The EMBO Journal vol.14 no.13 pp.3236–3246, 1995

Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs

A.Deirdre J.Scadden and Christopher W.J.Smith[1]

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

[1]Corresponding author

Nuclear pre-mRNA splicing has a fundamentally similar two-step mechanism to that employed by group II self-splicing introns. It is believed that nuclear pre-mRNA splicing involves a network of RNA–RNA interactions which form the catalytic core of the active spliceosome. We show here a non-Watson–Crick interaction between the first and last guanosine residues of a mammalian intron. As in *Saccharomyces cerevisiae*, substitution of the conserved guanosines at the 5′ and 3′ splice sites by A and C respectively, specifically suppresses step 2 splicing defects resulting from the individual mutations. No other combination of terminal nucleotides was able to restore splicing. We additionally provide independent evidence for an indirect interaction between other nucleotides of the consensus splice sites during step 2 of splicing. Substitution of the nucleotide in the +3 position of the 5′ splice site affects competition between closely spaced AG dinucleotides at the 3′ splice site, although the interaction is not via direct differential base pairing. Finally, we show that complete substitution of guanosine residues by inosine in a pre-mRNA has only a modest effect upon step 2 of splicing, although earlier spliceosome assembly steps are impaired. Predictions can thus be made about the precise configuration of the non-Watson–Crick interaction between the terminal residues.

Key words: inosine/RNA splicing/spliceosome

Introduction

Splicing of nuclear pre-mRNA is the process by which non-coding intervening sequences (introns) are accurately removed from the primary transcripts of eukaryotic genes. The removal of introns and concomitant ligation of coding sequences (exons) is a fundamentally important process for correct gene expression. Splicing of pre-mRNA is facilitated by a large number of protein and small nuclear RNA (snRNA) factors which together assemble with the pre-mRNA into a large ribonucleoprotein particle called the spliceosome (for reviews see Green, 1991; Moore *et al.*, 1993; Madhani and Guthrie, 1994).

Parallels may be drawn between the splicing of nuclear pre-mRNA and the autocatalytic excision of group II introns which have been found in organelles of fungi, higher plants and algae, and in bacteria (for review see Peebles *et al.*, 1986; Weiner, 1993). In both systems introns are removed via two successive transesterification reactions resulting in excision of the intron as a lariat with a 2′-5′ branched structure. Moreover, the stereochemical specificity of the two steps is identical in both group II and pre-mRNA splicing (Moore and Sharp, 1993; Padgett *et al.*, 1994). While the highly conserved secondary and tertiary structure of group II self-splicing introns provides the necessary signals for specificity and catalysis, very little information is encoded within nuclear pre-mRNA sequences. Short variably conserved consensus sequences at both intron–exon junctions are necessary, but not sufficient, for accurate splicing of pre-mRNA. The 5′ splice site in mammalian genes is defined by a single short consensus sequence (A/C)AG|GURAGU, while three elements are associated with the 3′ end of the intron; the branch-point (BP), YNYURAY, the polypyrimidine tract adjacent to the BP, and the 3′ splice site, YAG|G (vertical bars denote exon–intron boundaries, and invariant nucleotides are underlined). While numerous protein splicing factors are required for splicing of pre-mRNA, the mechanistic similarity between splicing of group II introns and nuclear pre-mRNA suggests that pre-mRNA splicing may also be fundamentally controlled by RNA–RNA interactions. Much effort has been made to characterize the network of RNA interactions which are potentially involved in splicing. Such RNA interactions include intermolecular interactions between different snRNAs and between snRNAs and pre-mRNA, and intramolecular interactions both within the individual snRNAs and within the pre-mRNA substrate. RNA interactions in all four categories are thought to be central to the splicing process and may play an important role in the catalytic core of the active spliceosome (Moore *et al.*, 1993; Madhani and Guthrie, 1994; Newman, 1994; Nilsen, 1994).

In contrast to the wealth of information on recognition of the 5′ splice site, which is recognized successively by U1, U5 and U6 snRNAs, and the BP which base pairs with U2 snRNA, relatively little is known about the RNA interactions involved in the specification of the 3′ splice site in step 2 of splicing. U5 snRNA interacts with the exon sequence adjacent to the 3′ splice site AG, but not with the intron itself (Newman and Norman, 1992). Although an interaction between U1 snRNA and the 3′ splice site is necessary in some *Schizosaccharomyces pombe* introns for step 1 of splicing, mutations in U1 snRNA were unable to suppress step 2 defects (Reich *et al.*, 1992). Thus, the interactions of the pre-mRNA with U5 and U1 snRNAs do not appear to be likely candidates for the precise recognition of the 3′ splice site AG during step 2 of splicing. In addition to these snRNA–pre-mRNA interactions, a non-Watson–Crick interaction between the highly conserved G residues in the first and last intron positions has been postulated. This was based upon the reciprocal suppression of step 2 blockage by 5′ splice site

G_1→A and 3′ splice site G_{-1}→C mutants in a yeast actin intron (Parker and Siliciano, 1993). It was proposed that a non-Watson–Crick A–C base pair is able to partially restore the normal G–G interaction. An interaction between the invariant first and penultimate nucleotides (G1 and A886 respectively) was similarly described in group II self-splicing introns; some double mutants at these positions demonstrate a reciprocal suppression of step two splicing defects (Chanfreau and Jacquier, 1993). The finding that an interaction between intron boundaries is required for the second step in both group II and nuclear pre-mRNA introns strengthens the idea that both systems employ similar mechanisms, albeit with differences in the details of the nucleotide interactions.

In this paper we describe experiments aimed at elucidating the nature of interactions between the 5′ and 3′ splice site consensus nucleotides of mammalian introns, and their role in defining the 3′ splice site during step 2 of splicing. (i) We first show that in a panel of all possible mutant combinations at the first and last positions of the intron, we could only detect spliced products that had arisen from the excision of introns with G at both ends (|G–G|) or with A at the 5′ end and C at the 3′ end (|A–C|). This confirms the genetic suppression data in *Saccharomyces cerevisiae* (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994). (ii) We found that mutations at the +3 position of the 5′ splice site were able to influence competition between closely spaced 3′ splice site AGs during step 2 of splicing. The effect was not allele specific and so is not consistent with specific base pairing of the 5′ splice site +3 nucleotide and the 3′ splice site −3 nucleotide. Nevertheless, this observation independently supports the concept that the consensus 5′ and 3′ splice sites are in close proximity during step 2 of splicing. (iii) We investigated the likely conformation of the G–G interaction by splicing RNAs in which guanosine was replaced by inosine. Although early spliceosome assembly steps were impaired in the inosine transcripts, both the catalytic steps were able to proceed. This result places restrictions upon the possible base pairing configuration between the terminal intron guanosines, and suggests that N1-carbonyl symmetric hydrogen bonding is the most likely interaction.

Results

Accurate in vitro 5′ splice site cleavage with all mutations at intron position 1

An interaction between the first and last G residues of an intron of *S.cerevisiae* was postulated on the basis of genetic suppression data (Parker and Siliciano, 1993). Specifically, a 3′ splice site AG→AC transversion was seen to suppress a second step splicing defect caused by a 5′ splice site G_1→A mutation. In order to investigate whether such an interaction similarly occurs between the first and last nucleotides of a mammalian intron, mutations were made in the consensus sequences at either end of a modified form of the intron between exons 2 and 3 of the rat α-TM gene (Smith and Nadal-Ginard, 1989; and Figure 1A). Modifications included the insertion of a spacer element between the 5′ splice site and BP to relieve steric interference, and the deletion of ~84 nucleotides between the polypyrimidine tract and the 3′ splice site AG. Splicing was analysed both in HeLa nuclear extracts *in vitro* and in transiently transfected COS cells *in vivo* .

