The EMBO Journal vol.10 no.9 pp.2635–2644, 1991

Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants

Gregory J.Goodall[1] and Witold Filipowicz

Friedrich Miescher-Institut, PO Box 2543, 4002 Basel, Switzerland
[1]Present address: Institute of Medical and Veterinary Science, Adelaide, South Australia, 5000, Australia

Communicated by A.-L.Haenni

We have found previously that the sequences important for recognition of pre-mRNA introns in dicot plants differ from those in the introns of vertebrates and yeast. Neither a conserved branch point nor a polypyrimidine tract, found in yeast and vertebrate introns respectively, are required. Instead, AU-rich sequences, a characteristic feature of dicot plant introns, are essential. Here we show that splicing in protoplasts of maize, a monocot, differs significantly from splicing in a dicot, *Nicotiana plumbaginifolia*. As in the case of dicots, a conserved branch point and a polypyrimidine tract are not required for intron processing in maize. However, unlike in dicots, AU-rich sequences are not essential, although their presence facilitates splicing if the splice site sequences are not optimal. The lack of an absolute requirement for AU-rich stretches in monocot introns is reflected in the occurrence of GC-rich introns in monocots but not in dicots. We also show that maize protoplasts are able to process a mammalian intron and short introns containing stem–loops, neither of which are spliced in *N.plumbaginifolia* protoplasts. The ability of maize, but not of *N.plumbaginifolia* to process stem–loop-containing or GC-rich introns suggests that one of the functions of AU-rich sequences during splicing of dicot plant pre-mRNAs may be to minimize secondary structure within the intron.

Key words: introns/plant genes/RNA processing/RNA structure/transient expression in protoplasts

Introduction

Splicing of pre-mRNAs in the nucleus follows similar pathways in most eukaryotes studied to date. The reaction occurs in two transesterification steps which result in ligation of the exons and excision of the intron in the form of a branched lariat. The whole process is mediated by a complex particle called the spliceosome, which is assembled from several distinct small nuclear ribonucleoprotein particles (snRNPs U1, U2, U5 and U4/U6) and various protein factors in addition to the substrate RNA (reviewed by Green, 1986; Sharp, 1987; Guthrie and Patterson, 1988; Mattaj, 1990). Pre-mRNA splicing has been most extensively studied in vertebrates and in the yeast *Saccharomyces cerevisiae*. While there is much in common in the two splicing systems, there are some significant differences, particularly in the structural elements in the pre-mRNA that define the intron and the way these elements are recognized. In both vertebrates and yeast, the 5′ splice site is determined by the binding of U1 snRNP, mediated in part by base pairing interactions between the 5′ splice site and the 5′ end of the U1 RNA (Black *et al.*, 1985; Zhuang and Weiner, 1986; Siliciano and Guthrie, 1988). However, the way in which the 3′ region of the intron is defined is different in the two systems. In *S.cerevisiae* the hallmark of this region is a highly conserved branch point UACUAAC (reviewed by Green, 1986); its recognition by U2 snRNP is determined primarily by base pairing with a segment of U2 RNA (Parker *et al.*, 1987). Vertebrate introns do not have highly conserved branch point sequences, but contain instead a pyrimidine-rich tract usually positioned upstream of the 3′ splice site (Green, 1986; Sharp, 1987; Mattaj, 1990). Proteins interacting with the polypyrimidine tract are required for stable binding of U2 snRNP (Ruskin *et al.*, 1988; Zamore and Green, 1991; and references therein) and may also help to select the 3′ splice site later in the reaction (Gerke and Steitz, 1986; Tazi *et al.*, 1986; Ruskin *et al.*, 1988).

We have shown previously that the structural requirements for intron processing in plants differ from those in vertebrates and yeast. Using synthetic model genes transiently expressed in protoplasts from a dicot plant, *Nicotiana plumbaginifolia*, we demonstrated that the only sequence elements necessary for intron processing in this system are the splice sites and a high A+U nucleotide content in the intron. Neither conserved branch point sequences nor a polypyrimidine tract, similar to those found in yeast or vertebrates, were found to be essential (Goodall and Filipowicz, 1989; reviewed by Goodall *et al.*, 1991). These findings are consistent with the observations that all dicot plant introns contain a minimum of ~60% AU (Goodall and Filipowicz, 1989) and that mammalian introns, which are usually not AU-rich, are generally not spliced when their processing is tested in transgenic plants or protoplasts (Barta *et al.*, 1986; van Santen and Spritz, 1987; Wiebauer *et al.*, 1988; Pautot *et al.*, 1989). Apart from the enrichment in A+U nucleotides and a lack of 3′-proximal polypyrimidine tracts, plant introns resemble vertebrate introns. They have a minimum length of ~70 nt but can be thousands of nucleotides in length. The consensus sequences at the 5′ and 3′ splice sites, AG/GTAAG and TGCAG/G respectively, resemble the vertebrate consensus (reviewed by Goodall *et al.*, 1991). Indeed, many plant introns, in particular those that happen to contain polypyrimidine tract-like sequences, are faithfully processed in mammalian systems (Brown *et al.*, 1986; Hartmuth and Barta, 1986; van Santen and Spritz, 1987; Wiebauer *et al.*, 1988).

We are interested in understanding the mechanism of intron recognition in plants and, in particular, in establishing the function of the AU-rich sequences in this process. The latter question may be of more general importance. Pre-mRNA introns in *Drosophila*, nematodes, ciliates and slime molds are also very AU-rich (discussed by Goodall and Filipowicz, 1989; Csank *et al.*,1990), although it is not

© Oxford University Press

known whether this feature is important for splicing in these organisms. In view of the requirement of the AU-rich sequences for intron processing in dicot plants, it was rather surprising to find that in monocot plants, in marked contrast to dicots, introns can contain as little as 31% A+U (see Figure 1). Keith and Chua (1986) have observed previously that a maize and a wheat intron are not as efficiently spliced as a dicot intron in transgenic tobacco plants, suggesting the possibility of differences in pre-mRNA splicing between dicots and monocots. In this work, we have compared the requirements for intron processing in maize, a monocot, and *N.plumbaginifolia*, a dicot. We find that splicing in monocots is considerably more 'permissive' than splicing in dicots. Maize, but not *N.plumbaginifolia*, can splice natural and synthetic introns which are GC-rich, or introns which contain stem–loop structures. Furthermore, a mammalian GC-rich intron is faithfully processed in maize but not in *N.plumbaginifolia* protoplasts.

Results

Splicing in monocots does not require AU-rich introns

In marked contrast to dicot introns, which all contain >59% AU, the AU content of monocot introns is much more variable. About 20% of introns in monocots contain more GC than AU, and several are <40% AU (Figure 1). The difference in the AU bias between monocot and dicot introns is also seen if one plots the average occurrence of A+U or G+C nucleotides in the regions extending 50 bp upstream and downstream of each of the splice sites. The introns of dicot plants average 74% AU, compared with an average 55% AU in the flanking exon sequences. For monocot genes these numbers are 59% and 44%, respectively (Goodall and Filipowicz, 1989, and data not shown). Differences between monocots and dicots in respect to the AU content in genes and introns have also been noted by researchers studying the isochore organization of plant genomes (Montero *et al.*, 1990).

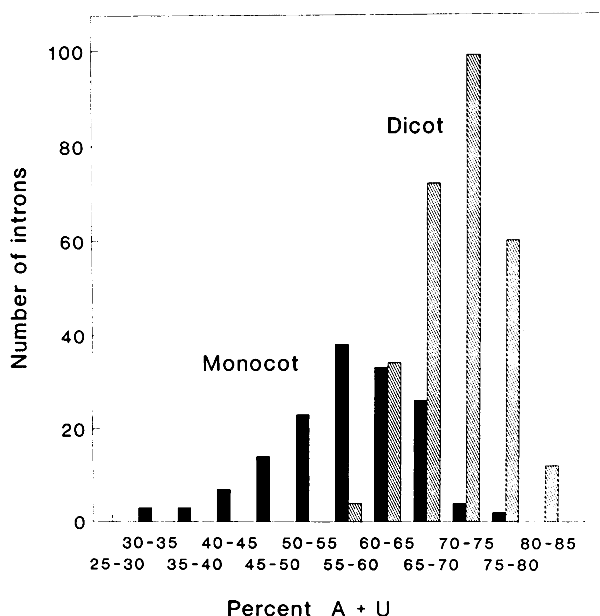

Fig. 1. Content of A+U nucleotides in 146 monocot (hatched bars) and 280 dicot (black bars) plant introns. All four dicot introns in the 55–60% class contain 59% AU. The most AU-poor monocot intron is the IVS1 of barley aleurain (31% AU; Whittier *et al.*, 1987).

