Nucleic Acids Research, 1996, Vol. 24, No. 4 619–627

Characterisation of intronic uridine-rich sequence elements acting as possible targets for nuclear proteins during pre-mRNA splicing in *Nicotiana plumbaginifolia*

Marek Gniadkowski[+], Maja Hemmings-Mieszczak, Ulrich Klahre[§], Hong-Xiang Liu and Witold Filipowicz*

Friedrich Miescher Institute, PO Box 2543, CH-4002 Basel, Switzerland

Received October 31, 1995; Revised and Accepted December 20, 1995

ABSTRACT

Introns of nuclear pre-mRNAs in dicotyledonous plants, unlike introns in vertebrates or yeast, are distinctly rich in A+U nucleotides and this feature is essential for their processing. In order to define more precisely sequence elements important for intron recognition in plants, we investigated the effects of short insertions, either U-rich or A-rich, on splicing of synthetic introns in transfected protoplasts of *Nicotiana plumbaginifolia*. It was found that insertions of U-rich (sequence UUUUUAU) but not A-rich (AUAAAAA) segments can activate splicing of a GC-rich synthetic intron, and that U-rich segments, or multimers thereof, can function irrespective of the site of insertion within the intron. Insertions of multiple U-rich segments, either at the same or different locations, generally had an additive, stimulatory effect on splicing. Mutational analysis showed that replacement of one or two U residues in the UUUUUAU sequence with A or C residues had only a small effect on splicing, but replacement with G residues was strongly inhibitory. Proteins that interact with fragments of natural and synthetic pre-mRNAs *in vitro* were identified in nuclear extracts of *N.plumbaginifolia* by UV cross-linking. The profile of cross-linked plant proteins was considerably less complex than that obtained with a HeLa cell nuclear extract. Two major cross-linkable plant proteins had apparent molecular mass of 50 and 54 kDa and showed affinity for oligouridilates present in synGC introns or for poly(U).

INTRODUCTION

Accurate splicing of nuclear pre-mRNAs requires that exon and intron sequences are effectively distinguished from each other and that appropriate pairs of 5′ and 3′ splice sites at the intron/exon junctions are precisely recognised and juxtaposed. Although some of the pre-mRNA *cis*-acting elements essential for accurate intron excision, such as the 5′ splice site (5′ss) or the intron 3′-terminal AG, are similar in all eukaryotes, others are either unique or their contribution differs significantly between different organisms (for reviews see 1–6). One of the characteristic features of introns in vertebrates and insects is the presence of the polypyrimidine tract usually positioned immediately upstream of the 3′ss (1,3). Early during spliceosome assembly the polypyrimidine tract is recognised by the splicing factor U2AF (7,8; and references therein). U2AF is required for the binding of the U2 snRNP to the branch point region the sequence of which is not highly conserved in vertebrate introns (1,3,9,10). Pre-mRNAs in mammals and in *Drosophila* may also contain *cis*-acting elements located in exons. These elements, known as splicing enhancers, play an essential role in splice site selection, during both constitutive and alternative splicing, and are recognised by members of the SR family of proteins or related splicing factors (5,11,12; and references therein).

The distinguishing feature of pre-mRNA introns in the yeast *Saccharomyces cerevisiae* is a highly conserved branch point sequence, UACUAAC, which base-pairs with the U2 snRNA (2,6,13). Moreover, most introns in yeast lack the extensive polypyrimidine tracts between the branch site and the 3′ss. The extended U2 snRNA-branch point base-pairing in yeast appears to compensate for the absence of the downstream auxiliary signal. Indeed, sequences positioned 3′ of the branch site do not significantly contribute to the first step of splicing in yeast (2,14; and references therein). Some yeast introns contain uridine-rich tracts in the 3′ss-proximal region. However, these tracts function during late stages of splicing (15,16).

Requirements for intron recognition in higher plants differ from those in vertebrates and yeast. The 5′ss and 3′ss in plant pre-mRNAs resemble their vertebrate counterparts but plant introns contain neither distinct 3′ss-proximal polypyrimidine tracts nor conserved branch point sequences similar to those found in vertebrate and yeast introns, respectively (reviewed in 4,17,18). Experiments with synthetic model introns have shown that these two signals are indeed

* To whom correspondence should be addressed

Present addresses: [+]Department of Genetics, University of Warsaw, 02-106 Warsaw, Poland and [§]The Rockefeller University, New York, NY 10021, USA

not essential for intron processing in transfected protoplasts of *Nicotiana plumbaginifolia* or maize (19,20). In two cases analysed, branching in plants was found to occur to A residues positioned 31 and 32 nt upstream of the 3′ss (21). A characteristic feature of plant introns is their AU-richness and this property is essential for splicing. In dicot plants studied more extensively, introns are on average ~15% more AU-rich than flanking exon sequences, with U residues usually contributing much more to the AU-richness than the A residues. In only a very few cases does the A+U content in introns fall below 60% and it is never <55% (19–20,22,23; reviewed in 4). Experiments with synthetic and natural introns have shown that the high A+U content is absolutely essential for efficient intron processing in dicot plant cells (19,20). Consistent with this, when heterologous introns were tested for splicing, only those which were AU-rich were processed at significant levels (19,20,22,24–26). Moreover, synthetic or natural non-intron sequences can function as introns as long as they are AU-rich (19,20,27,28). The requirement for AU-rich introns is not confined to dicot plants. Most introns in monocot plants are AU-rich and this property facilitates splicing (20,28). Introns in *Drosophila*, nematodes, ciliates, slime moulds, and some other organisms are also AU-rich (19,22,23,29,30), and in nematodes AU-richness has been shown to be important for both *cis*- and *trans*-splicing (31,32).

It is not known how AU-rich sequences, usually distributed along the whole length of the intron, contribute to intron recognition or splicing efficiency. Since hairpins introduced into introns strongly inhibit splicing in dicot plant cells, the role of AU-richness could be to minimise secondary structure in introns (20,33). A more plausible possibility is that AU-rich segments act as binding sites for heterogenous nuclear RNP (hnRNP) proteins (for recent reviews on mammalian and yeast hnRNP proteins, see 34,35) or other protein factors that help to delineate sequences to be excised as introns. The latter possibility was originally suggested by the observation that insertions of AU-rich segments of 30 nucleotides (nt) or longer restore efficient splicing of an artificial GC-rich intron, suggesting a positive effect of AU-rich segments on splicing rather than its inhibition by GC-rich sequences (19). This notion is further supported by the finding that the 5′ss and 3′ss selected for splicing in tobacco cells are those present at regions showing a transition from AU-rich to GC-rich sequences, while splice sites embedded within the AU-rich intron sequence are not efficiently utilised (36,37). Analysis of the 3′ss-proximal AU-rich region indicated that U residues may contribute more to the definition of the 3′ss than A residues (36).

In this work, in order to define more precisely sequence elements important for intron recognition in plants and the possible mode of their function, we have investigated the effects of short insertions, either U-rich or A-rich, on splicing of synthetic GC-rich introns in protoplasts of *N.plumbaginifolia*. We find that short U-rich segments such as UUUUAU or multimers thereof, but not A-rich sequences, can activate splicing of the GC-rich intron. The U-rich elements can function irrespective of whether they are inserted in the proximity of the 5′ or 3′ss or in the middle of the intron, indicating that they function differently from polypyrimidine tracts of metazoa. We have found that nuclear extracts of *N.plumbaginifolia* contain a limited set of proteins which interact with fragments of natural and synthetic pre-mRNAs *in vitro*. Two major cross-linkable plant proteins of apparent molecular mass of 50 and 54 kDa have affinity for oligouridilates which makes them good candidates for factors involved in intron recognition in plants.