In the first series of constructs ($pTMN_1$), the invariant G_1 residue of the intron was mutated to every other nucleotide, and the sequence CACAG was present at the 3′ splice site. When the wild-type GU dinucleotide was present at the 5′ splice site, splicing of the intron *in vitro* proceeded efficiently using the CAG as the 3′ splice site, as expected (Figure 1B). When the remaining $pTMN_1$ constructs (A_1, C_1 and U_1) were spliced *in vitro*, step 1 of splicing was seen to proceed accurately in all cases, albeit with reduced efficiency for the C_1 mutant. The accuracy of cleavage at the 5′ splice site was indicated by the uniform size of the excised 5′ exon intermediate from all mutants. Although cleavage at the 5′ splice site occurred efficiently, lack of detectable products and the concomitant accumulation of splicing intermediates indicated that splicing was blocked after step 1. This observation was in agreement with previous analyses which showed that some mutations in the conserved dinucleotide GU at the 5′ splice site did not abolish 5′ cleavage, but instead had a striking effect on cleavage and splicing at the 3′ splice site (Wieringa *et al.*, 1983; Newman *et al.*, 1985; Aebi *et al.*, 1986; Parker and Siliciano, 1993). Unlike some previous studies however, the block in splicing was after step 1 for all three mutants. For the $pTMC_1$ and $pTMU_1$ mutants a second lariat intermediate was detected running with slightly higher mobility. These lariats were produced by use of a branch-point adenosine seven nucleotides upstream of the wild-type (data not shown). In contrast with other introns studied (e.g. Aebi *et al.*, 1986), the transcript used here does not contain cryptic 5′ splice sites which can compete *in vitro* with the mutated wild-type splice site.

We were unable to observe any spliced product arising from use of the AC dinucleotide at the 3′ splice site of the $pTMA_1$ mutant. The sequence context of the AC dinucleotide was favourable for splicing; a C preceded the AC (as found in 72% of mammalian introns, Shapiro and Senapathy, 1987), and the AG following the AC dinucleotide was optimal for base pairing with U5 snRNA (Newman and Norman, 1992; Sontheimer and Steitz, 1993). Previous reports had shown that A_1 mutations in a yeast *rp51A* intron led to use of AC or AA dinucleotides just upstream of the usual AG dinucleotide (Chanfreau *et al.*, 1994). However, these observations, like the previous yeast suppression data (Parker and Siliciano, 1993), were made under steady-state conditions *in vivo*. We therefore carried out further analyses by transient transfection of COS cells.

In order to analyse splicing *in vivo*, the $pTMN_1$ series of mutants were transferred into an SV40 expression vector (Smith and Nadal-Ginard, 1989). The resultant $pSVTMN_1$ series of constructs were transfected into COS cells. RT-PCR analysis of transiently expressed RNA from all four constructs gave rise to a product of 226 bp which was the size expected when the RNA was correctly spliced. The exact site of exon ligation in the spliced α-TM $pSVTMN_1$ mutant transcripts was unambiguously identified by sequence analysis of cloned PCR products (Figure 1D). When the wild-type 5′ splice site was present (GU), splicing occurred using the trinucleotide CAG as the 3′

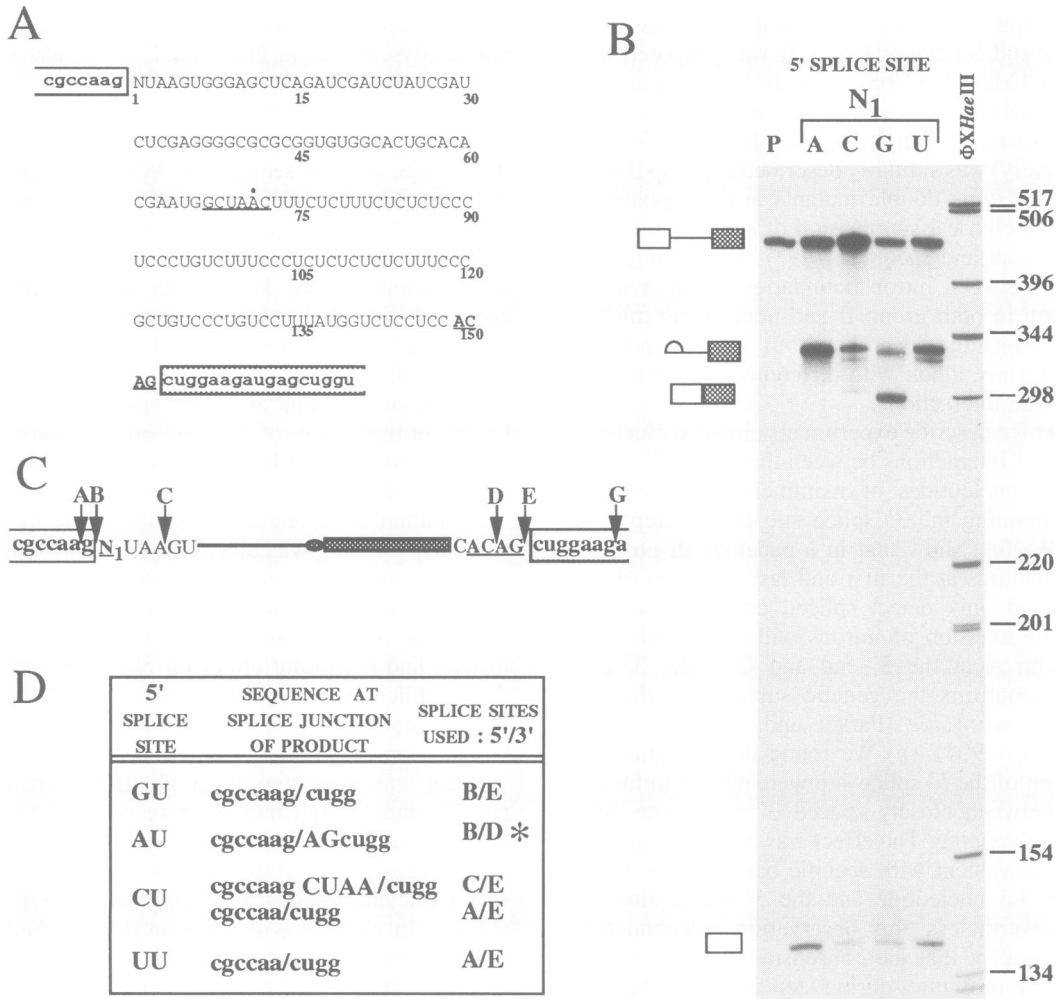

Fig. 1. Mutation of the conserved guanosine residue at the 5′ end of a mammalian intron. (**A**) The sequence corresponding to the $pTMN_1$ pre-mRNA series. Features to note are the position of the 5′ exon which is indicated by the open box and lower case text, the mutation at the 5′ splice site consensus sequence denoted by N and the sequence **ACAG** which replaces the wild-type 3′ splice site. The 3′ exon as seen in the wild-type transcript is indicated by a shaded box. The BP sequence (underlined) is at nucleotides 67–72, and the polypyrimidine tract at nucleotides 73–120. (**B**) Constructs $pTMN_1$ were spliced under normal conditions for 4 h; splicing intermediates and products are indicated by the symbols to the left of the figure. Although all four transcripts went through step 1 of splicing, a spliced product is only detected in the lane G_1 (wild-type). The lane marked P contains unprocessed $pTMA_1$ pre-mRNA. (**C**) A schematic diagram of the pre-mRNA corresponding to the $pSVTMN_1$ constructs. An open box and lower case text indicate the position of the 5′ and 3′ exons as seen in the wild-type transcript. The BP and polypyrimidine tract are represented by a hatched circle and rectangle respectively. The 5′ and 3′ splice sites actually used for splicing *in vivo* are shown by an arrow and labelled alphabetically. (**D**) A summary of the splice sites used when the RNA transcripts corresponding to the constructs $pSVTMN_1$ were spliced *in vivo*. The sequences were derived from cloned RT-PCR products. In the case of $5'C_1$/ACAG, seven clones were sequenced which used splice site C/G and two which used A/G. Only with the A_1 mutant were spliced products detected which had used the AU 5′ splice site in conjunction with the AC at the 3′ splice site (asterisked).

splice site. In the $pSVTMC_1$ mutant, the AG dinucleotide at the 3′ splice site was used in conjunction with two different cryptic 5′ splice sites; one was at the GU four nucleotides downstream of the wild-type 5′ splice site and the other at the newly created GC, one nucleotide upstream of the wild-type 5′ splice site. The use of a GC dinucleotide as a 5′ splice site has been observed previously (Aebi *et al.*, 1986). When U was in position 1 of the intron, the newly created GU one nucleotide upstream of the wild-type 5′ splice site was used in splicing in conjunction with the AG at the 3′ splice site. Such use of cryptic 5′ splice sites upon mutation of the invariant GU has been reported previously (Wieringa *et al.*, 1983; Aebi *et al.*, 1986; Cortes *et al.*, 1993). In the construct $pSVTMA_1$, the AC dinucleotide of the sequence CACAG at the 3′ splice site was used exclusively as the 3′ splice site in conjunction with the AU dinucleotide at the 5′ splice site.