To investigate whether the lower content of AU in monocot introns is symptomatic of a real difference in the mode of intron recognition, we have compared the processing of introns originating from monocot and dicot genes in transfected maize and *N.plumbaginifolia* protoplasts. The dicot introns used were IVS1, 2, 4 and 5 of the bean phaseolin gene and IVS 1 of the soybean leghemoglobin gene, while the monocot introns were IVS1 and 2 of actin and IVS9 and 10 of *waxy* genes of maize. The AU content of the dicot introns ranges from 69 to 77% and that of monocot introns from 40 to 62% (Table I). Splicing was analyzed by RNase A/T1 mapping using RNA isolated from transfected protoplasts and ^{32}P-labeled RNA probes complementary to the unspliced form of the RNA. The results are presented in Figure 2 and Table I. All the introns tested were accurately and efficiently spliced in maize protoplasts. On the other hand, only the AU-rich maize actin and dicot introns, were spliced in dicot protoplasts. The GC-rich monocot introns, maize *waxy* IVS9 (40% AU) and IVS10 (42% AU), were not processed at all in this system (see also Figure 5B).

We have previously used model synthetic genes, *syn*7 and its derivatives, to establish that AU-rich sequences are required for intron processing in dicots (Goodall and Filipowicz, 1989). In *syn* genes the introns and the flanking exons are composed of arbitrary sequences apart from the putative branch point and the splice sites. We have compared the activity of the synthetic introns in protoplasts of maize and *N.plumbaginifolia*. The construct *syn*7, containing an 85 nt intron with 75% AU, was efficiently spliced in both types of protoplasts. Replacement of AU-rich sequences within the intron by GC-rich segments (genes *syn*13, 23 and

Table I. Efficiency of processing of the plant introns in protoplasts of *N.plumbaginifolia* and maize

Intron	% AU	% of splicing[a]	
		Nicotiana	Maize
Dicot introns			
Bean phaseolin IVS 1	71	66	83
2	69	54	74
4	74	68	81
5	77	66	81
Soybean leghemoglobin IVS 1	74	74	91
Monocot introns			
Maize actin IVS 1	62	42	80
2	59	78[b]	89
Maize *waxy* IVS 9	40	ND[c]	95
10	42	ND[c]	95

[a]We have verified that the level of splicing in both *N.plumbaginifolia* (Goodall and Filipowicz, 1989) and maize (Figure 3B, lanes *syn*17; data not shown) protoplasts is not affected by variations is transfection efficiencies. It is unlikely that observed differences in the percentage of splicing are caused by different stabilities of unspliced RNAs in two types of protoplasts. In marked contrast to the mammalian cells, unspliced pre-mRNAs accumulate in *Nicotiana* and maize protoplasts and plants (Keith and Chua, 1986; Goodall and Filipowicz, 1989; Nash *et al.*, 1990; Luehrsen and Walbot, 1991; this work).
[b]Very efficient splicing of this intron which contains only 59% AU is readily explained by its nearly optimal splice sites, TG/GTAAG and TGCAG/G.
[c]Splicing not detectable.

24) progressively eliminated intron processing in *N.plumbaginifolia* but had only a relatively small, though significant, effect on processing in maize protoplasts (Figure 3A). Hence, GC-rich introns containing as much as 73% GC (intron *syn*24) can be faithfully processed in monocot cells.

A 3′ polypyrimidine tract and a conserved branch point are not required for splicing in maize

The data presented above indicate that splicing in monocots, contrary to that in dicots, shows no absolute requirement for AU-rich sequences in the intron. We have tested whether a 3′ splice-site proximal polypyrimidine tract, known to be essential for splicing in vertebrates, or the conserved branch-point essential for intron processing in yeast, are of importance for intron excision in maize. Previous work has shown that neither of these signals is essential for splicing in dicots (Goodall and Filipowicz, 1989). The synthetic intron *syn*35, which has a 3′-proximal 15 nt polypurine tract interrupted by a single C residue, was processed in maize as efficiently as in *N.plumbaginifolia* (Figure 3B). The intron *syn*46, which is a derivative of the A-rich intron *syn*17, contains a perfect 15 nt polypurine tract of different sequence from *syn*35. This was also efficiently spliced in both types of protoplasts (Figure 3B).

The *syn*7 intron contains a potential branch site sequence CUAAC 30 nt upstream of the 3′ splice site. Modification of this sequence has little effect on splicing in *N.plumbaginifolia*, and this effect can only be demonstrated when neighboring A residues, likely to act as alternative branch acceptors, are eliminated from the central region of the intron (Goodall and Filipowicz, 1989). We have tested the activity in maize protoplasts of the construct *syn*36, which has no adenosines beyond position 15 in the intron and in which a putative branch point sequence is changed from CUAAC to UCUCU. Intron *syn*36 was spliced with identical