MATERIALS AND METHODS

Plasmid constructions

Unless indicated otherwise, all DNA manipulations described in this and other sections were carried out according to (38). Sequences of all inserts were verified by dideoxy sequencing.

The synGC and synGCm constructs are derivatives of a vector pDELbm, a modified version of pDELb (19). To obtain pDELbm, the T7 promoter-polylinker-SP6 promoter region of pGEM1 was PCR-amplified, using T7- and SP6-promoter-specific primers (Promega), and cloned into pDELb precut with *Sma*I and *Pst*I and treated with T4 polymerase to blunt the *Pst*I end. SynGC intron, a derivative of intron syn24, was first constructed in plasmid pGS24 (19) by replacing its *Kpn*I–*Cla*I and *Nco*I–*Xho*I regions with new synthetic oligonucleotides, yielding pGSgc. The *Xba*I–*Pst*I fragment of pGSgc was then cloned into pDELbm cut with the same enzymes, yielding pSynGC. pSynGCm was obtained by replacing the *Nco*I–*Xho*I fragment of pSynGC with the sequence CATGGCCGC-CCCGCGGACTGCAGGTACGAGCGC (introduced changes are underlined). To obtain synGC or synGCm derivatives containing U- or A-rich islands, short synthetic oligonucleotides (see Fig. 2C) were 5′-end-phosphorylated and cloned into phosphatase-treated pSynGC precut with either *Cla*I, *Mlu*I or *Sac*II. Clones containing single inserts or different combinations of multimeric inserts, in either orientation, were identified by DNA sequencing. Derivatives containing point mutations in the U-island of pMluUm were prepared by cloning appropriate synthetic oligonucleotides into the *Mlu*I site of pSynGCm.

In vitro RNA synthesis

The following templates were used for *in vitro* synthesis of either cold or ^{32}P-labelled RNAs: (i) for syn7 RNA, pGS7 linearised with *Pst*I (19); (ii) syn7/cDNA RNA, pGEM2/cDNA linearised with *Pst*I (39); (iii) syn7/IVS, pGS7c (derivative of pGS7 with removed *Xba*I–*Cla*I fragment) linearised with *Nco*I; (iv) human β-globin (hβ) RNA, phβ-2, linearised with *Bam*HI (22); (v) Leghemoglobin (Leg) RNA, pHbII linearised with *Hin*cII [pHbII is the pGEM1-based plasmid containing the 0.78 kb *Acc*I–*Hin*dIII fragment subcloned from pLb-1 which contains the soybean leghemoglobin c3 gene insert (22)]; (vi) phaseolin (Phas) RNA, pGphas3 linearised with *Hin*cII (pGphas3 is a pGEM2-based plasmid containing the 0.71 kb *Eco*RI–*Xba*I fragment of the French bean phaseolin gene subcloned from pDEphas (20); (vii) Waxy9 RNA, pGEM1-based plasmid containing the 0.75 kb *Bam*HI–*Pvu*II *waxy* gene fragment (20), linearised with *Pst*I; (viii) for synthesis of synGC-specific RNA and its variants containing U-island insertions, pSynGC and its derivatives were linearised with *Xho*I. For additional information about RNA transcripts, see the legend to Figure 4. pSynGC- and pSynGCm-derived plasmids used for preparation of complementary probes were linearised with *Eco*RV.

In vitro transcriptions were performed as described (40). RNAs used for RNase A/T1 mapping and UV cross-linking were labelled with [α-^{32}P]CTP (sp. act. 80 Ci/mmol) and [α-^{32}P]UTP (100–200 Ci/mmol), respectively. All non-radioactive and radiolabelled RNAs were purifed by electrophoresis in 8 M urea/PAGE.

Transient expression and splicing analysis

Transfections of leaf protoplasts of *N.plumbaginifolia*, isolation of protoplast RNA and RNase A/T1 mapping were performed as

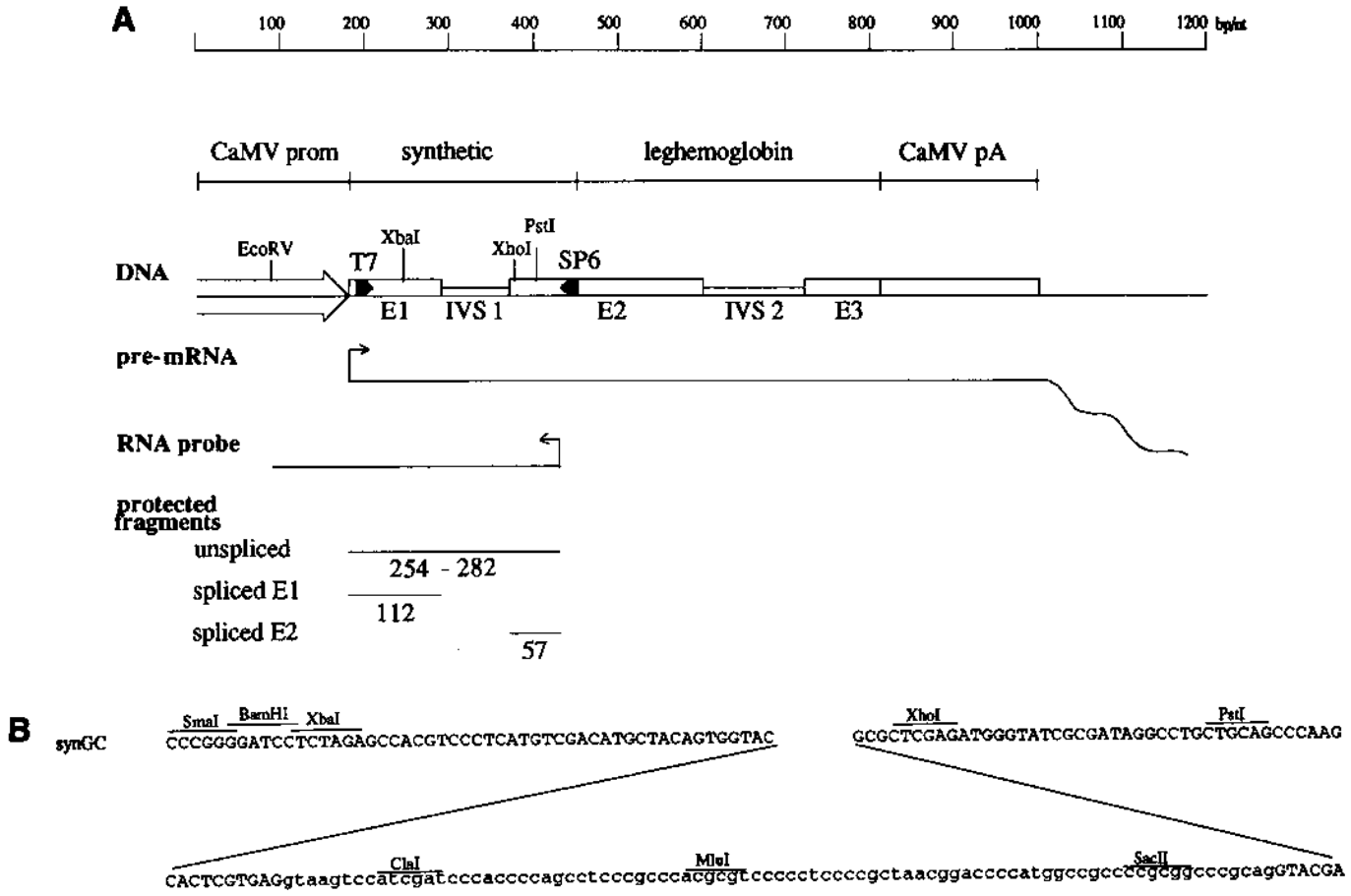

Figure 1. Structure of pSynGC and the RNase A/T1 mapping strategy. (**A**) The Cauliflower Mosaic Virus 35S promoter (CaMV prom), coding, and polyadenylation (CaMV pA) regions of pSynGC. Restriction sites, intron/exon organisation and location of the T7 and SP6 promoters are indicated. The origin of different pSynGC regions is shown above. The pre-mRNA transcript, the anti-sense RNA probe and the sizes of RNA fragments protected during the RNase A/T1 mapping, diagnostic of unspliced RNA and spliced exons 1 (E1) and 2 (E2), are shown below. The poly(A) tail is indicated by a wavy line. (**B**) Sequence of the synthetic synGC intron (lower case) and parts of the flanking exons (upper case). A potential branch point sequence, ctaac, was incorporated 30 nt upstream of the 3′ss. Relevant restriction sites are indicated.

