A 62,000 molecular weight spliceosome protein crosslinks to the intron polypyrimidine tract

Jin Wang and Thoru Pederson*
Cell Biology Group, Worcester Foundation for Experimental Biology, Shrewsbury, MA 01545, USA

Received July 16, 1990; Revised and Accepted September 21, 1990

ABSTRACT

Incubation in HeLa nuclear extract of a ^{32}P-labeled 61 nucleotide-long RNA corresponding to the lariat branch site/polypyrimidine tract/3′ splice site of the first intron of human β-globin pre-mRNA led to the crosslinking of a single protein of ~62,000 mol. wt. (p62). p62 corresponds to a polypyrimidine tract-binding protein recently described by Garcia-Blanco *et al.* (*Genes & Dev.* 3: 1874–1886, 1989). Crosslinking of p62 to the 61 nt RNA was highly sequence specific. No p62 crosslinking was observed with a 60 nt pGEM vector RNA, a 63 nt RNA antisense to the 61-mer or a 72 nt U2 RNA sequence. p62 crosslinking to the 61 nt RNA was competed by unlabeled 61 nt RNA, by β-globin pre-mRNA containing intron 1, and by poly(U) and poly(C), but was competed to a lesser extent or not at all by pGEM RNA, a β-globin RNA lacking intron 1, or poly(A). Experiments with mutated RNAs revealed that neither the lariat branch site adenosine nor the 3′ splice site were required for p62 crosslinking to polypyrimidine tract-containing RNA. Elimination of the polypyrimidine tract reduced p62 crosslinking, as did mutation of a polypyrimidine tract UU dinucleotide to GA. However, replacement of a pyrimidine-rich tract immediately adjacent (3′) to the lariat branch site with a 57% A+G pGEM vector RNA sequence also significantly reduced p62 crosslinking, indicating the involvement of both this pyrimidine-rich region and the classical polypyrimidine tract adjacent to the 3′ splice site. The sites of protein interaction were further defined by RNase H protection experiments, the results of which confirmed the patterns of p62 crosslinking to mutant RNAs. p62 crosslinking was efficiently competed by a DNA oligonucleotide having the same sequence as the 61 nt RNA, showing that p62 requires neither ribose 2′ OH groups nor uracil bases for its interaction with the polypyrimidine tract. p62 was not crosslinked to double-stranded 61 nt RNA. Q-Sepharose chromatography of HeLa nuclear extract yielded an unbound fraction (QU) in which p62 was the only polypyrimidine tract-crosslinkable protein and a bound fraction (QB) in which, surprisingly, several non-p62 proteins were crosslinkable to the polypyrimidine tract RNA. Yet, when the two Q-Sepharose fractions were combined, p62 strongly out-competed the otherwise-crosslinkable QB proteins for polypyrimidine tract RNA crosslinking. This indicates that p62 may have the highest affinity and/or crosslinking efficiency for the intron polypyrimidine tract of any HeLa nuclear protein.

INTRODUCTION

Pre-mRNA splicing involves numerous protein-RNA interactions. Among the nuclear proteins that have been shown to interact with pre-mRNA are the A1, C-group and D hnRNP proteins (1,2) and at least three other intron-binding proteins (3–6). The highly conserved polypyrimidine tract at the 3′ end of the introns in higher eukaryotes is an especially active site as regards interaction with proteins. The aforementioned hnRNP proteins have been shown to bind to this region, as well as elsewhere along the pre-mRNA (1), and another intron-binding protein of apparent 62,000 mol. wt., binds specifically to the polypyrimidine tract (5). In addition, a splicing factor termed U2AF (for U2 snRNP auxiliary factor), which is required for the binding of U2 snRNP to the lariat branch site (7), contains a ~65,000 mol. wt. protein component that binds to the polypyrimidine tract/3′ splice site (6). In this report, we describe the identification and characterization of p62, a HeLa nuclear protein that specifically crosslinks to the polypyrimidine tract of mammalian pre-mRNA introns. Our results confirm, but also extend, those reported previously (5,6). We have in particular scrutinized the RNA sequence specificity of the p62:polypyrimidine tract interaction.