These observations using a mammalian intron were in agreement with the data from *S.cerevisiae* introns (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994). Mutation of the 5′ splice site GU dinucleotide to AU caused a block in step 2 of splicing at the wild-type AG dinucleotide. However, use of a nearby AC dinucleotide as the 3′ splice site enabled the block in step 2 of splicing to be partially overcome. All other splicing events observed *in vivo* involved use of cryptic 5′ splice sites in which excised introns had a G residue at both ends. Use of these

cryptic 5′ splice sites presumably reflects a very minor proportion of the RNA processed *in vivo*, since the 5′ cleavage at these sites was not readily detectable *in vitro* (Figure 1B).

Splicing occurs only in introns terminating in 5′/G–G/3′ or 5′/A–C/3′

The preceding data demonstrated that an AC dinucleotide could be recruited as a 3′ splice site in the presence of an A at position 1 of the intron. We further investigated the specificity of the putative interaction by constructing the full panel of mutants with N_1 at the 5′ splice site and CANGG at the 3′ splice site (constructs pSVTMN$_1$/3′ANGG). In these constructs the AN dinucleotide is in an optimal sequence context to be used as a 3′ splice site, being preceded by a consensus C and followed by GG, which favours base pairing with U5 snRNA (Newman and Norman, 1992). Transiently expressed RNA was analysed by sequencing RT-PCR products to determine the precise splice sites used (Figure 2B and C), and by RNase protection to obtain a quantitative estimate of the efficiency of splicing of each mutant (Figure 2D).

Sequence analysis of RT-PCR products showed that no other combination of nucleotides was able to restore the accurate splicing seen with the terminal G residues in the wild-type intron, with the single exception of /A–C/ (Figure 2C, see asterisk). A wild-type 5′ splice site (G_1U) in combination with the sequence CANGG at the 3′ splice site led to efficient splicing using the GU dinucleotide at the 5′ splice site and an AG dinucleotide as the 3′ splice site, either within the CANGG sequence when N = A or G, or the AAG located seven nucleotides into the exon when N = U. In each case the AG nearest to the BP was used. With an AU dinucleotide at the 5′ splice site, accurate splicing only occurred with an AC dinucleotide at the 3′ splice site. With all other A_1, C_1 and U_1 mutants, spliced products arose by use of 5′ cryptic GU or GC splice sites in conjunction with the first available AG dinucleotide at the 3′ splice site, as described above.

The splicing of /A–C/ introns and the use of the cryptic GU and GC 5′ splice sites were undetectable *in vitro* (Figure 1B). We therefore directly analysed the *in vivo* spliced RNAs by RNase protection using an antisense probe containing sequences from part of the intron and the downstream exon. The probe produced pairs of protected bands due to an A-rich sequence in the TM exon (marked 'A' in the schematic representation of protected fragments, Figure 2D). The lower of the two bands for spliced RNA is immediately adjacent to a band due to endogenous COS cell TM mRNA (present in 'mock' but not tRNA control lanes). Nevertheless, the upper band clearly shows that the quantity of spliced product with the /A–C/ mutant is higher than with any of the other A_1, C_1 and U_1 mutants in which cryptic 5′ splice sites were used.

The analysis of splicing of these mutant constructs confirms the specificity of the interaction between the 5′ and 3′ splice sites; the interaction between the A at the 5′ splice site and the C at the 3′ splice site is unique in its ability to at least partially restore the proposed interaction between the G residues in the first and last positions of the intron (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994; Ruis *et al.*, 1994).

Mutations at the 5′ splice site +3 nucleotide affect competition between competing 3′ splice sites

We have previously shown that closely spaced AGs are able to compete in step 2 of *in vitro* splicing on the basis of the preceding nucleotide, with a hierarchy of competition of CAG~UAG>AAG>GAG (Smith *et al.*, 1993). Having demonstrated a non-Watson–Crick interaction between the terminal guanosines of a mammalian intron, we decided to test the hypothesis that the hierachy of 3′ splice site competititon was due to an extended interaction between the 5′ and 3′ splice sites (Figure 3A). If the consensus 5′ and 3′ splice sites are aligned in an anti-parallel fashion, the postulated non-Watson–Crick interaction between terminal guanosines could be followed by an invariant A–U base pair, followed by an interaction between the pyrimidine at the –3 position of the 3′ splice site with the consensus purine at the +3 position of the 5′ splice site. If such an interaction occurs, it should be possible to specifically alter the competition between closely spaced AG dinucleotides at the 3′ splice site by changing the nucleotide at the +3 position of the 5′ splice site.

We created a series of hybrid constructs containing a 3′ splice site CAG with competing NAG trinucleotides immediately upstream, in combination with all four variants at the 5′ splice site +3 position (constructs pSVTM5′N$_3$/3′–3NAG, Figure 3B). These constructs were transiently transfected into COS cells and the spliced RNA was analysed by RT-PCR (Figure 3C). When either A or G preceded the upstream AG, exclusive use of the downstream CAG was observed. Use of both AG dinucleotides was seen when the nucleotide preceding the upstream AG was C or U. Thus, these *in vivo* observations confirm the competitive hierarchy previously observed *in vitro* (Smith *et al.*, 1993). The ratio of usage of the competing AG dinucleotides at the 3′ splice site was clearly influenced by the nucleotide at position 3 of the 5′ splice site. When the sequence UAGCAG is present at the 3′ splice site, maximal use of the UAG was observed when 5′ splice site N_3 = A, and use of CAG was favoured when N_3 = G or C (Figure 3C). However, identical shifts in ratios between use of the two AGs in the CAGCAG sequence at the 3′ splice site were observed. The lack of allele specificity demonstrates that there is not a direct interaction between the nucleotides at the +3 position of the 5′ splice site and −3 position of the 3′ splice site. This is consistent with a recent systematic analysis of the *S.cerevisiae* actin intron in which interaction between the invariant 5′ splice site U_2 and the 3′ splice site A_{-2} bases was ruled out (Ruis *et al.*, 1994; and see Figure 3A). While ruling out differential base pairing with the 5′ splice site consensus as the basis of competitive hierarchy of 3′ splice site AGs, the preceding data nevertheless support the concept that the 5′ and 3′ splice sites are in close proximity during step 2 of splicing.

Splicing of inosine substituted pre-mRNA

The preceding data provided two independent lines of evidence for an interaction between the 5′ and 3′ consensus splice sites of mammalian introns during step 2 of splicing. We next proceeded to investigate the configuration of the interaction between the terminal guanosines of the intron by splicing transcripts in which inosine (I) was incorpor-