described (20,40). The efficiency of splicing was determined also as described (33). Values given represent the means of at least three independent experiments.

Preparation of plant and HeLa cell extracts

Isolation of plant nuclei. A cell suspension of *N.plumbaginifolia* (obtained from Dr I. Negrutiu, University of Lyon, France) was cultured in NP medium [MS medium (see 40) with altered vitamins: 0.5 mg/ml Ca-pantothenate and 2.5 mg/ml thiamin–HCl] and diluted 1:5 once a week. Protoplasts ($3–4 \times 10^8$) were prepared from 400 ml of cell suspension collected 5 days after subculturing. Pelleted cells (50 ml) were incubated overnight at 28°C in the dark with an equal volume of enzyme solution [5 mM MES, 5 mM $CaCl_2$, 0.47 M sucrose, 1.5% Cellulase Onozuka R10, 0.5% Macerozyme R10 (both from Yacult Honsha Co., Japan) (pH 5.6, 556 mOs)]. Protoplasts were filtered through a 100 µm sieve. Filtrate aliquots (40 ml) were overlayered with 5 ml of W5 solution (40) and centrifuged at 600 r.p.m. for 10 min. Protoplasts, floating and present at the interphase, were collected and washed twice with W5. Protoplasts were homogenised in 30 ml of Buffer H (20 mM MES pH 6.0, 5 mM EDTA, 0.15 mM spermine, 0.5 mM spermidine, 10 mM β-mercaptoethanol, 1 mM PMSF, 1 mg/ml leupeptin, 1 mg/ml antipain) with a few strokes in a Dounce homogeniser (pestle A). The slurry was filtered through 60-, 30/40- and 10/15 µm mesh nylon sieves. The nuclei were pelleted by centrifugation at 1200 *g* for 5 min, taken up in 50 ml of Buffer H and filtered through a 10/15 µm mesh nylon sieve. The sedimentation/filtering step was repeated and the pellet taken up in 20 ml of buffer L (20 mM HEPES pH 7.5, 50 mM KCl, 2 mM $MgCl_2$, 0.1 mM EDTA, 10% glycerol, 2 mM DTT, 1 mM PMSF, 1 mg/ml leupeptin, 1 mg/ml antipain).

Preparation of plant nuclear extracts. Nuclear extracts were prepared by two methods. Except when stated otherwise, the extracts used were prepared according to method 1, resembling that reported by Manley *et al.* (41). Nuclei were resuspended in a Dounce homogeniser and further diluted with Buffer L to 10 A_{260} U/ml. A 1/10 volume of 4.1 M $(NH_4)_2SO_4$ was added dropwise. After 30 min at 0°C and 90 min centrifugation at 100 000 *g*, the supernatant was collected and the protein precipitated by addition of $(NH_4)_2SO_4$ (0.33 g/ml). After centrifugation for 30 min at 20 000 *g*, the pellet was resuspended in a small volume of Buffer D (as Buffer L but without leupeptin and antipain) and dialysed against Buffer D.