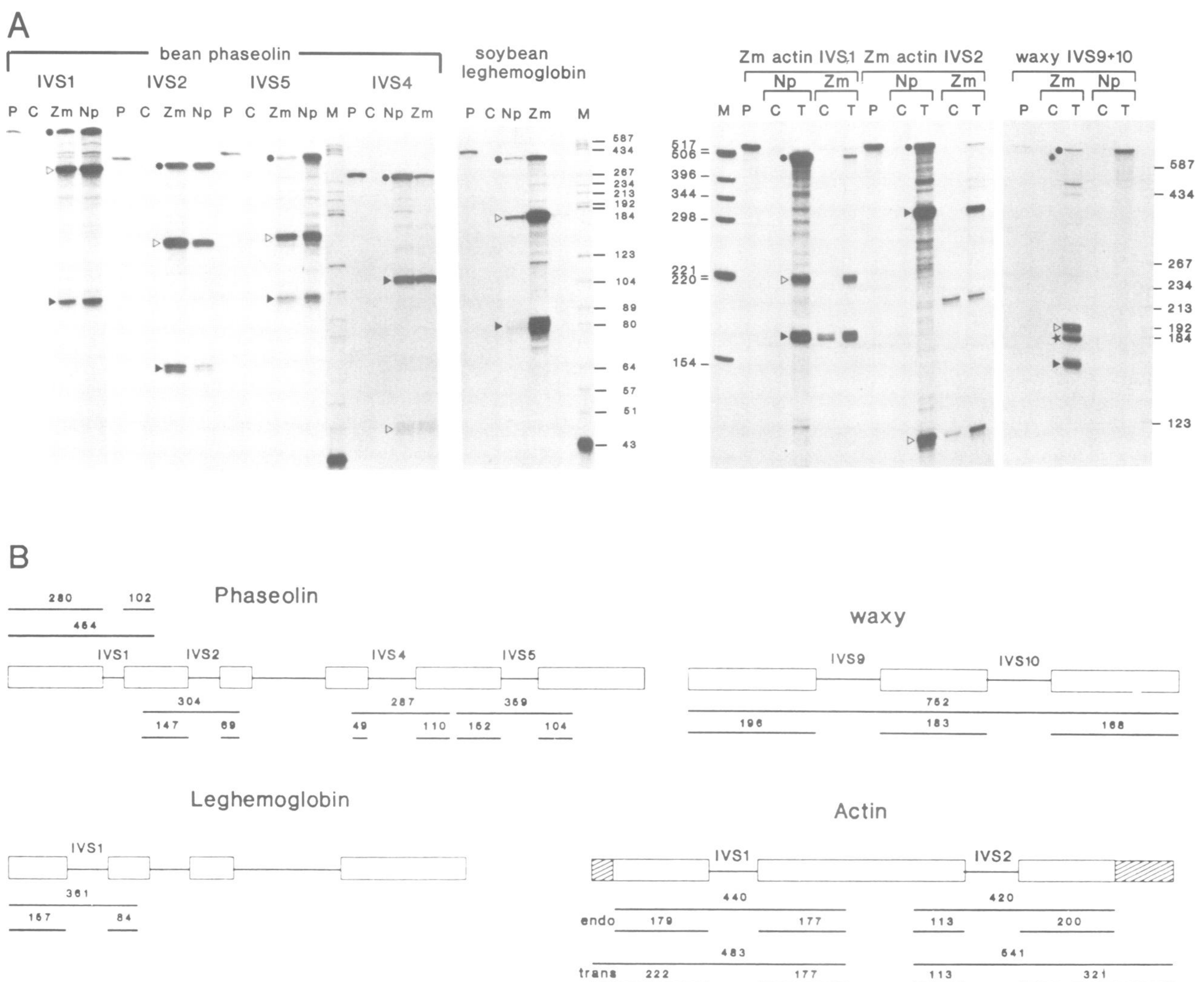

Fig. 2. Processing of dicot and monocot introns in protoplasts of *N.plumbaginifolia* and maize. (**A**) RNase protection analysis. Left panel: lanes P, undigested probe; lanes C, analyses with RNA from mock transfected maize protoplasts; lanes Zm and Np, RNA from transfected maize and *N.plumbaginifolia* protoplasts, respectively. Right panel: lanes P, undigested probe; lanes C and T, RNA from control and transfected protoplasts, respectively. Lanes M, size markers (end-labeled *Hin*fI or *Hae*III digest of pBR322). Fragments protected by unspliced RNA are marked with dots, while those protected by the upstream and downstream exons of spliced RNA are indicated by the open and closed triangles, respectively. The fragment corresponding to the middle exon of *waxy* is marked with asterisk. Fragments seen in control (C) lanes of the actin mRNA mapping with maize RNA are due to the endogenous actin mRNA (see panel B). (**B**) Schemes of the transfected genes or gene fragments showing the sizes of expected probe fragments protected by unspliced and spliced RNAs. Hatched boxes in the actin gene scheme indicate adjacent vector sequences present in the transcript of transfected (trans) gene but not of the endogenous (endo) actin gene. The probes extend into these sequences and hence distinguish between the endogenous and exogenous transcripts.

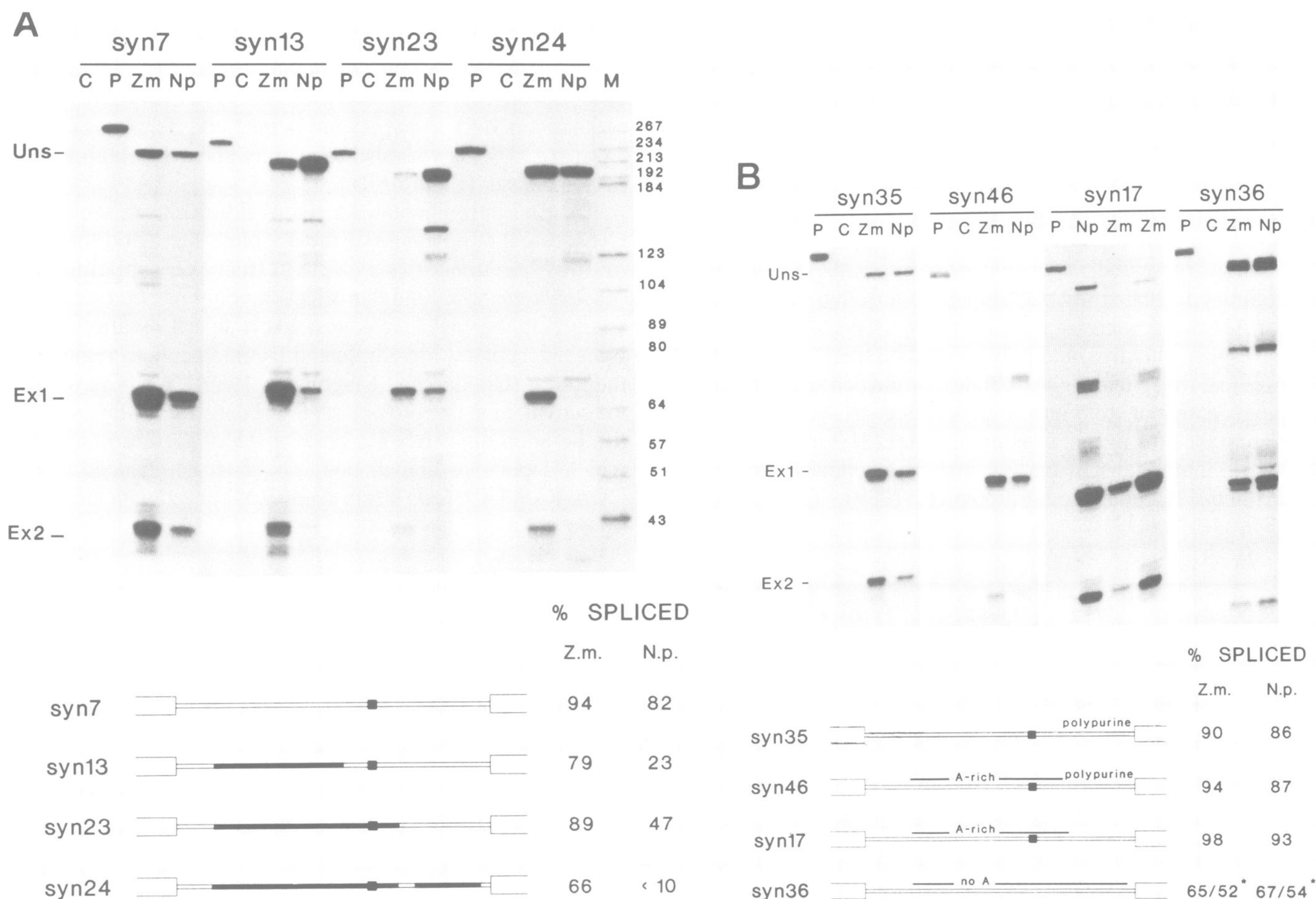

Fig. 3. RNase protection analysis of the splicing of synthetic introns. The synthetic introns are shown schematically below, along with their splicing efficiencies in maize (Z.m.) and *N.plumbaginifolia* (N.p.). For intron sequences, see Goodall and Filipowicz, 1989, and Materials and methods. Lanes P, undigested probe; lanes C, analyses with RNA from mock transfected maize protoplasts; lanes Zm and Np, RNA from transfected maize and *N.plumbaginifolia* protoplasts; lane M, size markers. The approximate sizes of expected protected probe fragments are 191 nt (unspliced RNA, uns), 63 nt (exon 1) and 43 nt (exon 2). The probe region protected by exon 2 contains less ^{32}P and thus gives a fainter signal than the exon 1-protected fragment. Identity of the additional 120–150 nt fragments seen in some analyses is not known. (**A**) Analysis of introns of various A+U content. The black bars represent GC-rich segments replacing AU-rich sequences of the *syn*7 intron. Black boxes represent the putative branch point sequence. (**B**) Analysis of introns designed to test the requirement for a polypyrimidine tract and importance of the branch point sequence. The region in *syn*17 and *syn*46 in which the A-rich sequence replaces the U-rich sequence of *syn*7 is indicated. The *syn*36 intron has no A residues at positions 16–83. Purine-rich intron 3′ regions of the *syn*35 and *syn*46 are GGGAAAAACGGAAAATGCAG and GGAAAGAAGAAAGAATGCAG, respectively. The two values listed for *syn*36 (asterisk) indicate that the two protected exon fragments were not found in equimolar amounts, the exon 1 fragment being in excess (see text).

efficiencies in maize and *N.plumbaginifolia* (Figure 3B), indicating that in monocots too, the criteria for branch point selection are rather relaxed and residues other than A are likely to be used for lariat formation. In both systems cleavage at the 5′ splice site occurred more efficiently than at the 3′ splice site. This could reflect the partial usage of a cryptic 3′ acceptor site positioned further downstream in the transcript, or could be due to the accumulation of the intron–exon 2 intermediate after the first step of splicing (see Goodall and Filipowicz, 1989). Hornig *et al.* (1986) have found that the second step of splicing is inhibited when branching occurs to G or U residues.