MATERIALS AND METHODS

Plasmids

Two plasmids, pPY61 and pPY61′, containing the lariat branch site/polypyrimidine tract/3′ splice site of the first intron of the human β-globin gene, were constructed by inserting chemically synthesized DNA fragments into the EcoRI and HindIII sites of pGEM1. (All oligodeoxynucleotides were synthesized on an Applied Biosystems model 8700 DNA synthesizer by Dr. Vipin Kohli.) After EcoRI linearization, SP6 RNA polymerase transcribed from pPY61 a 61 nucleotide RNA containing 12 linker and pGEM-templated nucleotides at its 5′ end, followed

* To whom correspondence should be addressed

by 45 β-globin intron nucleotides corresponding to the lariat branch site/polypyrimidine tract/3′ splice site, and ending with four extra linker-derived nucleotides at its 3′ end (see Figure 9A). Transcription of EcoRI-linearized pPY61′ produced the same 61 nucleotide RNA but containing a UU → GA mutation at nucleotides 22 and 23 after the lariat branch site (see Figure 9A).

pPY53 was derived from pPY61′ by excising a XbaI/EcoRI fragment and reinserting it into XbaI/EcoRI-digested pGEM1. SP6 transcription of EcoRI-cut pPY53 produced a 53 nucleotide RNA in which a 14 nt pGEM1 sequence replaced the lariat branch site region (see Figure 9A). In some experiments, β-globin or adenovirus RNAs were transcribed from the plasmids SP62-HβAΔ44, SP64-HβΔ6, SP64-Hβ-IVS1,2 and pBSAd10 as previously described (2).

Transcription

RNA was transcribed with SP6 or T7 RNA polymerases with [α-^{32}P]-UTP or [α-^{32}P]-CTP at 2 mCi/ml. The specific activity of the transcripts ranged from $2-7\times10^8$ dpm/μg. Unlabeled RNAs were transcribed in the presence of 500 μM ATP, CTP, GTP and UTP.

RNA-protein binding and UV crosslinking

The desired RNA was added to binding mixture containing 60% (vol/vol) HeLa nuclear extract (8) 0.5 mM ATP, 3.2 mM $MgCl_2$ and 20 mM creatine phosphate unless otherwise specified. The final concentration of ^{32}P substrate RNA was typically about 40 ng/ml. After incubation for 30 minutes at 30°C (unless otherwise specified), the samples were placed on ice and irradiated with 254 nm wavelength light (4000 μW/cm^2, 15 minutes, 4°C). The samples were then incubated at 37°C for 60 minutes with RNase A (100 μg/ml) and micrococcal nuclease (800 units/ml). SDS polyacrylamide gel sample buffer was then added, the samples were boiled for 3 minutes and loaded on a 12% polyacrylamide gel. ^{32}P-nucleotide crosslinked proteins were visualized by autoradiography.

Detecting ^{32}P-nucleotide crosslinked proteins in splicing complexes

^{32}P-labeled adenovirus pre-mRNA ($2-5\times10^5$ dpm) was incubated in 100 μL of 30% (vol/vol) HeLa nuclear extract at 30°C for 30 minutes and splicing complexes were resolved by electrophoresis on a non-denaturing 4% polyacrylamide-0.5% agarose gel (polyacrylamide:bisacrylamide = 60) buffered with 50 mM Tris-glycine, pH 8.8 (9). After autoradiography of the wet gel for 16 hours at 4°C, the A and B splicing complexes were excised with a razor blade and irradiated in the gel slice as described above. For other experiments, the adenovirus pre-mRNA was labeled to approximately a 10-fold lower specific activity. After incubation in nuclear extract and electrophoretic separation of splicing complexes as above, the gel slices containing A or B complexes were incubated in 200 μl of 0.5M KCl in buffer D for 10 min. at 4°C. (Buffer D is 20mM HEPES, pH 8.0, 100 mM KCl, 2mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 20% (vol/vol) glycerol.) The sample was then centrifuged for 1 min. at 12,000×g and the supernatant was removed. 100 μl of buffer D was added, followed by $2-4\times10^5$ dpm of ^{32}P-labeled 61 nt RNA. After incubation for an additional 30 min. at 21°C, the sample (gel slice and bathing buffer) was transferred to a small Petri dish for UV irradiation as detailed above. RNase A was then added to 500 μg/ml and micrococcal nuclease to 800 U/ml and the samples were incubated for two hours at 37°C. The gel slices were then transferred to the wells of a 12% polyacrylamide gel and the proteins were electrophoresed out under standard SDS running buffer conditions.