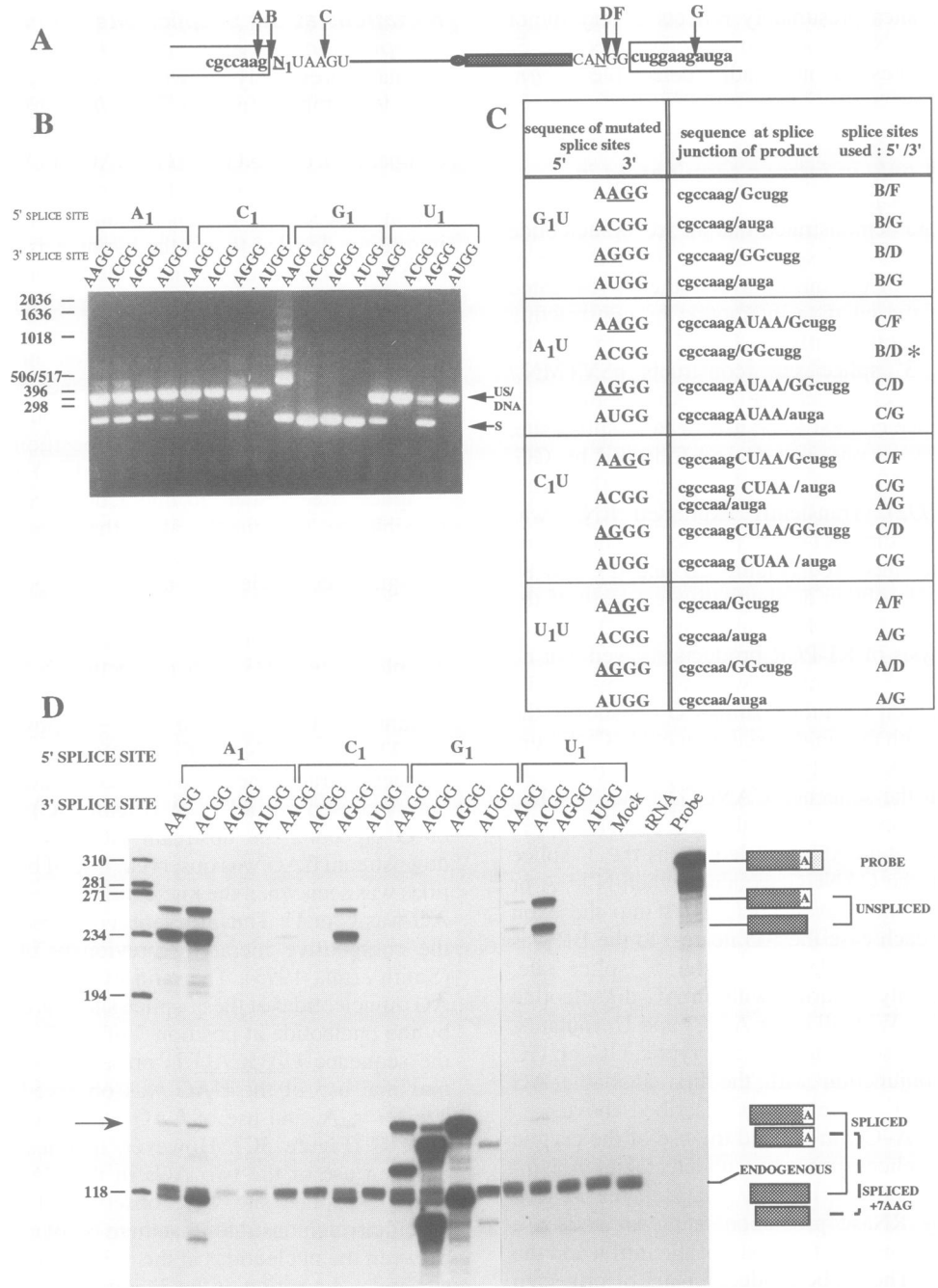

Fig. 2. Specificity of interactions between intron terminal nucleotides. (**A**) A schematic diagram of the pre-mRNA corresponding to the pSVTMN$_1$/ANGG constructs which have all possible combinations of potential intron terminal nucleotides. The boxes with lower case text indicate the position of the exons, the degenerate nucleotides at the 5′ and 3′ splice sites are underlined, and the BP and polypyrimidine tract are represented by a hatched circle and rectangle respectively. The various 5′ and 3′ splice sites used for splicing *in vivo* are shown by an arrow and labelled alphabetically. (**B**) RT-PCR of RNA harvested from COS cells transfected with the constructs pSVTMN$_1$/ANGG. 'S' indicates amplified products corresponding to spliced RNA, while 'US/DNA' denotes amplified products which are derived from both unspliced RNA and/or plasmid DNA. (**C**) A summary of the splice sites used when the RNA transcripts corresponding to the constructs pSVTMN$_1$/ANGG were spliced *in vivo*. Sequences were derived from cloning of the spliced RT-PCR products shown in (B). In the case of 5′C$_1$/ACGG, nine clones were sequenced which used splice sites C/G and only one which used A/G. The asterisk denotes the only case of excision of a non |G–G| intron. (**D**) RNase protection analysis of RNA harvested from COS cells transfected by the constructs pSVTMN$_1$/ANGG. The probe used for RNase protection analysis is shown schematically at the right of the figure. The hatched box represents the part of the probe which is complementary to the exon. The 'A' denotes a U-rich region of the probe which is particularly sensitive to digestion by RNase A, resulting in pairs of protected fragments for each detected RNA as indicated by brackets. 'Spliced' and 'spliced +7AAG' describe the pairs of protected fragments which result from use of a 3′ splice site within either the intron or exon sequence, respectively [as shown in (A)]. The arrow to the left of the figure points out the upper protected fragment which is seen when the 3′ splice site is either AC or AG (D and F) as shown in (A). The 'unspliced' protected fragments represent unprocessed pre-mRNA which has leaked out of the nucleus. Treatment of the RNA samples with DNase I prior to incubation with the probe verified that the large protected band was derived from RNA. The observed bands correspond to the protection products predicted from (C), with the exception of an additional band in the lane 5′G$_1$/AAGG running between the two expected products; the origin of this band is not clear.

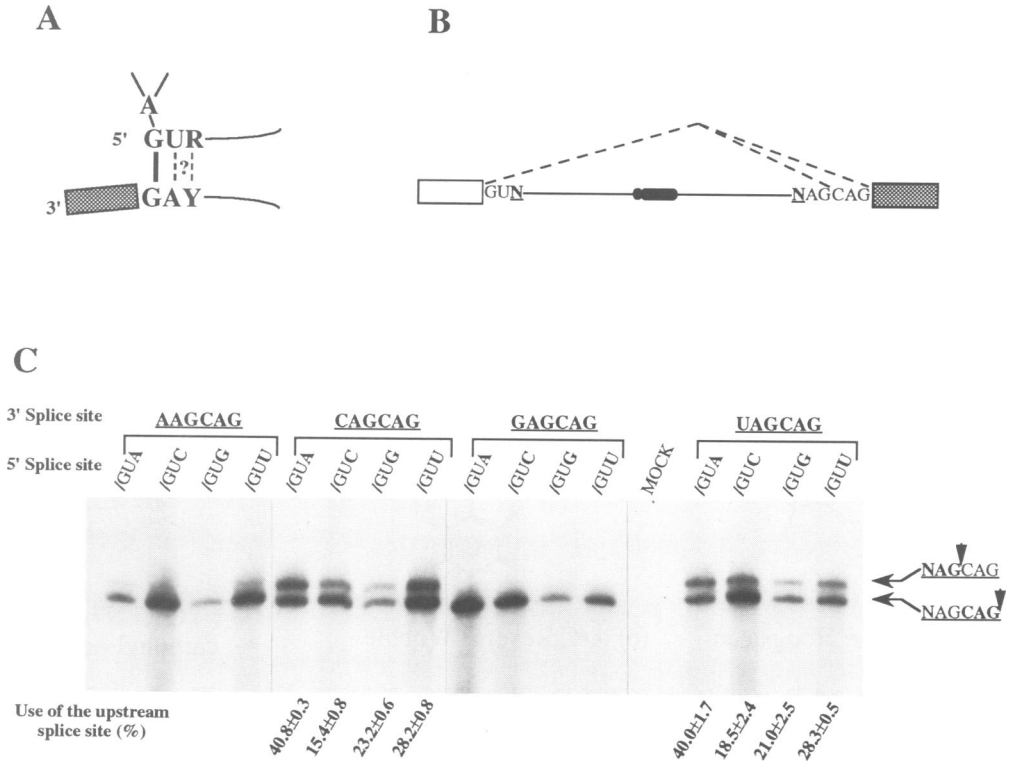

Fig. 3. Competition between 3′ splice sites AGs is influenced by the 5′ splice site sequence. (**A**) A schematic diagram of the proposed extended interaction between the conserved nucleotides in the 5′ and 3′ splice site consensus sequences prior to step 2 of splicing. The bold line represents the non-Watson–Crick base pair between the first and last guanosine residues, while the potential additional interactions are indicated by dotted lines. (**B**) A schematic diagram of constructs pTM5′N_3/3′–3NAG showing the two possible splicing pathways using competing 3′ splice site AGs. (**C**) RT-PCR analysis of RNA harvested from COS cells transfected with the pTM5′N_3/3′–3NAG constructs. Use of both closely spaced AG dinucleotides as the 3′ splice site is evident for both the pTM5′N_3/3′–3CAG and pTM5′N_3/3′–3UAG constructs. The ratio of use of the AGs is affected by the nucleotide in the N_3 postion of the 5′ splice site. As shown by the percentage use of the upstream splice site, the effect on competition between closely spaced 3′ splice sites by the N_3 nucleotide of the 5′ splice site is similar for both the pTM5′N_3/3′–3CAG and pTM5′N_3/3′–3UAG constructs. The figures represent the average value for use of the upstream splice site determined for triplicate transfections of the pTM5′N_3/3′–3NAG constructs.

ated in place of guanosine. Inosine differs from guanosine in lacking the exocyclic N2-amino group, which provides two potential hydrogen bond donor groups. Of the four possible G–G base pairs involving at least two hydrogen bonds (Figure 4; and Tinoco, 1993), one requires both N2-amino groups, two require the N2-amino of only one guanosine, while one (N1-carbonyl symmetric) does not involve N2-amino groups and so should be fully replaceable by inosine.