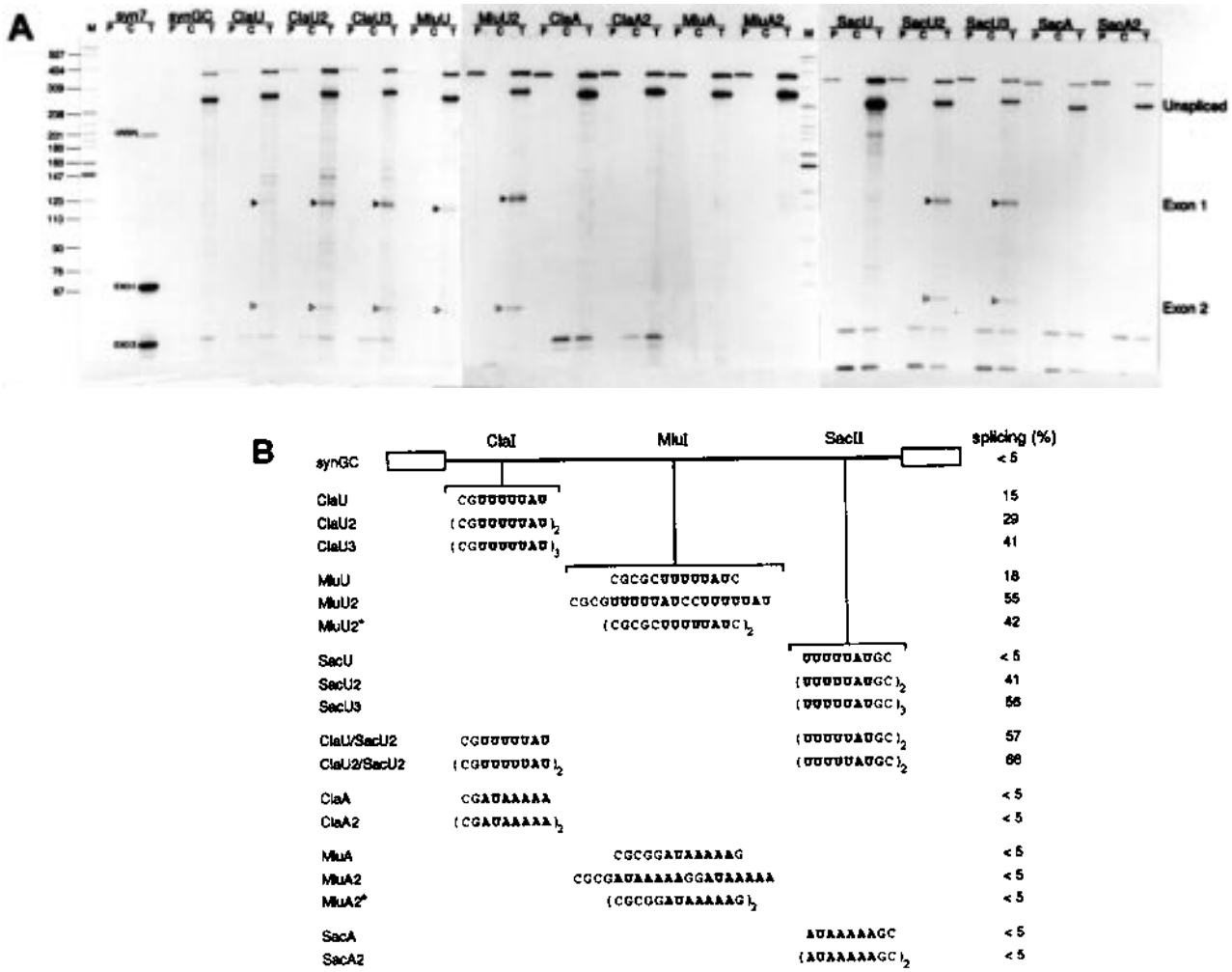

Figure 2. Splicing of synGC introns containing U- or A-rich island insertions. (**A**) RNase A/T1 mapping of RNA isolated from transfected protoplasts. Constructs used for transfections are indicated at the top and are schematically shown in (**B**). Lanes P, aliquot of undigested probe used for mapping; lanes T and C, mapping reactions carried out in the presence of RNA isolated from transfected protoplasts and *E.coli* tRNA, respectively. Lanes M, size markers (pBR322 cut with *Hpa*II). Fragments diagnostic of unspliced (UNSPL) and spliced (EX1 and EX2; additionally marked by filled-in and open triangles, respectively) RNAs are indicated. Probes and protected fragments specific for syn7 RNA are as described (19); EXO1 and EXO2, fragments diagnostic of spliced syn7 RNA. The protected fragments corresponding in length to the full-size probe, seen in lanes T, represent read-through transcripts covering the whole plasmid circle. Synthetic introns located proximal to the 3′ end of these long transcripts do not appear to undergo processing (unpublished results, together with G. Simpson). Approximately 145 nt-long protected bands seen in the ClaU, ClaU2 and ClaU3 lanes most probably represent degradation products of unspliced RNA-specific fragments. RT–PCR analysis provided no evidence of alternatively spliced RNAs (data not shown). (**B**) Schematic representation of constructs used for protoplast transfections. Sequences inserted into *Cla*I, *Mlu*I and *Sac*II sites are indicated, with U- and A-rich islands shown in bold. The efficiency of splicing of individual introns, as quantifed by RNase mapping (see A; mappings of some RNAs are not shown), is given on the right. Quantification of splicing is unreliable when the efficiency is <5%.

The usual yield was 1 ml of extract (~5 mg protein). The alternative method (method 2) was similar to that described by Dignam *et al.* (42). The nuclear pellet was resuspended in 3 vol. of Buffer D containing 0.42 M KCl and the suspension stirred gently for 30 min on ice and centrifuged for 30 min at 25 000 *g*. The resulting supernatant was dialysed against Buffer D.

Nuclear extracts from HeLa cells. Splicing nuclear extracts were prepared as described (42).

UV cross-linking

Unless indicated otherwise, binding reactions (10µl) were carried out in Buffer D (see above) and contained 10–20 µg of nuclear extract protein, 5–12 × 10^4 c.p.m. (6–8 fmol) of labelled RNA (heat denatured at 95°C and quenched on ice prior to addition), and 10 ng of yeast tRNA (it has been verified that addition of tRNA has no effect on the pattern of cross-linked proteins). In competition experiments containing 5 × 10^4 c.p.m. of labelled syn7 RNA, tRNA was omitted and an excess (always calculated in moles of nucleotides) of unlabelled RNA or homopolymers (purchased from Sigma) was added prior to addition of the protein. Further details are indicated in the figure legends. Incubation was for 10 min at 0°C. UV irradiation (254 nm, Camag UV handlamp 29230) was performed on ice for 10 min in Eppendorf tubes positioned 10 cm below the UV source. Following irradiation, RNase A was added to a final concentation

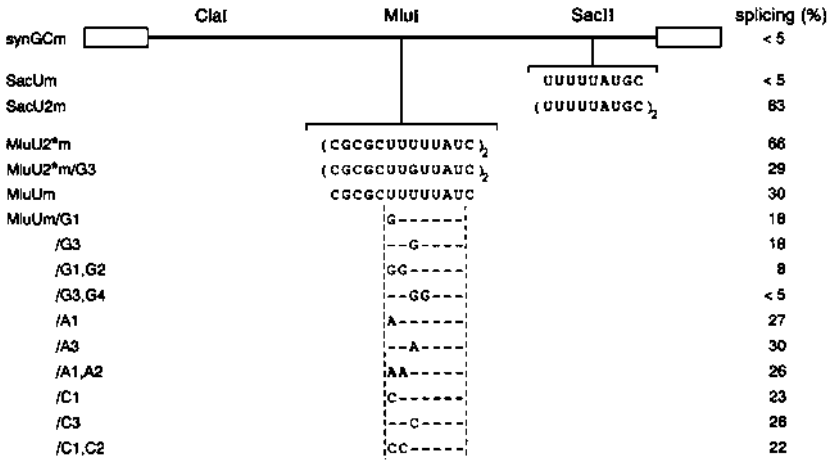

Figure 3. Splicing of synGCm introns containing insertions of the 'wild-type' and mutated U-rich islands. SynGCm is a derivative of the synGC intron having a modified 3′ss. Constructs used for transfections are indicated on the left and the efficiency of their splicing is given on the right. Mutations introduced to the UUUUUAU sequence of MluUm are indicated. Dashes represent unchanged nucleotides. For other details see legend to Figure 2.

of 0.1 mg/ml and samples incubated for 15 min at 30°C. Proteins were fractionated on an SDS–12% polyacrylamide gel.

RESULTS

Insertions of U-rich but not A-rich elements at different locations activate splicing of the GC-rich intron

A model pre-mRNA used to investigate the effect of U- or A-rich sequence elements on splicing is a derivative of the syn7 pre-mRNA (19; see Fig. 1). This pre-mRNA, called synGC, contains a synthetic intron (IVS1) which is 75% GC-rich and is devoid of AU-rich stretches of ≥4 nt. In addition, synthetic exons that flank the intron contain T7 and SP6 promoters to facilitate the *in vitro* synthesis of sense and antisense RNAs. The 5′ss of the synGC intron conforms to the consensus derived for dicot plant introns (AG/GTAAGT; ref. 4), while the 3′ss (CGCAG/GT) deviates from the consensus (TGCAG/GT) at position -5 (Fig. 1B). Plasmids expressing the synGC pre-mRNA or derivatives thereof were transfected into protoplasts of *N.plumbaginifolia.* Splicing efficiency was analysed by RNase A/T1 mapping using RNA isolated from transfected protoplasts and ^{32}P-labelled RNA probes complementary to the unspliced form of RNA. Consistent with previous findings, demonstrating that pre-mRNAs containing GC-rich introns are not spliced in dicot plant cells, the efficiency of synGC intron processing was <5% (Fig. 2). The control syn7 intron which is 75% AU-rich was spliced with an efficiency of 85% (Fig. 2A).

We investigated the effect of short, U-rich (sequence UUUUUAU) or A-rich (sequence AUAAAAA) insertions on splicing of the synGC pre-mRNA. These sequences (referred to as U- or A-rich islands) or multimers thereof, flanked by a few G or C nucleotides in order to facilitate cloning, were inserted either close to the 5′ss or 3′ss or in the middle of the intron (insertions *Cla*I, *Sac*II and *Mlu*I, respectively; see Figs 1B and 2B). Insertions of one, two or three U-rich islands near the 5′ss resulted in splicing of 15, 29 and 41% of the pre-mRNA, respectively. The insertion of a single U-rich island in the middle of the intron resulted in 18% of splicing, while pre-mRNAs with two U-rich islands separated by either two (intron MluU2) or six (intron MluU2*) G+C nucleotides were spliced 55 and 42%, respectively. The presence of one U-rich island at the *Sac*II site, 11 nt upstream of the acceptor AG, did not activate splicing, but pre-mRNAs with two and three islands at this location were efficiently spliced (41 and 56%, respectively; Fig. 2). Insertions of U-rich islands at two different locations had an additive stimulatory effect. Introns containing two U islands at the *Sac*II site and either one (intron ClaU/SacU2) or two (intron ClaU2/SacU2) islands inserted at the *Cla*I site were spliced at levels of 57% and 66%, respectively (Fig. 2, and data not shown), i.e. more efficiently then introns bearing insertions at only one site.