Nuclear extract fractionation

5.0 ml of HeLa nuclear extract (8), equivalent to 1.2×10^9 cells, were adjusted to 300 mM KCl in buffer A (buffer A: 20 mM HEPES, pH 7.9, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM dithiothreitol, 10% (vol/vol) glycerol. The sample was loaded at 4°−8°C onto a Q-Sepharose (Pharmacia) column (1.5×10 cm) pre-equilibrated with buffer A (flow rate 15 ml/hr). The column was washed with 5 bed volumes of 300 mM KCl-buffer A, and the protein-rich fractions were pooled as the unbound fraction ('QU'). The column was then washed with 3 bed volumes of 1M KCl in buffer A and the eluted material was termed the bound fraction ('QB'). The QU and QB fractions were concentrated 3−5 fold by centrifugation through Amicon 10 microconcentrator columns and were then dialyzed against a 200-fold excess of buffer D for 8−12 hours before use.

RNase H protection experiments

^{32}P-labeled 61 nt RNA substrate (10^5 dpm, approx. 2 ng) was incubated in 15 μL of QU fraction at 30°C for 10 min (total reaction volume = 25 μL). The desired oligodeoxynucleotide was then added to a concentration of 40 μg/ml and the incubation was continued for another 30 min. The reactions were then digested with proteinase K (200 μg/ml, 10 min., 37°C) in the presence of 0.5% SDS and extracted once with phenol:chloroform followed by one re-extraction of the aqueous phase with chloroform. After ethanol precipitation, the RNAs were electrophoresed on a 10% polyacrylamide gel containing 7M urea and visualized by autoradiography.

RESULTS

Strategy

We and others have studied the interaction of nuclear proteins with pre-mRNA by RNA-protein crosslinking and immunoselection methods (1−6, 10−12). In the present investigation, we studied the crosslinking of HeLa nuclear proteins to a 61 nucleotide-long RNA that encompasses the lariat branch site and polypyrimidine tract of the first intron of human β-globin pre-mRNA. By eliminating flanking RNA regions, we sought to define the core RNA sequence requirements for protein crosslinking to this region of pre-mRNA in the absence of long-range secondary structure or higher order spliceosomal organization.

A 62,000 mol. wt. nuclear protein crosslinks to the intron polypyrimidine tract

We cloned into pGEM1 a segment of the human β-globin gene's first intron spanning the lariat branch site to the 3′ splice site (see Methods and Materials for details). Transcription of this plasmid with SP6 RNA polymerase produced a 61 nucleotide RNA representing this region of β-globin pre-mRNA, whereas T7 RNA polymerase transcription yielded a 63 nucleotide antisense RNA as a useful control.

Incubation of ^{32}P-labeled 61 nt RNA in a HeLa nuclear extract under splicing-permissive conditions followed by UV irradiation led to crosslinking of the RNA to a 62,000 (apparent) molecular weight protein (Figure 1, lane 1). (For brevity, we shall refer to this protein simply as 'p62', recognizing that its electrophoretic mobility undoubtedly reflects a contribution of

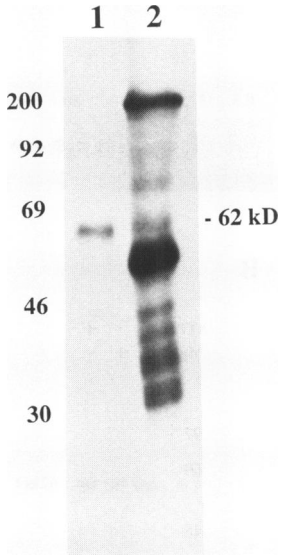

Figure 1. UV crosslinking of nuclear proteins to β-globin pre-mRNA and the 61 nt lariat branch site/polypyrimidine tract/3′ splice site RNA. Equimolar amounts of ^{32}P-labeled 61 nt RNA (lane 1) or β-globin pre-mRNA (lane 2) were incubated in HeLa nuclear extract (30 min., 30°C). After UV crosslinking and nuclease digestion (see Materials and Methods), the ^{32}P-nucleotide-crosslinked proteins were visualized by SDS-polyacrylamide gel electrophoresis and autoradiography. Molecular weight calibration markers indicated at the left of lane 1 (as mol. wt. $\times 10^{-3}$) were: 200, myosin heavy chain; 92, phosphorylase B; 69, bovine serum albumin; 46, ovalbumin; 30, carbonic anhydrase.