These data indicate that in monocots, as in dicots, the requirements for branch site are relaxed and a polypyrimidine tract is not necessary. These conclusions are supported by the analysis of monocot intron sequences, both AU-rich and GC-rich. Monocot introns do not show the 3′ end proximal enrichment in pyrimidine nucleotides characteristic of vertebrate introns. Similarly, no sequences resembling the conserved UACUAAC branch point region of yeast genes were identified in monocot or dicot introns (data not shown; see also Brown, 1986; Wiebauer *et al.*, 1988; Goodall *et al.*, 1991).

In monocots, optimal splice site sequences obviate the need for high AU content

If neither AU-richness nor polypyrimidine tracts nor conserved branch sites are required for intron processing in maize protoplasts, what are the signals which help to define intron sequences in this system? Since the only elements that were not yet tested are the splice sites, we have evaluated their importance for processing of GC-rich and AU-rich introns in maize. Comparison of the splice site sequences of the most GC-rich monocot introns (50–69% GC) with the most AU-rich introns (66–76% AU) available in the databases indicate that the GC-rich introns tend to have better matches to the 5′ and 3′ consensus sequences than the AU-rich introns. This is particularly evident at positions 4 and

Table II. Frequencies of consensus nucleotides at the splice sites of 27 AU-rich and 25 GC-rich monocot introns

Introns	5′ splice site							3′ splice site					
	−2	−1	1	2	3	4[a]	5	−5	−4[b]	−3	−2	−1	1
Total monocots	A_{60}	G_{80}	G_{100}	T_{100}	A_{69}	G_{5}	G_{50}	T_{62}	G_{48}	C_{84}	A_{100}	G_{100}	G_{59}
AU-rich	A_{67}	G_{78}	G_{100}	T_{100}	A_{70}	G_{11}	G_{48}	T_{63}	T_{44}	C_{74}	A_{100}	G_{100}	G_{56}
GC-rich	A_{44}	G_{88}	G_{100}	T_{100}	A_{66}	G_{0}	G_{88}	T_{56}	G_{76}	C_{92}	A_{100}	G_{100}	G_{76}

The subscripts denote the percentage of the consensus bases. The A+U content of the 27 most AU-rich introns present in the EMBL and GenBank databases is 66−76%; the 25 most GC-rich introns contain 31−50% AU.
[a]Strong bias against G at this position is characteristic for plant introns. Other nucleotides at this position are: A-38%, T-27%, C-29% (total monocot); A-56%, T-19%, C-15% (AU-rich); A-40%, T-12%, C-48% (GC-rich). Dicot introns have at this position: 57% A, 4% G, 24% T and 15% C (Goodall *et al.*, 1991).
[b] Other nucleotides at this position are: A-15%, T-26%, C-11% (total monocot); G-26%, A-22%, C-7% (AU-rich); A-4%, T-12%, C-8% (GC-rich).

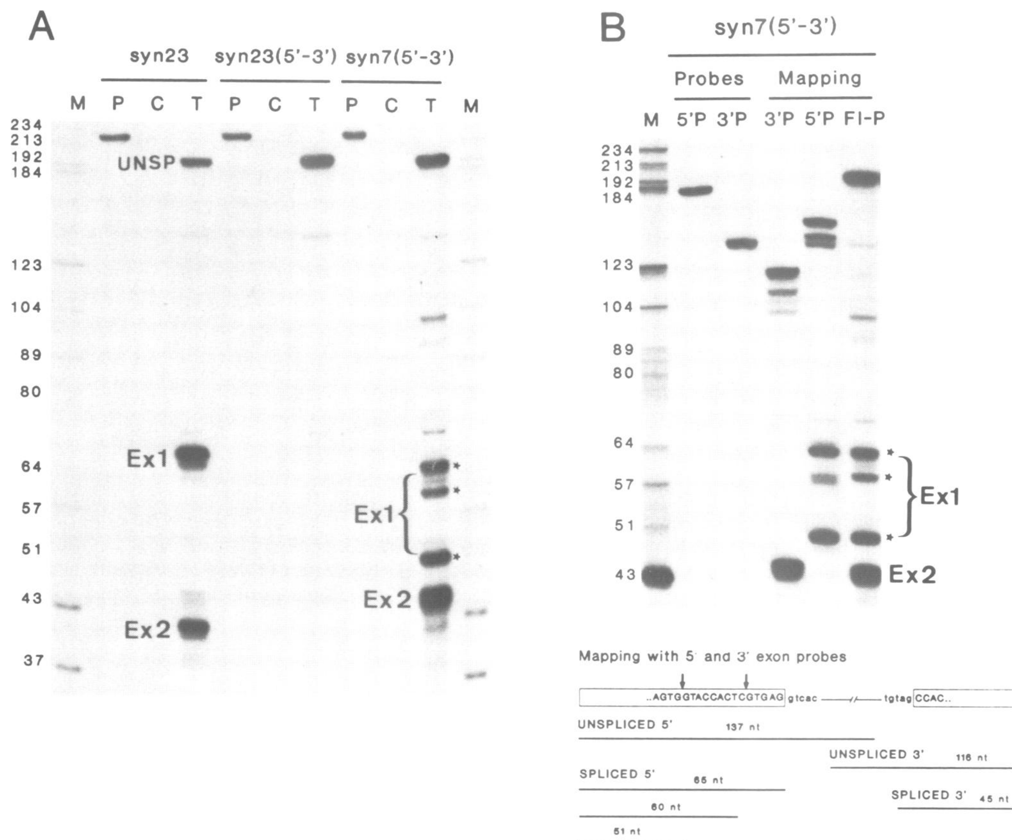

Fig. 4. Importance of the splice site consensus sequences for processing of a GC-rich synthetic intron in maize protoplasts. (**A**) RNase mapping of RNAs transcribed from the synthetic genes indicated at the top. Lanes P, undigested probe; lanes C and T, analyses with RNA from control and transfected protoplasts, respectively; lanes M, size markers. Small differences in the lengths of the corresponding exon-specific fragments seen between the 'wild-type' and the 5′−3′ mutant RNAs are due to the sequence differences in the splice site regions. (**B**) RNase mapping of *syn*7(5′−3′) transcript with probes specific for the 5′ and 3′ splice sites (probes 5′P and 3′P, respectively). Mapping with a full-length probe complementary to unspliced RNA (Fl-P) is also shown in the far right lane. Fragments protected by the 5′ and 3′ probes, diagnostic of unspliced and spliced RNAs, are shown schematically in the diagram below; the additional 5′ splice sites used are indicated by arrows. Different 5′ splice site-specific fragments are also marked with asterisks in the gel lanes in (A) and (B).

5 of the 5′ site and positions −4, −3 and 1 of the 3′ site (but not at position −2 of the 5′ site; Table II). In the 25 most GC-rich introns, in 13 cases the splice site, either 5′ or 3′, matches perfectly the consensus AG/GTA(no G)G or TGCAG/G; such adherence is seen only twice in the 27 AU-rich introns (Table II and data not shown).

The differences discussed above led us to suspect that splice sites with close matches to the consensus may indeed constitute crucial signals for recognition of GC-rich introns. We have tested therefore the effect of splice site mutations on the processing of GC-rich and AU-rich introns in maize. The introns were the synthetic introns *syn*23 (41% AU) and *syn*7 (75% AU), IVS1 of the human β-globin gene (55% AU) and IVS9 of the maize *waxy* gene (40% AU). The experiments are shown in Figures 4 and 5, and the results summarized in Table III.