The accuracy of splicing of introns containing U-rich island insertions at different locations was verified by reverse transcription–polymerase chain reaction (RT–PCR) analysis, using exon 1- and exon 2-specific oligonucleotides as primers. In all instances, products with the expected sizes of unspliced and spliced RNAs were identified; no products suggestive of alternatively spliced RNAs were apparent. The product diagnostic of spliced ClaU2/SacU2 RNA was isolated from a gel and cloned. Of 14 clones sequenced, all represented accurately spliced RNA (data not shown).

Although most dicot plant introns contain more U residues than A residues, ~10% have more As than Us (17,19). Therefore, we investigated whether A-rich sequences also have the potential to activate splicing when inserted into the synGC intron. Insertions of one or two A-rich islands at the *Cla*I, *Mlu*I or *Sac*II site had no effect on processing of the synthetic intron (Fig. 2). Taken together, these results demonstrate that insertion of U-rich but not A-rich elements can activate splicing of the GC-rich intron, and that U-rich elements can function irrespective of the position of insertion within the intron.

Effect of mutations in U-rich islands

In order to investigate the sequence or base composition requirements for insertions that activate splicing, we constructed a variant of the synGC intron, called synGCm. The sequence of the 3′ss-proximal nucleotides in synGCm was modified (CCCGCAG/GT changed to ACTGCAG/GT) to conform to the consensus derived for dicot plant introns (see above). The

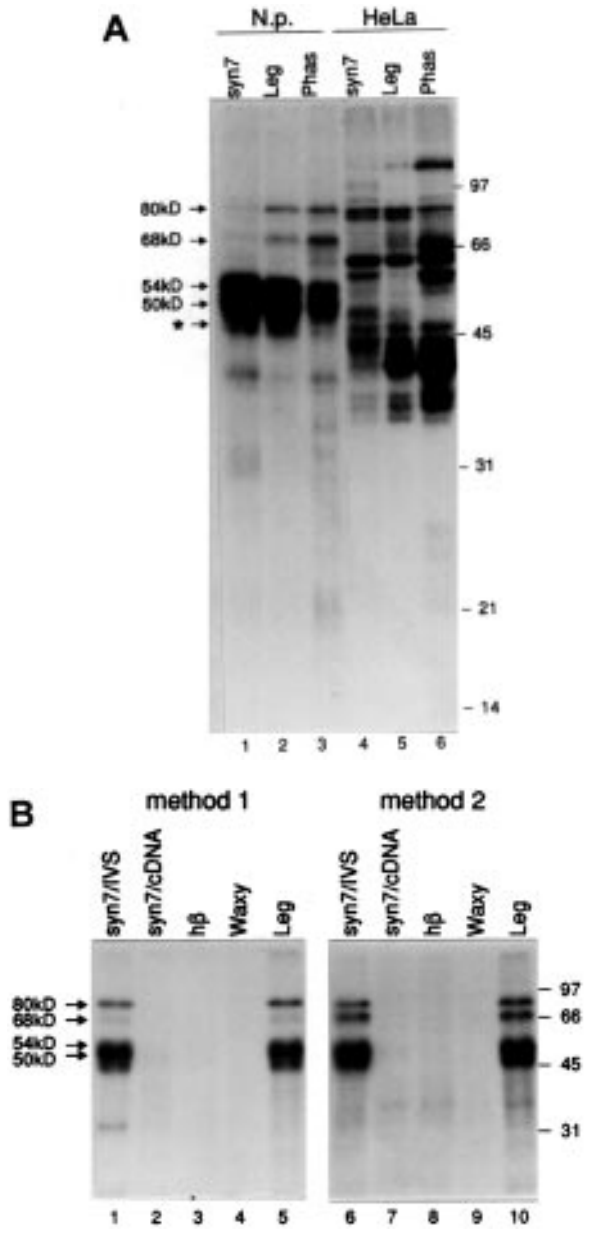

Figure 4. UV cross-linking of different RNAs to nuclear extract proteins. (**A**) Comparison of *N.plumbaginifolia* (lanes 1–3) and HeLa cell (lanes 4–6) nuclear extracts. (**B**) Comparison of the *N.plumbaginifolia* nuclear extracts prepared according to method 1 (lanes 1–5) and method 2 (lanes 6–10). The following ^{32}P-labelled RNAs were used: syn7 (206 nt long, containing a 85 nt-long intron); syn7/cDNA (121 nt; no intron); syn7/IVS, 70 nt long containing a 55 nt long internal *Cla*I–*Nco*I fragment of the syn7 intron; Leg (173 nt, containing a 100 nt long intron); Phas (473 nt; 72 nt intron); hβ (482 nt; 130 nt intron); waxy9 (341 nt, 110 nt intron) (indicated lengths of natural RNAs do not include adjacent polylinker sequences). Labelled RNA (5–12 × 10^5 c.p.m. corresponding to ~8 pmol of each RNA) was used per assay in (**A**) while an equal number of counts (10^5 c.p.m.) was added to each reaction in (**B**). Major cross-linkable proteins identified in plant extracts are indicated by arrows. (*) Intensity of the band marked with an asterisk varied between different extract preparations. Positions of protein size markers (in kDa) are indicated on the right.

synGCm intron was very inefficiently processed, but derivatives of it containing one or two U-rich islands inserted at the *Mlu*I site were spliced 30 and 66%, respectively (Fig. 3), i.e. more efficiently than their synGC-based counterparts (see Fig. 2). As in the case of intron synGC, insertion of a single U-rich island into the *Sac*II site of synGCm did not activate splicing, but insertion of two islands at the same location resulted in very efficient splicing (63%; Fig. 3).

The intron MluUm was subjected to mutagenesis (Fig. 3). Replacement of the first or third U residue in the sequence UUUUUAU with a G reduced splicing efficiency by ~50% [in a construct containing two U-rich islands, a similar mutation in each of them (intron MluU2*m/G3) also decreased activity by 50%; Fig. 3]. Simultaneous mutation of two U residues to G at different positions (introns MluUm/G1,G2 and MluUm/G3,G4) had a more drastic effect and diminished splicing close to the background level. In contrast, replacement of one or two U residues with A had only a negligble effect on the ability of the island to rescue splicing of the synGCm intron (introns MluUm/A1, /A2 and /A1,A2); similar replacements with C residues had an intermediate effect and decreased the stimulatory activity of the island by 25–30% (Fig. 3). These results indicate that efficient activation of splicing requires contiguous uridine stretches, which can be as short as two or three residues provided that intervening or neighbouring nucleotides are not G residues. They also suggest that U-rich sequences interrupted by few A residues may be as effective in rescuing splicing as the prototype UUUUUAU island. We did not directly assay whether insertions of C-rich sequences into synGC or synGCm introns were able to activate processing. However, this is unlikely since these introns already contain multiple stretches of C residues (see Fig. 1B).