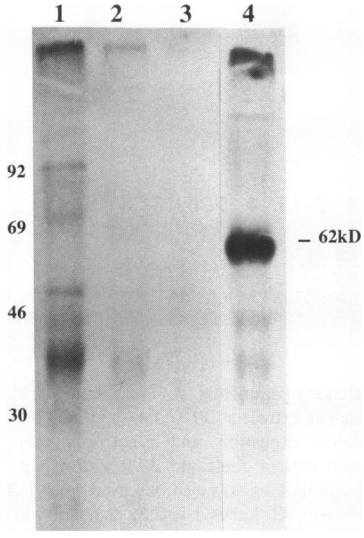

Figure 2. RNA sequence specificity of p62 crosslinking. Equimolar amounts of ^{32}P-labeled RNAs were incubated in nuclear extract, crosslinked and analyzed as in Figure 1. Lane 1: pGEM2 RNA (60 nt); lane 2: RNA antisense to the lariat branch site/polypyrimidine tract/3′ splice site sequence (63 nt); lane 3: a RNA corresponding to nucleotides 1 to 72 of human U2 small nuclear RNA; lane 4: the 61 nt RNA used in Figure 1, lane 1.

crosslinked nucleotides and that its true molecular weight has not been determined.) A considerably more complex pattern of proteins was crosslinked to human β-globin RNA containing the first two exons and the first intron (Figure 1, lane 2; see also ref. 2). However, a minor band at approximately 62,000 mol. wt. was observed in this more complex pattern, which was also

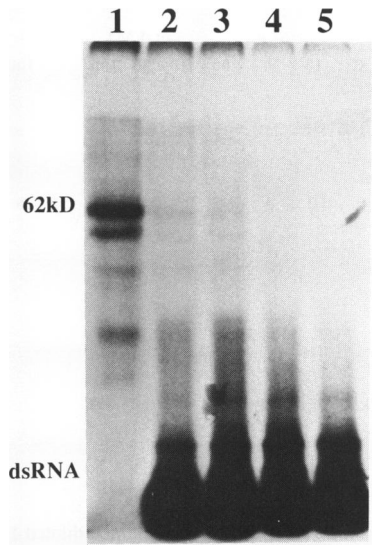

Figure 3. p62 does not crosslink to the intron polypyrimidine tract in double-stranded RNA. ^{32}P-labeled 61 nt RNA was incubated with an excess of unlabeled 63 nt antisense RNA for 20 min. at 30°C in buffer D. Nuclear extract (3μL) was then added and incubation was continued for 30 min., followed by crosslinking, nuclease digestion and electrophoresis. Lane 1: ^{32}P-labeled 61 nt RNA, no complementary RNA; lanes 2–5, ^{32}P-labeled 61 nt RNA pre-annealed with a 2, 4, 6 and 10-fold molar excess, respectively, of unlabeled complementary RNA prior to incubation with nuclear extract.

the case when adenovirus or mouse immunoglobulin pre-mRNAs were used (data not shown).

Crosslinking of p62 to the 61 nt RNA was highly sequence-specific. No crosslinking occurred with a 60 nt pGEM1 RNA transcript, a 63 nt RNA that is antisense to the 61 nt RNA, or a RNA representing nucleotides 1–72 of human U2 RNA (Figure 2). p62 was not crosslinkable to double-stranded 61 nt RNA (Figure 3), indicating that this protein has a preference for the single-stranded sequence embodied in the lariat branch site/polypyrimidine tract.

The specificity of crosslinking was investigated further by performing competition experiments. Crosslinking of p62 to the 61 nt RNA was competed by a 10-fold molar excess of unlabeled 61 nt RNA and more completely by a 30-fold excess or higher (Figure 4A). Some competition was observed with a 10-fold excess of pGEM2 RNA, but no additional competition was seen when this competitor's concentration was raised further (Figure 4A). tRNA displayed a weak level of competition, similar to that of pGEM2 RNA (data not shown). Interestingly, complete competition was observed with a 100-fold excess of a 46 nt oligodeoxynucleotide having the same sequence as the 61 nt RNA except lacking 7 nt and 8 nt at the 5′ and 3′ ends respectively (Figure 4, lane 10), indicating that p62 recognizes this nucleotide sequence in both RNA and DNA. Little or no competition was observed with two control oligodeoxynucleotides (Figure 4A, lanes 8 and 9). p62 crosslinking to the 61 nt RNA was competed more efficiently by a β-globin RNA containing the first intron than by a β-globin RNA lacking this intron (Figure 4B). Crosslinking of p62 was efficiently competed by poly(C) or poly(U), but not by poly(A) (Figure 5).