The GC-rich intron *syn*23 has splice sites that match the consensus perfectly and is spliced in maize with 89%

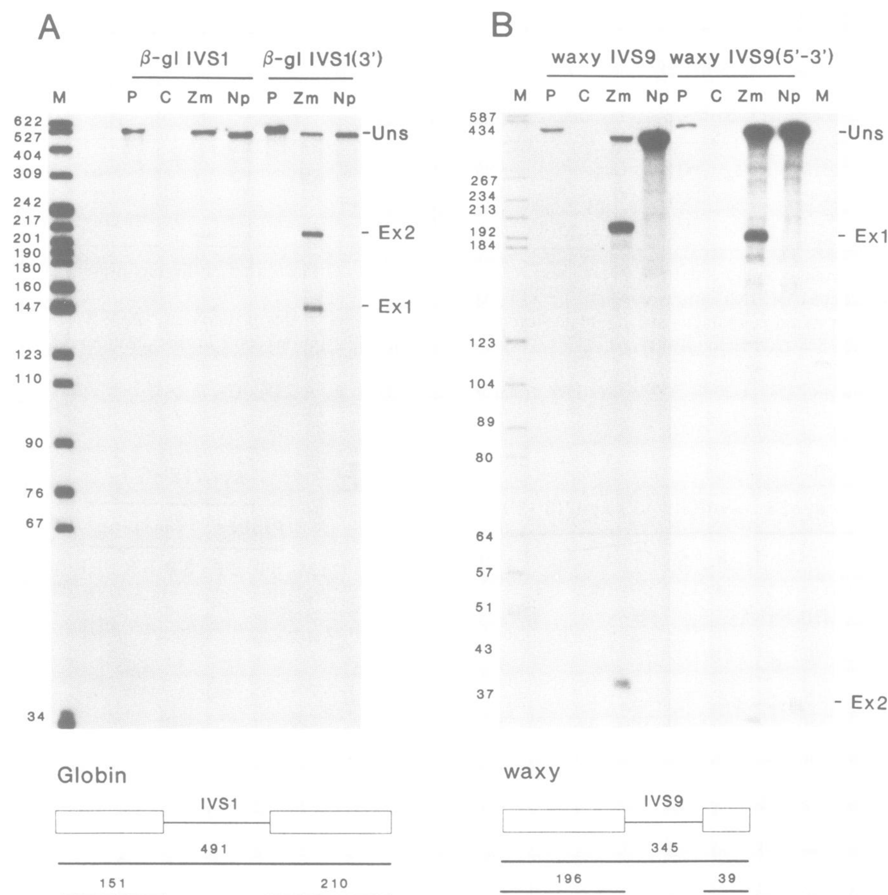

Fig. 5. RNase protection analysis of the splicing of human β-globin IVS1 (**A**) and *waxy* IVS9 (**B**), and their mutants. Plasmids used for transfection, containing the natural or mutated β-globin and *waxy* introns, were pDH51-hβ and pDH51-hβ1(3'), and pDEwx9−10 and pDEwx9(5'−3'), respectively. Lanes P, undigested probe; lanes C, analyses with RNA from mock transfected maize protoplasts; lanes Zm and Np, RNA from transfected maize and *N.plumbaginifolia* protoplasts; lane M, size markers (end-labeled *Hpa*II digest of pBR322 or *Hae*III digest of pBR322 plus 43 nt and 35 nt oligonucleotides). The fragments protected by unspliced and spliced RNAs are diagrammed below.

efficiency (Figure 4A). We modified the splice site regions, leaving the crucial 5'-GT and 3'-AG unchanged; AG/GTCAC replaced AG/GTAAG at the 5' site and TGTAG/C replaced TGCAG/G at the 3' site. Suboptimal splice sites having even less adherence to the consensus are present in some natural monocot introns (data not shown). The changes to the splice sites of *syn*23 inactivated the intron [Figure 4A, *syn*23(5'−3')]. When the same changes were made to the splice sites of the AU-rich intron *syn*7, the efficiency of splicing was not affected, although in addition to the normal 5' splice site, cryptic sites 14 nt and 5 nt upstream of the site AG/GTCAC were also used [Figure 4A, *syn*7(5'−3'); the identity of the cryptic 5' splice sites was confirmed by RNase protection assays using exon 1- and exon 2-specific probes (Figure 4B) and by S1 mapping using oligonucleotide probes complementary to the RNAs spliced at the predicted positions (data not shown)]. In total, 83% of *syn*7(5'−3') transcripts were spliced. We conclude that the presence of optimal splice sites is essential for processing of the GC-rich intron *syn*23 but not of the AU-rich intron *syn*7. Intronic AU-rich sequences may therefore be important for intron recognition in maize when suboptimal splice sites are present.

The IVS1 of the human β-globin gene contains 55% AU, too little to be spliced in dicots (Wiebauer *et al.*, 1988; Figure 5A), but within the range of AU content of monocot

Table III. Effect of splice site mutations on intron processing in maize protoplasts.

Intron	% AU	5' site	3' site	% splicing
*syn*23	41	AG/GTAAG	TGCAG/G	89
*syn*23(5'−3')[a]	41	AG/GTCAC	TGTAG/C	ND[b]
*syn*7	75	AG/GTAAG	TGCAG/G	94
*syn*7(5'−3')[a]	75	AG/GTCAC	TGTAG/C	83[c]
β-globin IVS1	55	AG/GTTGG	CTTAG/G	23
β-globin IVS1(3')	55	AG/GTTGG	CGCAG/G	77
waxy IVS9	40	CG/GTGAG	TTCAG/G	95
waxy IVS9(5'−3')	40	CG/GTTGG	CTTAG/G	19

Nucleotides deviating from the consensus are underlined.
[a]In mutated genes *syn*23 and *syn*7 a T at position +2 of the 3' splice site is, in addition, changed to a C residue.
[b]Splicing not detectable.
[c]Includes splicing at three different 5' sites (see Figure 4).

introns. Indeed, this intron is faithfully processed in maize protoplasts but the efficiency is only 23% (Figure 5A and Table III). The splice sites, AG/GTTGG and CTTAG/G, are suboptimal. We changed the 3' splice site of the β-globin intron from CTTAG/G to the near-optimal CGCAG/G (Table III). This mutation increased splicing efficiency in