We therefore synthesized RNA transcripts using either GTP or ITP; SP6 polymerase uses ITP accurately in place of GTP during transcriptional elongation. We used construct Py7, which has a synthetic intron between TM exons 2 and 3, and which proceeds through both steps of splicing efficiently (Figure 5A and B). The incorporation of ITP into RNA transcripts by SP6 polymerase was much less favourable than GTP and, in the absence of CAP analogue, was further reduced. Nevertheless, transcription with CAP but no ITP was undetectable, indicating that the relatively inefficient transcription with ITP did not reflect the use of the G-containing CAP analogue during transcription elongation (Yisraeli and Melton, 1989). Moreover, composition analysis of the I-containing Py7 transcripts by alkaline hydrolysis followed by TLC and autoradiography showed that G incorporation was undetectable (<0.5%, Figure 5C).

When Py7 RNA transcripts containing inosine were spliced *in vitro*, splicing intermediates and products were all detected, albeit at low levels compared with splicing of a wild-type G-containing Py7 transcript (Figure 5B). Thus, the presence of inosine at the highly conserved positions of the 5′ and 3′ splice sites failed to block step 2 of splicing. Significantly, analysis of splicing complexes using both glycerol gradients (Figure 5D) and native polyacrylamide gel electrophoresis (data not shown) showed clearly that spliceosome assembly was itself severely inhibited by substitution of G residues by I. Moreover, the reduction in the amount of 60S spliceosome formed was sufficient to largely account for the reduced amounts of splicing intermediates and products. The impairment of spliceosome assembly was not surprising considering that the base pairing interaction of the pre-mRNA with both U1 and U6 snRNAs would be weaker. In addition, the inosine substitution could potentially inhibit early spliceosome assembly by impairing the function of A/G-rich splicing enhancer sequences in both exons (S.Mayer, personal communication; Staknis and Reed, 1994). To circumvent the spliceosome assembly defect, we pre-incubated the pre-mRNAs with SR proteins, which activate early steps in spliceosome assembly (Staknis and Reed, 1994). Pre-incubation with SR proteins had little effect upon the guanosine-containing transcripts,

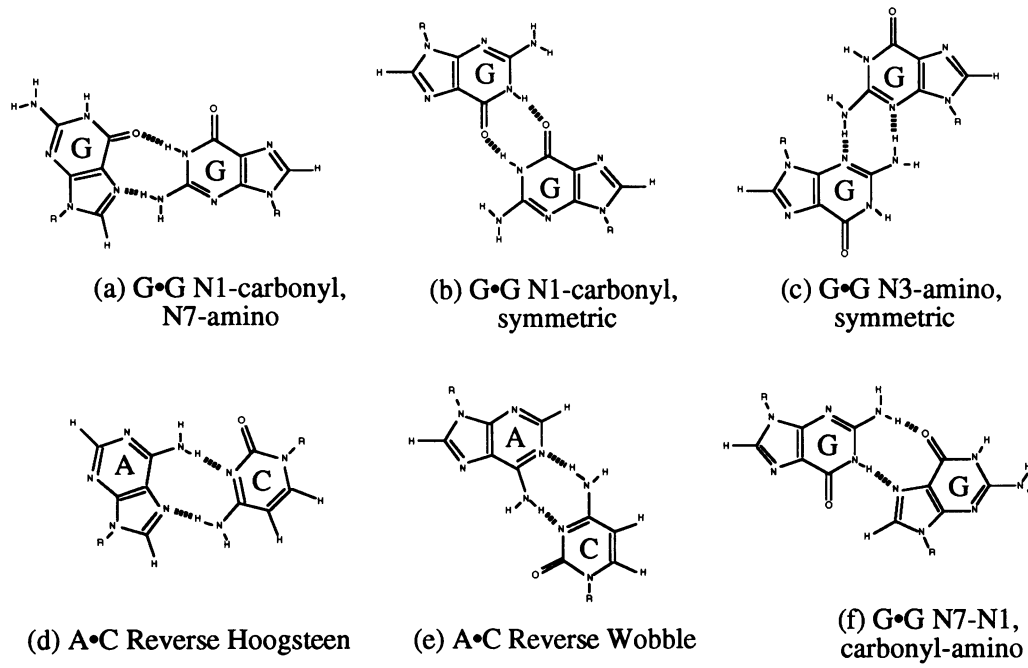

Fig. 4. Models of the potential base pairing configurations between the intron terminal guanosines. Models showing various configurations of non-Watson–Crick G–G interactions (**a**, **b**, **c** and **f**) and A–C interactions (**d** and **e**). The configurations shown are limited to those which involve two hydrogen bonds (Tinoco, 1993).

but greatly enhanced the appearance of both splicing products and intermediates with the inosine transcripts (Figure 5B, right panel). Under these conditions the ratio of lariat product to 5′ exon suggests that the inosine substitution has had only a modest effect upon step 2 of splicing. Given that other important RNA interactions may be partially disrupted by inosine substitution (for instance the Watson–Crick base pairing between G_5 of the intron and U6 snRNA), this suggests that inosine can reproduce the normal interaction between the terminal guanosines of the intron. One final observation may account for the lower level of spliced products observed with inosine-containing RNA. When inosine-containing linear molecules of RNA corresponding to the spliced mRNA product were incubated in nuclear extract, degradation of the RNA was seen at a highly accelerated rate when compared with the equivalent G-containing transcript (Figure 5E).

Together, these observations demonstrate that the presence of inosine at the first and last positions of a mammalian intron is able to duplicate the interaction which occurs between the terminal guanosine residues in a wild-type intron. This observation therefore precludes N3-amino symmetric hydrogen bonding as the mode of interaction between the terminal intron guanosines and supports the likelihood of an N1-carbonyl symmetric interaction (Figure 4; and see Discussion).

Discussion

In this paper we have described two lines of evidence that support the concept of a close interaction between the consensus splice sites at the boundaries of mammalian introns during step 2 of splicing. Furthermore, by demonstrating that substitution of inosine for guanosine allows both catalytic steps of splicing to proceed, we are able to make predictions about the precise conformation of this interaction.

Interaction between terminal nucleotides of a mammalian intron

The two lines of evidence in support of an interaction between the 5′ and 3′ splice sites during step 2 of splicing are as follows.

We have found that, as in budding yeast (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994; Ruis *et al.*, 1994), splicing that has been blocked at step 2 by mutation of either the first or last G of the intron can be restored, in part, by the presence of an A at position 1 in combination with a C at the last position of the intron (Figures 1 and 2). No other combination of nucleotides at the intron termini was able to restore splicing. However, even the IA–CI double mutant was not able to restore splicing to anywhere near wild-type levels. *In vitro*, no spliced product was detectable in the IA–CI introns (Figure 1) and, under steady-state conditions *in vivo* (Figure 2), the level of spliced RNA was <10% of wild-type levels (precise quantitation was not possible due to the interference of the endogenous TM protected band with one of the two authentic protected bands). The partial activity of the IA–CI intron is not surprising; neither of the possible A–C base pairing configurations is isomorphic with the various G–G base pairs. It is likely that the A–C pairing can only restore some of the properties of the normal G–G interaction (see below).