The ability of p62 to be crosslinked to the 61 nt RNA was detected within 30 seconds in the HeLa nuclear splicing extract (Figure 6A), required neither ATP nor Mg^{++} (Figure 6B) and was only slightly more extensive at 30°C than at 0°C (Figure

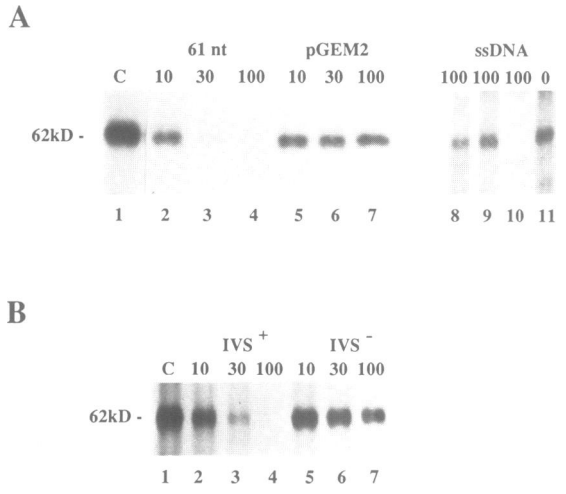

Figure 4. Competition experiments. Nuclear extract, diluted 10-fold with buffer D, was incubated with ^{32}P-labeled 61 nt RNA in the absence or presence of various unlabeled RNAs or oligodeoxynucleotides, followed by crosslinking, nuclease digestion and electrophoresis. *A*. Lane 1: no competitor; lanes 2–4: 10, 30 and 100-fold molar excess of unlabeled 61 nt RNA respectively; lanes 5–7, 10, 30 and 100-fold molar excess of pGEM2 RNA (60 nt); lane 8, 100-fold molar excess of a 20-mer control oligodeoxynucleotide (5′-GATCCAGGTAA-GTATCTGCA-3′); lane 9, 100-fold molar excess of a 37-mer control oligodeoxynucleotide (5′-AGGAACGGTGCACCCATATATAGGATCCTCTA-GAGTC-3′); lane 10, 100-fold molar excess of a 46-mer ologideoxynucleotide having the same sequence as the 61 nt polypyrimidine tract RNA except lacking 7 nt and 8 nt at the 5′ and 3′ ends respectively; lane 11, no oligodeoxynucleotide competitor. *B*. Lane 1: ^{32}P-labeled 61 nt RNA, no competitor; lanes 2–4: 10, 30 and 100-fold molar excess of β-globin pre-mRNA containing exon 1, intron 1 and exon 2 (SP6 transcript of BamHI-cut SP64-HbΔ6; lanes 5–7: 10, 30 and 100-fold molar excess of β-globin RNA containing exons 1 and 2 but lacking intron 1 (SP6 transcript of BamHI-cut SP64-Hβ-IVS 1,2).

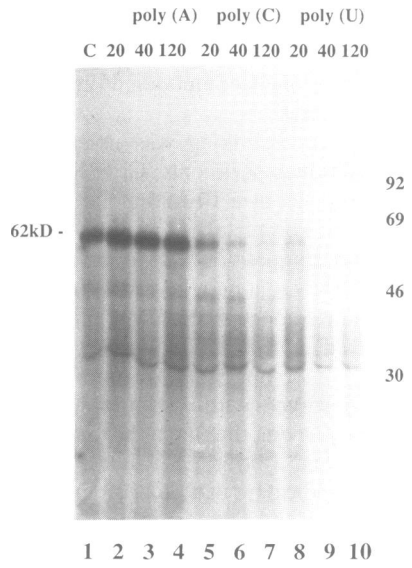

Figure 5. Competition by ribohomopolymers. Lane 1: ^{32}P-labeled 61 nt RNA alone; lanes 2–4: 20, 40 and 120-fold molar excess of poly(A); lanes 5–7, 20, 40 and 120-fold molar excess of poly(C); lanes 8–10: 20, 40 and 120-fold molar excess of poly(U).