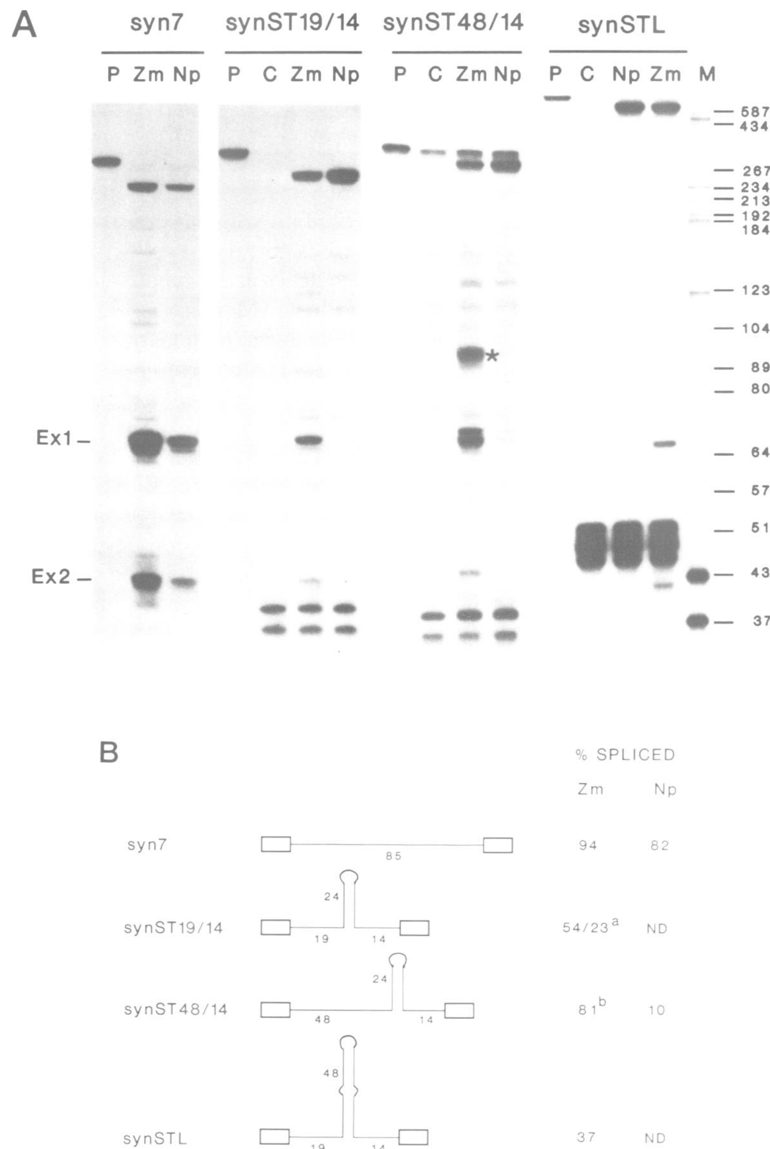

Fig. 6. Processing of synthetic introns containing stem–loop structures. (**A**) RNase protection analysis. Lanes P, undigested probe; lanes C, analyses with RNA from mock transfected maize protoplasts; lanes Zm and Np, RNA from maize and *N.plumbaginifolia* protoplasts. Mapping of *syn*7 RNA is the same as shown in Figure 3A. Exon 2-specific fragment diagnostic of splicing of *syn*ST48/14 RNA to the 'cryptic' 3' site GTTAG/C, located in the stem, is indicated with the asterisk. (**B**) Schematic structures of the stem–loop-containing introns (for details, see Materials and methods). Lengths of the single stranded regions (in nt) and of the base-paired stems (in bp) are indicated. Efficiency of splicing in maize (Zm) and *N.plumbaginifolia* (Np) is indicated.

[a]The two protected exon fragments were not found in equimolar amounts, the exon 1 fragment being in excess.

[b]About 25% of splicing occurred between the usual 5' and 3' splice sites, and 75% between the usual 5' site and a cryptic 3' site GTTAG/C located in the stem. ND, splicing not detectable.

maize to 77% but did not make the intron functional in *N.plumbaginifolia* (Figure 5A and Table III). As a control for this splice site repair experiment, we mutated the splice sites of the *waxy* gene IVS9 so that they are, with the exception of C at position −2 of the 5' site, identical to the splice sites of the β-globin IVS1 (Table III). This change caused a drastic reduction, from 95% to 19%, in splicing efficiency (Figure 5B and Table III). When only the 5' splice site or 3' splice site was modified, an intermediate level of splicing was found (data not shown). These results support the conclusion of the experiments with *syn*7 and *syn*23 introns, that optimal splice sites are essential for processing of GC-rich introns in maize protoplasts.

Short introns with secondary structure are spliced in maize but not in N.plumbaginifolia

The role for AU-rich sequences during splicing of dicot plant pre-mRNAs could either be to minimize secondary structure in introns, or to act as the binding site for one or more splicing factors (Goodall and Filipowicz, 1989). Having found that, unlike dicots, monocots can process GC-rich introns, it was of interest to compare the effect of secondary structure on intron processing in both systems. To this end, we have constructed synthetic introns that contain palindromes capable of forming stem–loops (Figure 6B), and analyzed their processing (Figure 6A).

*Syn*ST19/14 is a derivative of *syn*7 of identical length and

of a similar AU content, but it can potentially form a 24 bp stem with a 4 nt loop within the intron. The *syn*ST19/14 was spliced in maize, although with lower efficiency, but was not spliced at all in *N.plumbaginifolia*. Similarily, the intron *syn*STL, which contains two sets of palindromes capable of forming a 48 bp stem interrupted in the middle by a symmetrical 4 nt bulge, was partially (37%) spliced in maize but not in *N.plumbaginifolia* protoplasts. In the derivative *syn*ST48/14, the distance between the 5′ splice site and the stem−loop was increased to 48 nt. This intron was spliced in maize with much higher efficiency than *syn*ST19/14. About 25% of the *syn*ST48/14 splicing occurred between the usual 5′ and 3′ sites. The remaining 75% was between the usual 5′ site and a cryptic 3′ site positioned within the potential stem, as documented by the presence of the 89 nt protected fragment in the RNase A/T1 digest (Figure 6A, lane *syn*ST48/14-Zm) and by S1 mapping with oligonucleotide probes complementary to spliced RNA (data not shown). In marked contrast to maize, the *syn*ST48/14 intron remained almost inactive in *N.plumbaginifolia* protoplasts (Figure 6). We have found that synthetic introns of 130−150 nt in length, which contain the same stem−loop structure as introns *syn*ST19/14 and *syn*ST48/14, are efficiently processed in both maize and *N.plumbaginifolia* (data not shown). Hence, it is unlikely that the inactivity of introns *syn*ST19/14 or *syn*ST48/14 in *N.plumbaginifolia* protoplasts is due to inhibition of splicing by the sequences present in the stem−loop or due to some unspecific effects of the stem−loop on RNA processing.

The observation that stem−loops present in short synthetic introns completely inhibit splicing in *N.plumbaginifolia* but have only moderate effect in maize suggests that intron secondary structure can be melted, at least partially, or its formation prevented in monocot but not in dicot cells.

Discussion

We have shown previously that the only sequence elements necessary for splicing of synthetic model introns in dicots are the splice sites themselves and a high AU content in the intron (Goodall and Filipowicz, 1989). We offer here further support for this notion by showing that of all the natural plant introns tested in protoplasts of *N.plumbaginifolia*, the only ones that were inactive were the two GC-rich introns of maize. It should be noted that the maize and the wheat intron shown by Keith and Chua (1986) to be spliced with moderate efficiency in tobacco are 62% AU. The introns of the maize transposable elements *Ac* and *Spm*, which are known to produce functional mRNAs and to transpose in dicot plants, are also AU-rich (59−77% AU) (Pereira and Sädler, 1989; Masson *et al.*, 1989; and references therein).

The main finding of this work is the demonstration that the monocot maize can recognize and splice many introns that are spliced poorly or not at all in the dicot *N.plumbaginifolia*, indicating that monocots differ substantially from dicots in their mechanism of intron recognition. It is especially evident that introns that are GC-rich, like the *waxy* IVS9 and 10, or synthetic intron *syn*24, which are not spliced at all in *N.plumbaginifolia*, are spliced in maize. Furthermore, the human β-globin IVS1 (55% AU) and its splice site mutant, which are inactive in *N.plumbaginifolia*, are faithfully processed in maize. This is the first example of a heterologous non-plant intron being accurately processed by a higher plant. Hence, AU-rich sequences are not absolutely essential for intron processing in maize. The fact that all dicot introns are AU-rich, whereas GC-rich introns have been identified in wheat and barley as well as maize, suggests that our results with *N.plumbaginifolia* and maize are representative of dicots and monocots, or at least graminaceous monocots, in general. The ability of monocots to splice both AU-rich and GC-rich introns correlates well with the frequent occurrence of GC-rich genes and isochores in graminaceous monocots but not in dicots (Montero *et al.*, 1990).

However, although not absolutely required, AU-rich sequences do seem to play a role in promoting splicing in monocots. Splicing decreased progressively from 94 to 66% as the AU content of the synthetic intron was lowered from 75 to 27% (Figure 3A). More importantly, we have shown that the presence of AU-rich sequences is essential for processing of introns containing splice sites that strongly deviate from the consensus. This was demonstrated with a synthetic GC-rich intron (*syn*23), the GC-rich IVS9 of the maize *waxy* gene, and a human intron, β-globin IVS1 (Table III). The synthetic GC-rich intron was not spliced when the splice sites were suboptimal, but an AU-rich intron containing the same suboptimal splice sites was efficiently processed, indicating that these splice sites can be functional when AU-rich sequences are present. The conclusion that in the absence of internal AU-rich segments, the introns require better splice sites, presumably to provide strong interactions with the spliceosome components binding to them, is further supported by comparison of the splice site sequences in AU-rich and GC-rich monocot introns (Table II).