Identification of nuclear proteins interacting with U-rich sequences *in vitro*

We and others (18,19,36) have proposed that intronic AU- or U-rich sequences may be recognised by specific proteins that help to delineate intron borders or to attract components of the splicing machinery. To investigate whether plant nuclear extracts contain proteins which can interact with pre-mRNA transcripts, the technique of UV-cross-linking was employed. In the experiment shown in Figure 4A, uniformly [α-^{32}P]UTP-labelled RNAs containing synthetic intron syn7, intron 2 of the soybean leghemoglobin (Leg) gene and intron 1 of the bean phaseolin (Phas) gene were used (in each RNA the intron is flanked by exon sequences; see legend to Fig. 4A). All these introns are AU-rich (71–75% AU) and are efficiently processed in transfected protoplasts of *N.plumbaginifolia* (20). The RNAs were incubated with nuclear extracts prepared either from *N.plumbaginifolia* or HeLa cells. The major cross-linkable proteins identified in plant extracts had apparent molecular mass of 50 and 54 kDa; other less intensely labelled proteins of 68 and 80 kDa were also found to be reproducibly cross-linked to the synthetic and natural plant RNAs (lanes 1–3). The intensity of the ~45 kDa band varied with different extract preparations; as did intensities of other minor bands of ~38 and ~30 kDa (Fig. 4A, see also below). The profile of cross-linked proteins seen with the plant extract was significantly less complex than that seen with the HeLa cell nuclear extract (lanes 4–6). In the latter case, patterns of cross-linked proteins differed considerably between individual RNA substrates (Fig. 4A). This is consistent with previous findings that HeLa cell nuclei contain a large population of hnRNP proteins which interact with pre-mRNAs with evident sequence preferences (reviewed in 34,35,44).

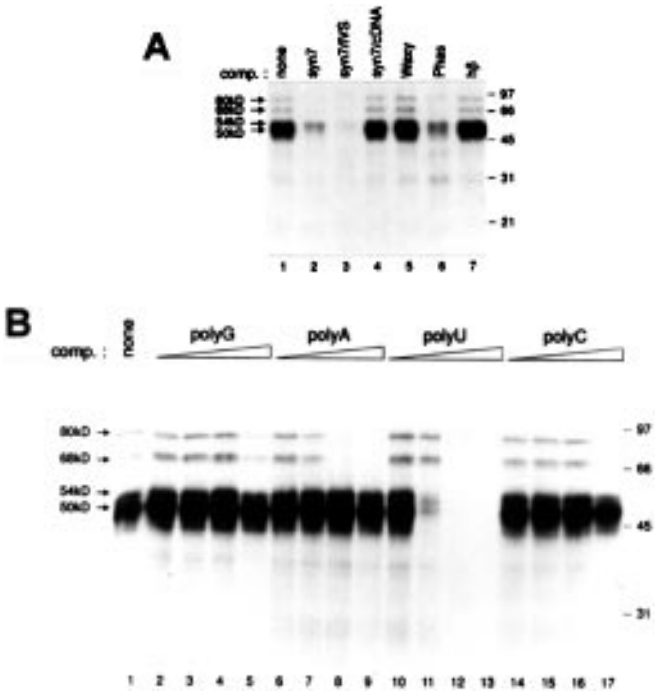

Figure 5. Competition of different RNAs (**A**) and homopolymers (**B**) for cross-linking of the ^{32}P-labelled syn7 RNA to nuclear proteins. Cold RNAs used as competitors are described in the legend to Figure 4. The RNAs were added at 500-fold and homopolymers at 3-, 30-, 300- and 3000-fold excess (calculated in moles of nucleotides) over syn7 RNA. Similar results were obtained when different competitor RNAs were added at 500 molar excess over syn7 RNA (data not shown).

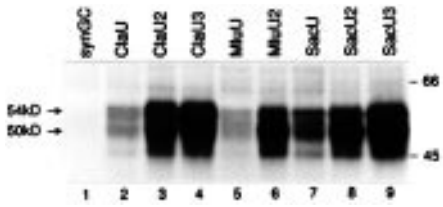

Figure 6. Cross-linking of the U-island-containing synGC RNAs to nuclear proteins. Each sample contained 6 fmol (~5 × 10^4 c.p.m.) of [α-^{32}P]UTP-labelled RNA transcribed *in vitro* from the plasmids indicated.

The experiment presented in Figure 4B is similar to that shown in Figure 4A but compares two types of extracts prepared from nuclei of *N.plumbaginifolia* (see legend to Figure 4B, and Materials and Methods), and four additional RNA substrates were tested. Syn7/cDNA corresponds to a cDNA version of syn7, devoid of the intron, while syn7/IVS is an internal fragment of the syn7 intron. Waxy contains intron 9 of the maize *waxy* gene and hβ contains intron 1 of the human β-globin gene. The latter two introns, being GC-rich (60 and 55% GC, respectively), are not processed in *N.plumbaginifolia* protoplasts (20). The results show that both methods of extract preparation yield comparable patterns of cross-linked proteins. As in the experiment shown in Figure 4A, Leg RNA acted as an efficient substrate but Waxy and hβ transcripts, containing non-functional introns, did not become cross-linked to *N.plumbaginifolia* proteins. Comparison of syn7/IVS and syn7/cDNA RNAs revealed that it is intron and not exon sequences of syn7 which become cross-linked to the proteins (Fig. 4B, compare lanes 1 and 2, and lanes 6 and 7). Additional experiments showed no significant differences in cross-linking of the syn7 intron internal fragments and intron fragments which extend into exons and encompass either the 5′ss or the 3′ss (data not shown). Thus, neither splice site seems to be essential for binding of proteins identified by the cross-linking assay.

Competition experiments were performed to eliminate the possibility that the lack of protein labeling with the syn7/cDNA-specific RNA, and Waxy and hβ RNAs, is not due to their inability to interact with the proteins but rather reflects a lower content of U residues in these RNAs, making the cross-linking inefficient. Addition of an excess of unlabelled syn7, syn7/IVS or Phas RNA efficiently competed for binding of the ^{32}P-labelled syn7 RNA to plant proteins; no competition was observed with syn7/cDNA, Waxy or hβ RNAs (Fig. 5A).

Nucleotide binding specificity of the cross-linked plant proteins was analysed in reactions carried out in the presence of increasing concentrations of different homopolymers. Binding of the 50 and 54 kDa proteins to the labelled syn7 RNA was efficiently competed by poly(U) but not by poly(A), poly(G) or poly(C). Cross-linking of the 68 and 80 kDa proteins was competed equally well by poly(U) and poly(A) but at concentrations much higher than that required for inhibition of the 50 and 54 kDa protein binding by poly(U). At still higher polymer concentrations, cross-linking of the 68 and 80 kDa proteins was inhibited also by poly(C) and poly(G) (Fig. 5B).

Consistent with the demonstration that the 50 and 54 kDa proteins have a high affinity for poly(U), these two proteins could be cross-linked to RNA fragments representing the synGC introns with U-rich islands inserted at different locations (Fig. 6). The 50 and 54 kDa proteins were labelled more strongly when incubated with RNAs containing multiple U-rich islands than with RNAs containing single islands. It should be noted that the RNA containing the SacU intron, which did not undergo processing *in vivo* (see Figs 2 and 3), interacted with both the 54 and 50 kDa proteins *in vitro* (Fig. 6, lane 7). SynGC RNA (Fig. 6, lane 1) and its derivatives containing one or two A-rich islands at different locations (data not shown) were not cross-linked to 50 and 54 kDa proteins; these RNAs also did not compete with cross-linking of these proteins to the labelled syn7 RNA (data not shown).