6C). Crosslinking of p62 was not sensitive to prior micrococcal nuclease digestion of the HeLa nuclear extract (Figure 6C). The crosslinking of p62 was extremely sensitive to KCl concentration,

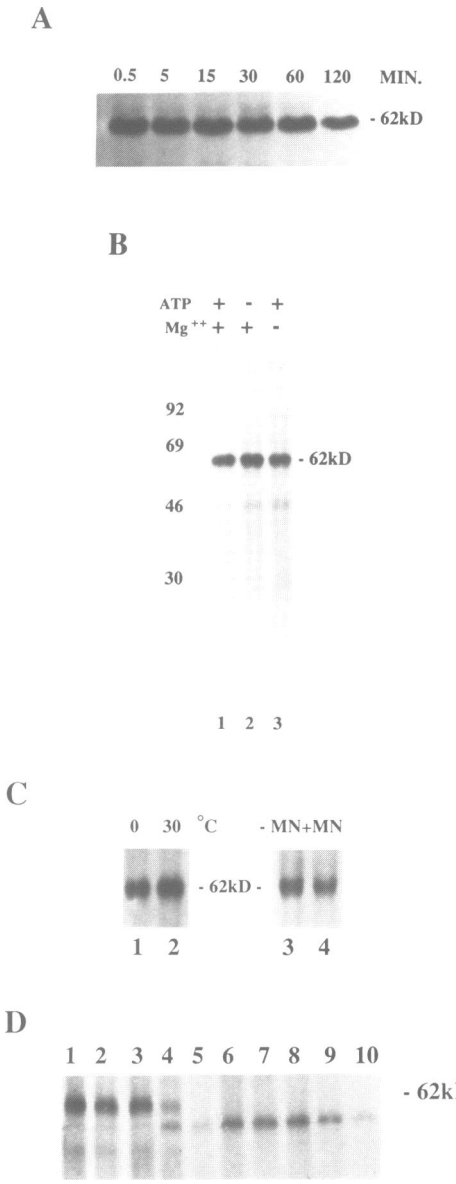

Figure 6. p62 crosslinking parameters. *A*. Time course. ^{32}P-labeled 61 nt RNA was incubated in nuclear extract at 30°C for the times indicated, followed by crosslinking, nuclease digestion and electrophoresis. *B*. ATP/Mg^{++} independence. Nuclear extract contained ATP and/or Mg^{++} as indicated. A similar level of p62 crosslinking was also observed when both ATP and Mg^{++} were omitted (not shown). *C*. Lanes 1 and 2, incubation in nuclease extract at the temperatures indicated; lanes 3 and 4: incubation in nuclear extract that was pre-incubated for 30 min. at 30°C in either the absence (lane 3) or presence (lane 4) of micrococcal nuclease (500 units/ml), followed by addition of 2 mM EGTA (1 mM $CaCl_2$ present throughout in both cases). *D*. Dependence of p62 crosslinking on KCl concentration. Incubation of ^{32}P-labeled 61 nt RNA was in nuclear extract containing varying KCl concentrations: 15 mM KCl (lane 1), 30 mM, 60 mM, 100 mM, 150 mM (lane 5), 200 mM, 250 mM, 300 mM, 400 mM and 500 mM (lane 10).

decreasing sharply at 100 mM KCl (Figure 6D). At 100 mM KCl and higher, we noted the crosslinking of a ~57,000 mol. wt. protein. Whether this is represents a proteolytic fragment of p62 produced and/or crosslinked to the RNA at the higher KCl concentrations or is a different protein altogether has not been determined.

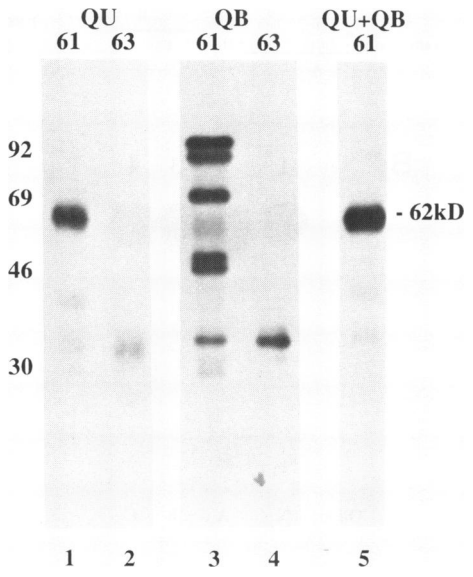

Figure 7. Q-Sepharose fractionation of p62. Nuclear extract was subjected to chromatography an Q-Sepharose (see Materials and Methods) and crosslinking of proteins to ^{32}P-labeled 61 nt RNA or to ^{32}P-labeled 63 nt antisense RNA was assayed in the Q-Sepharose unbound (QU) and bound (QB) fractions, and in a 1:1 recombined QU and QB fraction, as indicated by the code at the top of each lane.