In the yeast actin and *rp51a* introns (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994) it was found that IA–AI introns could also splice, with an efficiency even lower than the IA–CI introns. Moreover, there is evidence in mammalian introns for mutual suppression of G→A mutations at the terminal nucleotides of a dihydrofolate reductase intron (Carothers *et al.*, 1993). We were unable

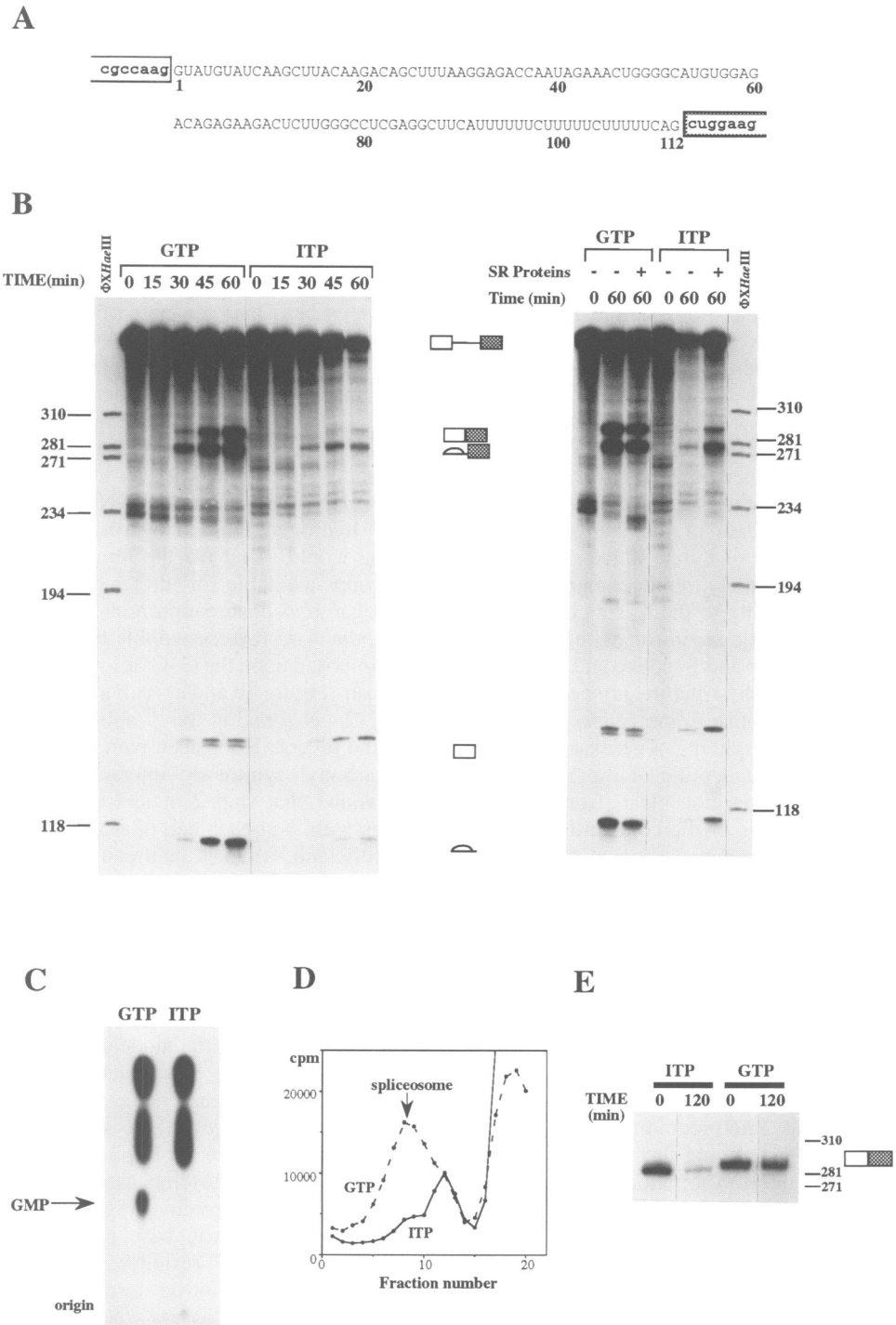

Fig. 5. Step 2 of splicing is not blocked in inosine-containing pre-mRNAs. (**A**) The sequence of the Py7 pre-mRNA. Features to note are the position of the exons which are indicated by boxes and lower case text, and the long pyrimidine tract adjacent to the 3′ splice site. The BP has not been mapped but is presumed to be A_{84} or A_{91}. (**B**) In the left hand panel, Py7 RNA transcripts synthesized using either GTP or ITP were spliced *in vitro* for 0–60 min. Splicing intermediates and products are indicated by the symbols to the right of the panel. Both lariat product and spliced mRNA were detected when I-containing transcripts were spliced, although at a reduced level compared with splicing of the wild-type transcript. In the right hand panel, Py7 RNA transcripts containing either guanosine or inosine were spliced for 60 min ± SR proteins. The SR proteins were incubated with the RNA transcripts for 30 min prior to the addition of nuclear extract. In the presence of SR proteins, splicing of the inosine-containing RNA transcripts is greatly enhanced. Splicing intermediates and products are indicated by the symbols to the left of the panel. (**C**) Thin-layer chromatography was used to analyse the GMP content of Py7 transcripts synthesized using either guanosine or inosine (see Materials and methods for conditions). An arrow indicates the position of 3′-GMP as determined by the migration of unlabelled standards. (**D**) RNA transcripts containing either G or I were incubated in nuclear extract for 60 min, and splicing complexes were then analysed on 0–30% glycerol gradients. A small amount of the splicing complex has formed on the I transcript at 60 min, although most of the counts represented degraded RNA at the top of the gradient. (**E**) RNA transcripts corresponding to spliced mRNA (ΔKP) were synthesized using either GTP or ITP and incubated in nuclear extract for 0–120 min, and then electrophoresed under denaturing conditions on polyacrylamide gels. After 120 min the transcript containing inosine is substantially more degraded than the guanosine transcript.

to detect any such excision of IA–AI introns (Figure 2). Possibly splicing of such introns was occurring in our experiments but at a level low enough to be disguised by the inefficient splicing using cryptic GU 5′ splice sites.

The second line of evidence for an association between the 5′ and 3′ splice sites during step 2 of splicing was the effect of varying the 5′ splice site +3 nucleotide upon the step 2 competition between closely spaced AGs (Figure 3). The initial rationale for these constructs was to test the hypothesis that the 5′ splice site +3 nucleotide (purine consensus) might directly interact with the consensus pyrimidine preceding the 3′ splice site AG, thus determining the competitiveness of the AG. This simple hypothesis was clearly disproven. Nevertheless, in two series of constructs (pSVTM5′N_3/3′–3UAG and CAG, Figure 3) we found reproducible variations in the usage of the competing AGs according to the nucleotide at the 5′ splice site +3 position. In our constructs, the BP, to which the 5′ splice site is covalently attached at step 2 of splicing, is over 170 nucleotides upstream of the competing 3′ splice sites. Thus, the influence of the 5′ splice site upon 3′ splice site competition is unlikely to reflect a gratuitous effect due to obligatory proximity to the 3′ splice site, as could be argued for a conventional intron with an ~18 nucleotides separation of the BP and 3′ splice site. Rather, this data argues that the branch structure is specifically retained in the region of the spliceosome active site and plays a role through step 2 of splicing. The exact nature of the interaction of the +3 nucleotide of the 5′ splice site with the 3′ splice site is not clear from these experiments. Likewise, the basis of the competitive hierarchy of NAG trincucleotides remains open to question. U6, U2 and U5 snRNAs are all possible candidates for interacting with the –3 nucleotide at the 3′ splice site to determine competitiveness (Madhani and Guthrie, 1994; Nilsen, 1994).

These data therefore support the concept that the two terminal nucleotides of the intron interact directly at step 2 of splicing. The G_1 of the intron could be thought of as providing a template for the correct positioning of the 3′ terminal G in the active site for the second transesterification reaction. This interpretation is consistent with the stereochemical data showing that the active site of the spliceosome has distinct conformations during steps 1 and 2 (Moore and Sharp, 1993); the G_1 nucleotide would be a substrate for step 1 of the reaction, but an active site template for step 2.

Likely base pairing configurations between intron terminal guanosines

Our experiments with inosine-containing RNAs (Figure 5) do not themselves directly address the occurrence of a direct interaction between the terminal intron nucleotides. Nonetheless, assuming such an interaction, the demonstration that step 2 of splicing is not blocked by inosine substitution allows informed discrimination between the four possible G–G base pairing interactions that involve at least two hydrogen bonds (Figure 4; and Tinoco, 1993). First, the N3-amino symmetric interaction (c) can be confidently ruled out since inosine is unable to form either of the hydrogen bonds. Only the N_1-carbonyl symmetric interaction (b) can be fully replicated by I–I base pairing, and so is a good candidate for the configuration of the G–G interaction. In either of the remaining two possibilities (a and f) inosine would be able to form only one of the two hydrogen bonds, which would be a much weaker interaction than the equivalent G–G base pair (Saenger, 1984). If the G at either the 5′ or 3′ splice site is involved in base triple interactions, inosine may be able to restore these further interactions despite forming only one of the usual hydrogen bonds between the intron terminal bases.