Some aspects of intron recognition are similar in monocots and dicots. Using different synthetic introns we have demonstrated that, as in dicots, the 3′-proximal polypyrimidine tract is not essential for splicing in maize protoplasts and that the criteria for branch site selection appear to be quite relaxed. Even when all A residues from intron positions 15 to the 3′-terminal AG were eliminated, the synthetic intron (*syn*36) was faithfully, though less efficiently, processed. A similar result, suggestive of branching to nucleotides other than A (Hornig *et al.*, 1986; Hartmuth and Barta, 1988) was previously obtained with *N.plumbaginifolia* protoplasts (Goodall and Filipowicz, 1989). In support of the above conclusions, analysis of sequences of monocot introns, either AU-rich or GC-rich, did not reveal the 3′ end-proximal enrichment in pyrimidine nucleotides characteristic of vertebrate introns or the presence of sequences resembling the UACUAAC branch point region of yeast introns.

Another feature that distinguishes splicing in monocot and dicot protoplasts is the ability to process introns containing stem−loop structures. Synthetic introns *syn*ST19/14, ST48/14 and STL, in which internal stem−loops would effectively shorten the intron below the ~70 nt minimum required for splicing (Goodall and Filipowicz 1990), are not spliced in *N.plumbaginifolia*, despite being AU-rich. However, the same introns are processed in maize, indicating that monocots, in contrast to dicots, either have a mechanism to inhibit the formation of secondary structure in introns, or are able to some extent to unwind helices in introns prior to splicing. In this respect, splicing in monocots resembles the mammalian reaction *in vivo*. In mammalian cells,

constitutive splicing patterns are not strongly affected by palindromes that could potentially sequester the 5′ or 3′ splice site in the secondary structure (Solnick and Lee, 1987; Eperon *et al.*, 1988); this could be due to hnRNP proteins binding to nascent RNAs and preventing formation of the secondary structure (Eperon *et al.*, 1988), or due to the presence of RNA helicases (reviewed by Wassarman and Steitz, 1991). However, in the mammalian system, secondary structure may have a strong effect on RNA splicing *in vitro* or on the choice of alternative splice sites *in vivo* (Solnick, 1986; Solnick and Lee, 1987; Fu and Manley, 1987; Eperon *et al.*, 1988). On the other hand, the inhibitory effect of stem−loops on splicing in *N.plumbaginifolia* is reminiscent of the yeast *S.cerevisiae* in which sequestration of the splicing signals by complementary sequences has a strong effect on splicing *in vivo* (Yoshimatsu and Nagawa, 1989). We have recently constructed synthetic introns with the 5′ splice site located in a stem. Splicing of these introns is also much more strongly affected in *N.plumbaginifolia* protoplasts than in maize (our unpublished results).

We have suggested (Goodall and Filipowicz, 1989) that the role of AU-rich sequences during splicing of dicot plant pre-mRNAs could either be to minimize secondary structure in introns, or to act as the binding sites for hnRNP proteins or other protein factors required for splicing. The observation that introns containing stem−loops are inactive in protoplasts of *N.plumbaginifolia*, despite being AU-rich, offers some support for the former possibility. Similarly, it is possible that the ability of maize protoplasts to process GC-rich introns, which have the potential to form more stable secondary structures, is related to the ability of monocots to prevent the formation of stems or to unwind them prior to splicing.

In conclusion, splicing in maize appears to be more 'permissive' than splicing in dicot plants or in vertebrates and yeast. The maize splicing machinery is able to process faithfully introns that do not contain signals known to be absolutely essential for intron processing in other organisms. The relaxed requirements of splicing in monocots provide a plausible explanation for the observations that different transposable element or retro-element sequences, when expressed as part of pre-mRNA transcripts in maize, are frequently excised as introns (reviewed by Weil and Wessler, 1990).

Materials and methods

Plasmids for transient expression

The plasmids pDH51-Lb, carrying the soybean leghemoglobin c3 gene, and pDH51-hβ, carrying the human β-globin gene, have been described (Wiebauer *et al.*, 1988). The plasmid pDEmac, bearing part of the maize actin gene, including introns 1 and 2, was constructed as follows: a 960 bp *Hin*dIII fragment from the maize actin gene (Shah *et al.*, 1983) was inserted into the *Hin*dIII site of the vector pGEM-2 (Promega), generating pSPmac. A *Sma*I−*Xho*II fragment from this plasmid, containing pGEM-2 and actin sequences, was then inserted between the *Sma*I and *Bam*HI sites of the expression vector pDELb (Goodall and Filipowicz, 1989). The transcript expressed from this plasmid contains maize actin sequences flanked upstream by 42 nt of bacterial sequences and downstream by leghemoglobin and cauliflower mosaic virus sequences. pDEwx9−10, bearing introns 9 and 10 and flanking exon sequences from the maize *waxy* gene was constructed by inserting a 752 bp *Bam*HI−*Pvu*II fragment from the plasmid pwx11 (Klösgen *et al.*, 1986) between the *Pst*I and *Bam*HI sites of the expression vector pDEDH (Goodall and Filipowicz, 1989). To generate the plasmid pDEphas, an *Eco*RI−*Bam*HI fragment from the plasmid AG-pPVPh3.8 (Slightom *et al.*, 1983), containing the entire phaseolin coding region except for 111 bp of exon 1, was inserted between the *Sma*I and *Bam*HI sites of the polylinker of pDEDH.

The plasmids containing synthetic genes *syn*7, *syn*13, *syn*17, *syn*23, *syn*24, *syn*35 and *syn*36 have been described (Goodall and Filipowicz, 1989). The gene *syn*46 is a derivative of *syn*17 in which the 3′-terminal intron sequence GGTTTTATGATATCATGCAG was replaced by GGAAAGAAGAAA-GAATGCAG. Synthetic genes with splice site mutations, *syn*7(5′−3′) and *syn*23(5′−3′), were obtained by replacement of *Kpn*I−*Cla*I (5′ site region) and *Eco*RV−*Xho*I (3′ site region) fragments of *syn*7 and *syn*23 with appropriate synthetic DNA fragments (for sequences of the resulting mutants, see Table III). Plasmids containing synthetic genes with intronic stem−loops (*syn*ST series) were constructed by replacing, in one or several steps (Goodall and Filipowicz, 1989), the region between the *Cla*I and *Eco*RI sites of *syn*7 with synthetic oligonucleotides. In *syn*ST19/14 the 24 nt sequence AATT-ATTTATTTCTTCTTTTCCAA (intron position 20−43) of *syn*7 is replaced with the sequence AAAATTATGGAAAATGTTAGCCAA which is complementary to intron nucleotides at position 48−71. The inverted repeats have a potential to form a 24 bp stem with 4 nt (TTTT) loop. *Syn*ST48/14 is a derivative of *syn*ST19/14 having the 29 nt sequence CGAAAAGAA-GAAATAAATAATTTAAATAT inserted at intron position 10/11. *Syn*STL is a derivative of *syn*ST19/14 having a 56 nt sequence GGCATCCAGGG-TAGACTAGTAAAGGGCCCTTTACTAGTCTACCCTGGATGCCTT-TT, containing 24 nt inverted repeats, inserted at intron position 47/48. Structures of *syn*ST and *syn*STL introns are schematically shown in Figure 6.

Point mutations in the splice sites of IVS1 of the human β-globin gene [to generate the pDH51-hβ1(3′) derivative of pDH51-hβ; see above] and of IVS9 of the *waxy* gene [to generate the pDEwx9(5′−3′) derivative of pDEwx9−10] were introduced by performing PCR amplifications (Marshallsay *et al.*, 1990) using primers that bore the desired mutant sequence and extended to the neighboring restriction sites. The original sequences were then replaced with the amplification products as described (Marshallsay *et al.*, 1990). The identity of these and all other constructs used in this work was confirmed by sequence analysis.