DISCUSSION

The finding that U- but not A-rich islands can trigger splicing of the inactive synGC intron is consistent with the observation that in natural plant introns U residues generally contribute more to the A+U nucleotide bias than A residues. On average, dicot plant introns are ~41% U and 30% A, and runs of U residues, either uninterrupted or interspersed with few A residues, are a common feature (reviewed in 4,17,18,45). These U-rich stretches are usually found randomly distributed throughout the intron sequence (but see below). Our results extend the findings of Lou *et al.* (36) who studied properties of the 3′ss-proximal AU-rich elements in the maize *Adh*1 gene intron expressed in tobacco cells. Replacement of uridines in these elements by A residues decreased utilisation of the adjacent 3′ss and activated a cryptic 3′ss located upstream in the intron. Inactivation of two AU-rich elements, by either U to A or U to C mutations, had more effect than did mutation of a single element, suggesting that individual elements function cooperatively, which also appears to be the

case for the U-rich islands studied in this work (Figs 2 and 3). In addition, Carle-Urioste *et al*. (46) have recently identified an intra-intronic U-rich element, the integrity of which may be important for efficient splicing of the *bronze* gene intron in maize cells.

Using another set of synthetic introns we have shown previously that replacement of a central region in the AU-rich (50% U and 25% A) syn7 intron with an A-rich sequence still allows efficient splicing. This led to the conclusion that A-rich sequences may be as efficient as U-rich sequences in activating splicing in *N.plumbaginifolia* protoplasts (19). In view of the results presented here, indicating that U-rich stretches interspersed with a few A residues (e.g., sequences AAUUUAU or UUAUUAU; see Fig. 3) are as effective in promoting splicing as the prototype U-rich island (UUUUUAU), it is likely that segments composed of Us and As, still present in the synthetic A-rich intron, were responsible for its efficient splicing.

One of the most important findings of this work is that U-rich sequences can activate splicing irrespective of whether they are inserted near the 5′ss or 3′ss or in the middle of the synGC intron (Fig. 2). This indicates that U-rich islands in plant introns function differently from the metazoan polypyrimidine tracts which are usually located proximal to the 3′ss and which are always downstream of the branch point region (1–3). At an early step of spliceosome assembly, the polypyrimidine tract is recognised by the splicing factor U2AF, which facilitates binding of the U2 snRNP to the branch point region located upstream (3,5,7,9). Other proteins able to interact with the polypyrimidine tract have also been characterised. They may participate in later stages of splicing or play a regulatory role (5,8,47, and references therein). All available evidence indicates that U-islands in introns of the ClaU and MluU series activate splicing from positions upstream of the branch point. The U-islands in the ClaU and MluU introns are separated from the 3′ss by a minimum of 76 and 52 nt, respectively. We have recently demonstrated that A residues selected for branching during splicing *in vivo* of intron 3 of the *Arabidopsis* rubisco activase gene and of the intron syn7 are positioned 32 and 31 nt upstream of the 3′ss, respectively (21). It is most probable that A_{-31} is also selected for branching in the U-island-containing derivatives of synGC since these introns share many structural features with syn7 including a branch point consensus of identical sequence and location (CTAAC, positions –34/–30; see Fig. 1B, and ref. 19). Requirements for a minimal functional distance between the 5′ss and the branch point also argue in favour of branching in ClaU and MluU introns taking place downstream of U-islands. The 5′ss and the branch point have to be separated by 45–50 nt in order to allow productive interaction of U1 and U2 snRNPs with vertebrate pre-mRNAs, and a similar distance appears to be essential for splicing in plants (33, and refs therein). Branching to sequences upstream of U-islands would require that regions as short as 12 nt (introns of the ClaU series) or 39 nt (introns of the MluU series) are competent to simultaneously bind U1 and U2 snRNPs. Even if the U-islands themselves act as branch acceptors, a possibility we consider unlikely, the distance between the 5′ss and the branch point in the ClaU introns would be too short to allow splicing.

While single U-islands inserted at the *Cla*I and *Mlu*I sites can activate processing, a similar insertion at the *Sac*II site, close to the 3′ss, had no stimulatory effect. On the other hand, insertions of two or three U-islands at the *Sac*II site stimulated splicing very strongly (Figs 2 and 3). Inactivity of a single U-island at the *Sac*II site is rather unexpected since 3′ss-proximal regions (residues –21/–6) of plant introns are on average 6–8% more U-rich than regions further upstream (4). It is unlikely that the inactivity of a single U-island is due to its being too close to the acceptor AG since an identical sequence inserted at positions –27/–21, 9 nt further upstream than in the intron SacU, also did not activate splicing (unpublished result). It is possible that sequence requirements for protein binding in the vicinity of the 3′ss are different from requirements at the 5′-proximal and central parts of the intron. Contrary to the results obtained with nuclear extracts (Fig. 6), different proteins may bind to U-rich sequences present at different intron positions *in vivo*.

Accessibility of U-rich sequences within the intron may also be important for determining their activity. We have found (unpublished results) that the stimulatory effect of U-rich islands on processing of synthetic introns is diminished by insertions of complementary A-rich islands likely to result in the formation of hairpins which sequester U-rich sequences. Formation of secondary structure might explain why some introns do not undergo splicing in protoplasts despite the fact that they contain one or two short stretches of Us [e.g., intron syn19 (19) or intron 1 of the human β-globin gene (20)] which would be expected to act as functional U-rich islands.

UV cross-linking experiments have revealed that extracts prepared from nuclei of *N.plumbaginifolia* contain a relatively limited set of proteins which interact with natural and synthetic pre-mRNAs *in vitro*. The two major cross-linkable plant proteins have apparent molecular mass of 50 and 54 kDa. These two proteins, as well as a few others which are labelled less intensely (e.g., 68 and 80 kDa proteins), appear to interact with pre-mRNAs which undergo splicing in protoplasts of *N.plumbaginifolia* (syn7, Leg and Phas RNAs) but not with RNAs which are not spliced in this system (Waxy and hβ RNAs) or with the intron-less syn7/cDNA RNA (Figs 4 and 5A). These observations, together with the finding that the 50 and 54 kDa proteins have affinity for poly(U) (Fig. 5B) and also interact with oligouridilates present in the U-island-containing derivatives of synGC (Fig. 6), make these proteins good candidates for factors involved in intron recognition in plants. We have recently cloned a cDNA encoding the 50 kDa protein which cross-links to plant introns *in vitro*. The deduced protein sequence contains three RNP-type RNA binding domains (also known as RRM or RBD-CS domains; 34,44) but it does not appear to be an equivalent of the characterised mammalian or yeast hnRNP proteins (34,35). Neither does it represent the plant counterpart of the splicing factor U2AF, for which a cDNA clone has also been isolated from *N.plumbaginifolia* (unpublished results, together with G. Simpson and C. Domon). Since the 50 kDa protein is enriched in the nucleus (our unpublished results), it probably also differs from the 50 kDa cytoplasmic poly(U) binding protein implicated in the elicitor-induced destabilization of mRNA encoding a proline-rich protein in bean (48).

The profile of plant proteins identified by UV cross-linking in this work is considerably less complex than that seen with a HeLa cell nuclear extract (Fig. 4). More than 20 hnRNP proteins which associate with nascent pre-mRNAs in mammalian cell nuclei and which range in size from 34 to 120 kDa, have been characterised (34,35,44,49). The exact functions of most of the hnRNP proteins are not known. Some of them are implicated in constitutive or alternative splicing, while others may participate in other processing reactions in the nucleus or in RNA transport (reviewed

in 34,44). Several hnRNP-like proteins have also been recently characterised in the yeast *S.cerevisiae* (reviewed in 35). Although our results show that only a small number of proteins present in plant nuclear extracts are efficiently cross-linked to pre-mRNA fragments *in vitro*, they do not eliminate the possibility that plant cell nuclei contain additional hnRNP-like proteins which are either less abundant or less readily cross-linkable to the substrates used in this work.