The evidence from IA–CI introns does not point unambiguously to a particular configuration for the interaction between the terminal G residues. An A–C reverse Hoogsteen base pair can be partially superimposed upon the G–G N7-amino, carbonyl–N1 base pair (Figure 4, compare a and d). This type of A–C interaction would result in significant displacement of the $C_{1'}$–N of the cytosine, but the C would maintain an appropriately positioned carbonyl group as a potential hydrogen bond acceptor. As mentioned above, inosine could reconstitute the G–G N7-amino, carbonyl–N1 base pair with only a single hydrogen bond. Nevertheless, the 3′ splice site inosine would still have the remaining O6 carbonyl and N7 groups available for further bonding and would have the sugar–phosphate appropriately positioned. Alternatively, the A–C reverse wobble base pair can be partially superimposed upon the G–G or I–I N1-carbonyl symmetric base pair (Figure 4b and e) with a less severe displacement of the $C_{1'}$–N_1 bond at the 3′ splice site than would occur with the reverse Hoogsteen pair. In further support of the N1 carbonyl symmetric interaction (Figure 4b) is the observation that step 2 of splicing is partially active in introns with A at each end (Carothers *et al.*, 1993; Parker and Siliciano, 1993; Chanfreau *et al.*,1994). The only A–A and G–G base pairs that can be readily superimposed are the A–A N_1 amino symmetric and the G–G N_1 carbonyl symmetric. However, these are not strictly isomorphous, which may account for the weak ability of an A–A interaction to restore step 2 of splicing.

The available data do not point unequivocally to a unique possibility for the mode of interaction of the intron terminal guanosines. A consideration of IA–CI introns does not allow a clear choice between models 'a' and 'b' (Figure 4). However, the modest effect of inosine substitution upon step 2 of splicing *in vitro*, particularly when efficient spliceosome assembly has been restored by addition of SR proteins (Figure 5), strongly suggests N1-carbonyl symmetric base pairing between the conserved splice site G residues (Figure 4). Site specific substitution at the intron termini (Moore and Sharp, 1992) would help to assess whether step 2 of splicing is unaffected by inosine substitution, which would definitively support model 'b' (Figure 4) or whether there is a discernible effect, supportive of model 'a'.

Materials and methods

Constructs

Constructs used as templates for *in vitro* transcription or for transfection were all prepared by standard cloning procedures (Sambrook *et al.*, 1989). The parental construct, pG+30, contains α-TM exons 2 and 3 with a spacer element just downstream of the 5′ splice site to relieve steric interference (Smith and Nadal-Ginard, 1989). In the constructs $pTMN_1$, the mutated 5′ splice site was generated by the insertion of the sequence CAAGNTGAGTGGGAGCT between the *Sac*I and *Sty*I sites flanking the 5′ splice site. The 3′ splice site was mutated by the insertion

of the sequence CTCCTCCACAG between the *Acc*I and *Pvu*II sites, simultaneously deleting 93 nucleotides of the α-TM intron. The constructs $pTMN_1$/ANGG were subsequently generated by inserting the degenerate sequence CTCCTCCANGG between the *Acc*I and *Pvu*II sites of each of the four $pTMN_1$ constructs. The constructs $pTM5'N_3/3'$–3NAG were based on the constructs –3NAG described previously (Smith *et al.*, 1993). Additional mutations at the 5′ splice site of these constructs were made by insertion of the degenerate oligonucleotide CAAGGTNA-GTGGGAGCT between the *Sac*I and *Sty*I sites. Constructs for transfection ($pSVTMN_1$, $pSVN_1$/ANGG and $pSVTM5'N_3/3'$–3NAG) were generated by subcloning the *Hin*dIII fragment containing the α-TM exons and intron from the corresponding construct for *in vitro* transcription as described above. The construct Py7 contains α-TM exons 2 and 3 with a synthetic intron; the 5′ end partially derived from the human β-globin gene (Figure 5A). The construct ΔKP comprises spliced exons 2 and 3 of α-TM, generated by deletion between the *Kpn*I and *Pvu*II sites at the ends of the intron. The identity of all constructs was confirmed by sequencing.

In vitro transcription and splicing reactions

Standard ^{32}P-labelled RNA transcripts were transcribed from pGEM vectors with SP6 or T7 RNA polymerase (Promega) using a G(5′)ppp(5′)G dinucleotide primer as described previously (Smith and Nadal-Ginard, 1989; Mullen *et al.*, 1991). For synthesis of inosine-containing transcripts, transcription reactions containing only nucleotides A, C and U were pre-incubated for 20 min at 40°C before adding ITP. RNA transcripts for use as probes in RNase protection analysis were not capped, and after transcription were purified from 8 M urea–6% polyacrylamide gels by electroelution. HeLa cell nuclear extracts were prepared using the modifications of Abmayr *et al.* (1988). Standard splicing reactions contained 20–50 fmol [^{32}P]RNA transcript, 2.5 mM $MgCl_2$, 500 μM ATP, 20 mM creatine phosphate, 2500 U/ml RNasin, 12 mM Tris (pH 7.9), 12% (v/v) glycerol, 60 mM KCl, 0.12 mM EDTA, 0.3 mM DTT and 40–60% nuclear extract, and were incubated at 30°C. Splicing reactions were then subjected to proteinase K digestion and phenol–chloroform extraction. Reaction products were analysed by electrophoresis in 8 M urea–4–8% polyacrylamide gels, followed by autoradiography. SR proteins were prepared from sheep uterine smooth muscle according to the method of Zahler *et al.* (1992). The activity of the SR proteins was determined by their ability to activate splicing in a HeLa S-100 extract.

Composition analysis of RNA transcripts

[^{32}P]UTP-labelled RNA transcripts were digested with RNase A and RNase T1 (Domdey *et al.*, 1984) and the digestion products were fractionated by TLC on PEI plates using 1 M LiCl. This method was used to determine the IMP content of transcripts. Alkali hydrolysis (Kozak and Shatkin, 1977) was used to completely degrade [^{32}P]UTP-labelled RNA transcripts for subsequent determination of GMP content. Hydrolysed RNA was fractionated by TLC on PEI plates where the solvent was 0.65 M Na formate, pH 3.4.

Analysis of splicing complexes

Splicing reactions containing ~100 fmol [^{32}P]RNA transcript were set up under standard splicing conditions, but were not subjected to treatment with proteinase K or phenol–chloroform. Samples were layered onto 11 ml glycerol gradients [20 mM Tris-HCl, 1 mM $MgCl_2$, 60 mM KCl and 10–30% glycerol (v/v)] and subjected to centrifugation at 40 000 r.p.m. for 3 h at 4°C in a Beckman SW40 Ti rotor. Fractions of 500 μl were collected from the bottom to top of each gradient, and analysed by Cherenkov counting.

Cell culture and transfection

COS cells were grown under standard conditions in DMEM supplemented with 10% fetal calf serum. Transient transfection was carried out by calcium phosphate co-precipitation as described previously (Smith and Nadal-Ginard, 1989; Mullen *et al.*, 1991).