Transient expression in protoplasts and splicing analysis

Protoplasts from maize and *N.plumbaginifolia* leaves were transfected and incubated at 26°C for 24 h as described (Goodall and Filipowicz, 1989; Goodall *et al.*, 1990). Probes for RNase A/T1 protection analysis were SP6 RNA polymerase transcripts, labelled with [α-^{32}P]CTP or GTP (400 Ci/mmol), from pGEM-derived plasmids. The plasmids for soybean leghemoglobin and human β-globin probes (Wiebauer *et al.*, 1988) and for the synthetic genes *syn*7−*syn*46 (Goodall and Filipowicz, 1989) have been described. pGEM plasmids containing fragments of maize actin and *waxy* genes, bean phaseolin gene and the synthetic genes of the *syn*ST series were prepared by similar procedures. Plasmid used for the synthesis of the exon 1-specific *syn*7(5′−3′) probe was obtained by deletion of the downstream portion of the synthetic gene, the *Eco*RV−*Pst*I fragment, from the pGEM derrivative of *syn*7(5′−3′); for preparation of the exon 2-specific probe, the pGEM plasmid containing the *syn*7(5′−3′) insert was linearized with *Cla*I (see Goodall and Filipowicz, 1989). Gene-specific and mutant-specific probes were always used and in each case the probe was complementary to the unspliced form of RNA. All probes contained additional vector sequences. Probe synthesis, RNA isolation from protoplasts, RNAse A/T1 analysis and quantification of splicing with the Ambis radioactivity scanner were as described (Goodall and Filipowicz, 1989; Goodall *et al.*, 1990). For calculating the efficiency of splicing, correction was made for the different amounts of label in each probe fragment. The values given are usually the averages from at least three transfections. The range obtained for each test sequence was typically $<5\%$ of the mean. Unless indicated otherwise, the ratio of fragments specific for upstream and downstream exons was close to 1:1, as expected.

Nucleotide frequency compilations

Analyses of intron length distribution and of the nucleotide frequencies upstream and downstream of intron borders are based on 280 dicot and 146 monocot introns available in the EMBL (release 22) and GenBank (release 62) data libraries. The same collection of introns was used for derivation of the 5′ and 3′ splice site consensus sequences (see also Goodall *et al.*, 1991).

Acknowledgements

We thank J.Petruska and C.Kellenberger for excellent assistance and Drs B.Hohn, S.Connelly and R.Kole for critical reading of the manuscript.

References

Barta,A., Sommergruber,K., Thompson,D., Hartmuth,K., Matzke,M.A. and Matzke,A.J.M. (1986) *Plant Mol. Biol.*, **6**, 347–357.
Black,D.L., Chabot,T. and Steitz,J.A. (1985) *Cell*, **42**, 737–750.
Brisson,N. and Verma,D.P.S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4055–4059.
Brown,J.W.S. (1986) *Nucleic Acids Res.*, **14**, 9549–9559.
Brown,J.W.S., Feix,G. and Frendewey,D. (1986) *EMBO J.*, **5**, 2749–2758.
Csank,C., Taylor,F.M. and Martindale,D.W. (1990) Nucleic Acids Res., **18**, 5133–5141.
Eperon,L.P., Graham,I.R., Griffiths,A.D. and Eperon,I.C. (1988) *Cell*, **54**, 393–401.
Fu,X.-Y., and Manley,J.L. (1987) *Mol. Cell. Biol.*, **7**, 738–748.
Gerke,V., and Steitz,J.A. (1986) *Cell*, **47**, 973–984.
Goodall,G.J. and Filipowicz,W. (1989) *Cell*, **58**, 473–483.
Goodall,G.J. and Filipowicz,W. (1990) *Plant Mol. Biol.*, **14**, 727–733.
Goodall,G.J., Kiss,T. and Filipowicz,W. (1991) *Oxford Surveys Plant Mol. Cell Biol.*, **7**, in press.
Goodall,G.J., Wiebauer,K. and Filipowicz,W. (1990) *Methods Enzymol.*, **181**, 148–161.
Green,M.R. (1986) *Annu. Rev. Genet.*, **20**, 671–708.
Guthrie,C. and Patterson,B. (1988) *Annu. Rev. Genet.*, **22**, 387- 419.
Hartmuth,K., and Barta,A. (1986) *Nucleic Acids Res.*, **14**, 7513- 7528.
Hartmuth,K. and Barta,A. (1988) *Mol. Cell. Biol.*, **8**, 2011–2020.
Hornig,H., Aebi,M. and Weissmann,C. (1986) *Nature*, **324**, 589–593.
Keith,B. and Chua,N.-H. (1986) *EMBO J.*, **5**, 2419–2425.
Klösgen,W.B., Gierl,A., Schwarz-Sommer,S. and Sädler,H. (1986) *Mol. Gen. Genet.*, **203**, 237–244.
Luehrsen,K.R. and Walbot,V. (1991) *Mol. Gen. Genet.*, **225**, 81–93.
Marshallsay,C., Kiss,T. and Filipowicz,W. (1990) *Nucleic Acids Res.*, **18**, 3451–3458.
Masson,P., Rutherford,G., Banks,J.A. and Fedoroff,N. (1989). *Cell*, **58**, 755–765.
Mattaj,I.W. (1990) *Curr. Opin. Cell Biol.*, **2**, 528–538.
Montero,L.M., Salinas,J., Matassi,G. and Bernardi,G. (1990) *Nucleic Acids Res.*, **18**, 1859–1867.
Nash,J., Luehrsen,K,R. and Walbot,V. (1990) *Plant Cell*, **2**, 1039- 1049.
Parker,R., Siliciano,P.G., and Guthrie,C. (1987) *Cell*, **49**, 229–239.
Pautot,V., Brzezinski,R. and Tepfer,M.(1989) *Gene*, **77**, 133–140.
Pereira,A. and Sädler,H. (1989) *EMBO J.*, **8**, 1315–1321.
Ruskin,B., Zamore,P.D. and Green,M.R. (1988) *Cell*, **52**, 207–219.
Shah,D.M., Hightower,R.C. and Meagher,R.B. (1983) *J. Mol. Appl. Genet.*, **2**, 111–126.
Sharp,P.A. (1987) *Science*, **235**, 766–771.
Siliciano,P. and Guthrie, C. (1988) *Genes Dev.*, **2**, 1258–1267.
Slightom,J.L., Sun,S.M. and Hall,T.C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1897–1901.
Solnick,D. (1985) *Cell*, **43**, 667–676.
Solnick,D. and Lee,S.I. (1987) *Mol. Cell. Biol.*, **7**, 3194–3198.
Tazi,J., Alibert,C., Temsamani,J., Reveillaud,I., Cathala,G., Brunel,C. and Jeanteur,P. (1986) *Cell*, **47**, 755–766.
van Santen,V.L. and Spritz,R.A. (1987) *Gene*, **56**, 253–265.
Wassarman,D.A., and Steitz,J.A. (1991) *Nature*, **349**, 463–464.
Weil,C.F., and Wessler,S.R. (1990) *Annu. Rev. Plant Physiol.*, **41**, 527–552.
Whittier, R.F., D.A. Dean, D.A. and Rogers, J.C. (1987). *Nucleic Acids Res.*, **15**, 2515–2535.
Wiebauer,K., Herrero,J.-J. and Filipowicz,W. (1988) *Mol. Cell. Biol.*, **8**, 2042–2051.
Yoshimatsu,T. and Nagawa,F. (1989) *Science*, **244**, 1346–1348.
Zamore,P.D. and Green,M.R. (1991) *EMBO J.*, **10**, 207–214.
Zhuang,Y. and Weiner,A.M. (1986) *Cell*, **46**, 827- 835.

Received on April 29, 1991