Evidence that the 50 and 54 kDa proteins interact with AU-rich or U-rich intron sequences *in vivo* is still missing, but it is interesting to speculate how such sequences and proteins interacting with them could contribute to the processing of plant introns. U-islands are usually randomly distributed along the intron and proteins binding to them could delineate sequences to be excised as introns. The bound proteins could assist U-snRNPs or other splicing factors in identifying the splice sites at the transition regions between the AU-rich and GC-rich sequences. The 50 and 54 kDa proteins may represent specialised sequence-specific hnRNP proteins which, as already proposed for some mammalian hnRNP proteins (34,35), maintain discrete regions within pre-mRNAs in a conformation suitable for interactions with RNA splicing factors.

ACKNOWLEDGEMENTS

We thank J. Petruska for excellent technical assistance and Drs T. Hohn, P. King and H. Rothnie for critical reading of the manuscript.

REFERENCES

1 Green,M.R. (1991) *Annu. Rev. Cell Biol.*, **7**, 559–599.
2 Rymond,B.C. and Rosbash,M. (1992) In Jones, E.W., Pringle,J.R. and Broach,J.R. (eds) *The Molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression*, Vol II, Cold Spring Harbor Laboratory Press, NY, USA. pp. 143–192.
3 Moore,M.J., Query,C.Q. and Sharp,P.A. (1993) In Gesteland,R. and Atkins,J. (eds). *The RNA World*. Cold Spring Harbor Laboratory Press, NY, USA. pp. 303–358.
4 Filipowicz,W., Gniadkowski,M., Klahre,U. and Liu,H.-X. (1995) In Lamond,A. (ed.) *Pre-mRNA Processing*, R.G. Landes Publishers, Georgetown, TX, USA. pp. 65–77.
5 Kramer,A. (1995) In Lamond,A. (ed) *Pre-mRNA Processing*, R.G. Landes Publishers, Georgetown, TX, USA. pp. 35–64.
6 Madhani,H.D. and Guthrie,C. (1994) *Annu. Rev. Genet.*, **28**, 1–26.
7 Zamore,P.D., Patton,J.G. and Green,M.R. (1992) *Nature*, **355**, 609–614.
8 Singh,R., Valcarcel,J. and Green,M.R. (1995) *Science*, **268**, 1173–1176.
9 Ruskin,B., Zamore,P.D. and Green,M.R. (1988) *Cell*, **52**, 207–219.
10 Nelson,K.K. and Green,M.R. (1989) *Cell*, **42**, 335–342.
11 Lynch,K.W. and Maniatis,T. (1995) *Genes Dev.*, **9**, 284–293.
12 Wang,Z., Hoffmann,H.M. and Grabowski,P.J. (1995) *RNA*, **1**, 21–35.
13 Parker,R. Silciano,P. and Guthrie,C. (1987) *Cell*, **49**, 229–239.
14 Rymond,B.C., Torrey,D.D. and Rosbash,M. (1987) *Genes Dev.*, **1**, 238–246.
15 Petterson,B. and Guthrie,C. (1991) *Cell*, **64**, 181–187.
16 Umen J.G. and Guthrie,C. (1995) *Genes Dev.*, **9**, 855–868.
17 Goodall,G.J., Kiss,T. and Filipowicz,W. (1991) *Oxf. Surv. Plant Mol. Cell Biol.*, **7**, 255–296.
18 Luehrsen,K.R., Taha,S. and Walbot,V. (1994) *Prog. Nucl. Acid Res. Mol. Biol.*, **47**, 149–193.
19 Goodall,G.J. and Filipowicz,W. (1989) *Cell*, **58**, 473–483.
20 Goodall,G.J. and Filipowicz,W. (1991) *EMBO J.*, **10**, 2635–2644.
21 Liu,H.-X. and Filipowicz,W. (1996) *Plant J.*, in press.
22 Wiebauer,K, Herrero,J-J. and Filipowicz,W. (1988) *Mol. Cell. Biol.*, **8**, 2042–2051.
23 Csank,C., Taylor,F. and Martindale,D.W.(1990) *Nucleic Acids Res.*, **18**, 5133–5141.
24 Barta,A., Sommergruber,K., Thompson,D., Matzke,M.A., and Matzke,A.J.M. (1986) *Plant Mol. Biol.*, **6**, 347–357.
25 Hunt, A.G., Mogen, B.D., Chu, N.M. and Chua. N.H. (1991) *Plant Mol. Biol.*, **16**, 375–379.
26 Waigmann,E. and Barta,A. (1992) *Nucleic Acids Res.*, **4**, 20, 75–81.
27 Simpson,C.G. and Brown,J.W.S. (1993) *Plant Mol. Biol.*, **21**, 205–211.
28 Luehrsen,K.R. and Walbot,V. (1994) *Genes Dev.*, **8**, 1117–1130.
29 Blumenthal,T. and Thomas,J. (1988) *Trends Genet.*, **4**, 305–308.
30 Mount,S.M., Burks,C., Hertz,G., Stormo,G.D., White,O., and Fields,C. (1992) *Nucleic Acids Res.*, **20**, 4255–4262.
31 Conrad,R., Liou,R.F. and Blumenthal,T. (1993) *Nucleic Acids Res.*, **21**, 913–919.
32 Conrad,R., Lea,K. and Blumenthal,T. (1995) *RNA*, **1**, 164–170.
33 Liu,H.-X., Goodall,G.J., Kole,R. and Filipowicz,W. (1995) *EMBO J.*, **14**, 377–388.
34 Dreyfuss,G., Matunis,M.J., Pinol-Roma,S. and Burd,C.G. (1993) *Annu. Rev. Biochem.*, **62**, 289–321.
35 Swanson,M.S. (1995) In Lamond,A. (ed) *Pre-mRNA Processing*, R.G. Landes Publishers, Georgetown, TX, USA. pp. 17–33.
36 Lou,H., McCullough,A.J. and Schuler,M.A., (1993) *Mol. Cell. Biol.*, **13**, 4485–4493.
37 McCullough,A.J., Lou,H. and Schuler,M.A. (1993) *Mol. Cell. Biol.*, **13**, 1323–1331.
38 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
39 Connelly,S. and Filipowicz,W. (1993) *Mol. Cell. Biol.*, **14**, 5910–5919.
40 Goodall,G.J., Wiebauer,K. and Filipowicz,W. (1990) *Methods Enzymol.*, **181**, 148–161.
41 Manley,J.L., Fire,A., Cano,A., Sharp,P.A. and Gefter,M.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3855–3859.
42 Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
43 Skukla,R.R., Dominski,Z., Zwierzynski,T. and Kole,R. (1990) *J. Biol. Chem.*, **265**, 20377–20383.
44 Burd,C.G. and Greyfuss,G. (1994) *Science*, **265**, 615–621.
45 Sinibaldi,R.M. and Mettler,I.J. (1992) *Prog. Nucl. Acid Res. Mol. Biol.*, **42**, 229–257.
46 Carle-Urioste,J.C., Ko,C.H., Benito,M.-I. and Walbot,V. (1994) *Plant Mol. Biol.*, **26**, 1785–1795.
47 Gozani,O., Patton,J.G. and Reed,R. (1994) *EMBO J.*, **13**, 3356–3367.
48 Zhang,S. and Mehdy,M.C. (1994) *The Plant Cell*, **6**, 135–145.
49 Pinol-Roma,S., Choi,Y.D., Matunis,M.J. and Dreyfuss,G. (1988) *Genes Dev.*, **2**, 215–227.