueller,M.W., Allmaier,M., Eskes,R. and Schweyen,R.J. (1993) Transposition of group II intron aI1 in yeast and invasion of mitochondrial genes at new locations. *Nature*, **366**, 174–176.

Norton,J.C., Waggenspack,J.H., Varnum,E. and Corey,D.R. (1995) Targeting peptide nucleic acid–protein conjugates to structural features within duplex DNA. *Bioorg. Med. Chem.*, **3**, 437–445.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schmidt,W.M., Schweyen,R.J., Wolf,K. and Mueller,M.W. (1994) Transposable group II introns in fission and budding yeast: site-specific genomic instabilities and formation of group II IVS plDNAs. *J. Mol. Biol.*, **243**, 157–166.

Shi,Y. and Berg,J.M. (1996) DNA unwinding induced by zinc finger protein binding. *Biochemistry*, **35**, 3845–3848.

Yang,J., Zimmerly,S., Perlman,P.S. and Lambowitz,A.M. (1996) Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature*, **381**, 332–335.

Weiss,A., Keshet,I., Razin,A. and Cedar,H. (1996) DNA demethylation *in vitro*: involvement of RNA. *Cell*, **86**, 709–718.

Zimmerly,S., Guo,H., Perlman,P.S. and Lambowitz,A.M. (1995a) Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*, **82**, 545–554.

Zimmerly,S., Guo,H., Eskes,R., Yang,J., Perlman,P.S. and Lambowitz,A.M (1995b) A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell*, **83**, 529–538.

Received on July 21, 1997; revised on September 8, 1997