Detection of expressed RNA

Transiently expressed RNA was analysed by either reverse transcription followed by nested PCR (Figures 2B and 3) or RNase protection analysis (Figure 2D). Cytoplasmic RNA was harvested by the hot phenol method described previously (Smith and Nadal-Ginard, 1989; Mullen *et al.*, 1991). For RT-PCR (Figures 2B and 3), 5 μg of cytoplasmic RNA was denatured at 65°C for 10 min, snap cooled on ice and then mixed with 50 ng of RT primer SV3′3 in the presence of 2 mM dNTPs, 50 mM Tris-HCl, pH 8.3, 6 mM $MgCl_2$ and 40 mM KCl. Annealing was at 42°C for 1 h, after which ~10 U of AMV reverse transcriptase (Promega) was added, and incubation was carried out for a further 45 min at 42°C. For PCR, 1 μl of the RT reaction was used as template in 25 μl reactions. PCR buffer was as supplied by Promega, but with $MgCl_2$ at 1.0 mM. Unlabelled 5′ primer SV5′1 and 3′ primer E3R (or SV3′1) were added to 250 ng for standard PCR. For labelled PCR, unlabelled 5′ primer SV5′1 was added to 250 ng, the 5′ ^{32}P-labelled 3′ primer E3R was added to 10 ng, and the unlabelled 3′ primer E3R was added to 100 ng. A hot start protocol was used in which the 3′ primer and *Taq* polymerase (1 U; Promega) were added to the remainder of the reaction at 80°C. Cycling conditions were: 94°C, 30 s, 62°C, 30 s, 72°C, 1 min, 30–35 cycles. Samples were then phenol–chloroform extracted, ethanol-precipitated and separated on either 8 M urea–4–6% polyacrylamide gels or 2% agarose gels. The relative ratios of bands of labelled PCRs were determined using a Molecular Dynamics PhosphorImager. In control reactions, we determined that the relative ratios of bands corresponding to use of competing 3′ splice sites did not vary over 20–35 cycles of PCR. Sequences of oligonucleotides used for RT-PCR are: **SV3′3**, GGTACAGACCTGTGGCTGAGTTTGC; **SV5′1**, GAGCTATT-CCAGAAGTAGTGTGAGGAG; **E3R**, CTTGAGTTTCTTTTGCAGT-GACACC; and **SV3′1**, CTCACTGCGTTCCAGGCAATGCT. The probe used for RNase protection analysis was an RNA transcript synthesized *in vitro* using the construct $pTMA_1$/AUGG as a template linearized within the intron with *Xho*I. RNA samples from COS cells (20 μg) were mixed with probe (~50 000 c.p.m.) in hybridization buffer (80% formamide, 40 mM PIPES, pH 6.4, 400 mM NaCl and 1 mM EDTA) to 30 μl, denatured at 85°C for 5 min, and incubated for 8–10 h at 45°C. Samples were digested with RNase A (20 μg/ml) and RNase T1 (35 U/ml) in 350 μl digestion buffer (10 mM Tris–HCl, pH 7.5, 300 mM NaCl and 5 mM EDTA) for 30 min at 30°C. RNase protection reactions were subsequently subjected to proteinase K digestion and phenol–chloroform extraction. Reaction products were analysed by electrophoresis in 8 M urea–6% polyacrylamide gels, followed by autoradiography.

Determination of precise splice junctions

Unlabelled RT-PCR products amplified using SV5′1 and SV3′1 were digested with *Hin*dIII and subcloned into pBSII (KS+) (Stratagene). Double-stranded DNA sequencing was carried out according to the Sequenase method (USB) using the oligonucleotide TM2F, CCGGGT-GCTGGAGGAGCTGCAC, complementary to an α-TM sequence within exon 2. At least five randomly selected clones were sequenced in order to unambiguously determine the sequence of the splice junction. If more than one sequence was found, then between 5 and 25 further selected clones were sequenced.

Acknowledgements

We thank Andy Newman and Gavin Roberts for critically reading the manuscript, and Clare Gooding for providing the SR proteins. This work was supported by a grant from The Wellcome Trust to C.W.J.S. A.D.J.S. is supported by The Cambridge Commonwealth Trust and The New Zealand Federation of University Women (Inc.).

References

Abmayr,S.M., Reed,R. and Maniatis,T. (1988) Identification of a functional mammalian spliceosome containing unspliced pre-mRNA. *Proc. Natl Acad. Sci. USA*, **85**, 7216–7220.

Aebi,M., Hornig,H., Padgett,R.A., Reiser,J. and Weissman,C. (1986) Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell*, **47**, 555–565.

Carothers,A.M., Urlaub,G., Grunberger,D. and Chasin,L.A. (1993) Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol. Cell. Biol.*, **13**, 5085–5098.

Chanfreau,G. and Jacquier,A. (1993) Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J.*, **12**, 5137–5180.

Chanfreau,G., Legrain,P., Dujon,B. and Jacquier,A. (1994) Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3′ splice site selection in *S.cerevisiae*. *Nucleic Acids Res.*, **22**, 1981–1987.

Cortes,J.J., Sontheimer,E.J., Seiwart,S.D. and Steitz,J.A. (1993) Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5′ splice sites *in vivo*. *EMBO J.*, **12**, 5181–5189.

Domdey,H., Apostol,B., Lin,R.-J., Newman,A., Brody,E. and Abelson,J. (1984) Lariat structures are *in vivo* intermediates in yeast pre-mRNA splicing. *Cell*, **39**, 611–621.

Green,M.R. (1991) Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Biochem.*, **7**, 559–599.

Kozak,M. and Shatkin,A.A. (1977) Sequences of two 5′-terminal ribosome-protected fragments from Reovirus messenger RNAs. *J. Mol. Biol.*, **112**, 75–96.

Madhani,H.D. and Guthrie,C. (1994) Dynamic RNA–RNA interactions in the spliceosome. *Annu. Rev. Genet.*, **28**, 1–26.

Moore,M.J. and Sharp,P.A. (1992) Site-specific modification of pre-mRNA: the 2′-hydroxyl groups at the splice sites. *Science*, **256**, 992–997.

Moore,M.J. and Sharp,P.A. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*, **365**, 364–368.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

Mullen,M.P., Smith,C.W.J., Patton,J.G. and Nadal-Ginard,B. (1991) α-Tropomyosin mutually exclusive exon selection: competition between branch-point/polypyrimidine tracts determines exon choice. *Genes Dev.*, **5**, 642–655.

Newman,A. (1994) Activity in the spliceosome. *Curr. Biol.*, **4**, 462–464.

Newman,A.J. and Norman,C. (1992) U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell*, **68**, 743–754.

Newman,A.J., Lin,R.J., Cheng,S.C. and Abelson,J. (1985) Molecular consequences of specific intron mutations on yeast mRNA splicing *in vivo* and *in vitro*. *Cell*, **42**, 335–344.

Nilsen,T.W. (1994) RNA–RNA interactions in the spliceosome: unraveling the ties that bind. *Cell*, **78**, 1–4.

Padgett,R.A., Podar,M., Boulanger,S.C. and Perlman,P.S. (1994) The stereochemical course of group II intron self-splicing. *Science*, **266**, 1685–1688.

Parker,R. and Siliciano,P.G. (1993) Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature*, **361**, 660–662.

Peebles,C.L., Perlman,C.S., Mecklenburg,K.L., Petrillo,M.L., Tabor,J.H., Jarrell,K.A. and Cheng,H.L. (1986) A self-splicing RNA excises an intron lariat. *Cell*, **44**, 213–223.

Reich,C.I., VanHoy,R.W., Porter,G.L. and Wise,J.A. (1992) Mutations at the 3′ splice site can be suppressed by compensatory base changes in U1 snRNA in fission yeast. *Cell*, **69**, 1159–1169.

Ruis,B.L., Kivens,W.J. and Siliciano,P.G. (1994) The interaction between the first and last intron nucleotides in the second step of pre-mRNA splicing is independent of other conserved intron nucleotides. *Nucleic Acids Res.*, **22**, 5190–5195.

Saenger,W. (1984) C.R.Cantor, *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.

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.

Shapiro,M.B. and Senapathy,P. (1987) RNA splicing junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.*, **15**, 7155–7174.

Smith,C.W.J., Porro,E.B., Patton,J.G. and Nadal-Ginard,B. (1989) Scanning from an independently specified branch-point defines the 3′ splice site of mammalian introns. *Nature*, **342**, 243–247.

Smith,C.W.J. and Nadal-Ginard,B. (1989) Mutually exclusive splicing of tropomyosin exons enforced by an unusual lariat branch-point location: implications for constitutive splicing. *Cell*, **56**, 749–758.

Smith,C.W.J., Chu,T.T. and Nadal-Ginard,B. (1993) Scanning and competition between AGs are involved in 3′ splice site selection in mammalian introns. *Mol. Cell. Biol.*, **13**, 4939–4952.

Sontheimer,E.J. and Steitz,J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989–1996.

Staknis,D. and Reed,R. (1994) SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.*, **14**, 7670–7682.

Tinoco,I. (1993) Appendix 1: Structures of base pairs involving at least two hydrogen bonds. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Weiner,A.M. (1993) mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell*, **72**, 161–164.

Wieringa,B., Meyer,F., Reisser,J. and Weissmann,C. (1983) Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit beta-globin gene. *Nature*, **301**, 38–43.

Yisraeli,J.K. and Melton,D.A. (1989) Synthesis of long, capped transcripts *in vitro* by SP6 and T7 RNA polymerases. *Methods Enzymol.*, **180**, 42–50.

Zahler,A.M., Lane,W.S., Stolk,J.A. and Roth,M.B. (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.*, **6**, 837–847.

Received on March 6, 1995; revised on April 24, 1995