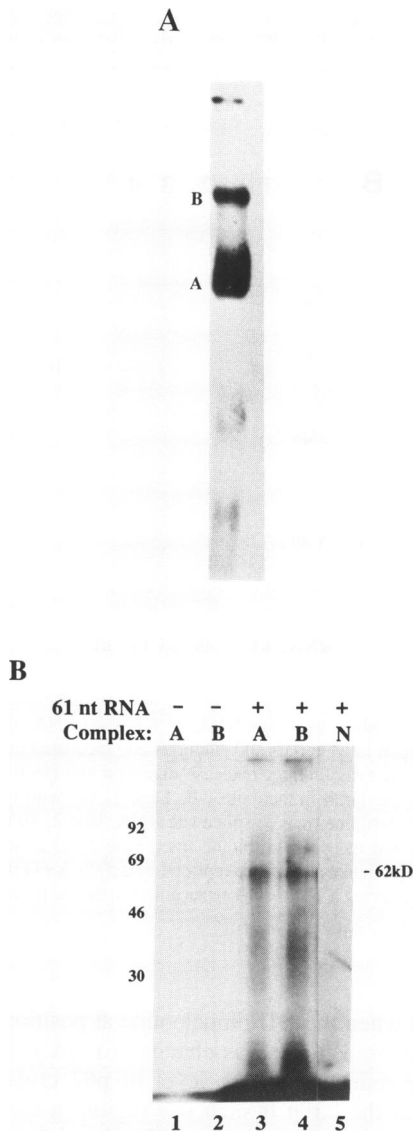

Figure 8. Presence of p62 in active splicing complexes. *A*. ^{32}P-labeled Ad10 pre-mRNA was incubated in HeLa nuclear extract for 30 min. at 30 °C under splicing permissive conditions, followed by electrophoresis of splicing complexes in a non-denaturing 4% polyacrylamide-0.5% agarose gel (9). *B*. A and B splicing complexes formed on low ^{32}P specific activity Ad10 pre-mRNA were displayed as in *A*. Gel slices containing A or B complexes were incubated in 500 mM KCl followed by lowering the ionic strength to 60 mM KCl. ^{32}P-labeled 61 nt RNA was added, followed by UV crosslinking, nuclease digestion and SDS-polyacrylamide gel electrophoresis (see Materials and Methods for details). Lane 1, A complex, no 61 nt ^{32}P-RNA incubated with gel slice prior to crosslinking; lane 2, B complex, no 61 nt ^{32}P-RNA incubated with gel slice prior to crosslinking; lane 3, A complex, ^{32}P-labeled 61 nt RNA incubated with gel slice prior to crosslinking; lane 4, B complex, ^{32}P-labeled 61 nt RNA incubated with gel slice prior to crosslinking. Autoradiographic exposure, 16 hours; lane 5, ^{32}P-61 nt RNA was incubated with gel slices corresponding to the positions of A and B splicing complexes but from a non-denaturing preparative gel of nuclear extract to which the adeno pre-mRNA had not been added.

p62 crosslinking in active splicing complexes

To investigate the possible relationship of p62 to mRNA splicing, we fractionated HeLa nuclear extract by Q-Sepharose chromatography. Neither the non-bound ('QU') nor bound ('QB') fractions supported splicing of human β-globin pre-mRNA, but recombined QU and QB displayed efficient splicing (J.W. and T.P., unpublished results). Crosslinking of p62 to the 61 nt RNA was observed in the QU fraction, but almost no p62 crosslinking occurred in the QB fraction (Figure 7). When the two Q-Sepharose fractions were recombined, only p62 was crosslinked to the 61 nt RNA (Figure 7). Thus, p62 crosslinks to the 61 nt RNA in a splicing non-permissive fraction (QU) and in a splicing-permissive reconstituted system (QU + QB), but does not crosslink to the 61 nt RNA in another splicing non-permissive fraction (QB).

We next asked whether p62 is a component of active splicing complexes. Adenovirus pre-mRNA splicing complexes were resolved on a non-denaturing polyacrylamide gel (9) as shown in Figure 8A. The A and B splicing complexes were cut out of the gel and subjected to UV-mediated RNA-protein crosslinking. After nuclease digestion and SDS-polyacrylamide gel electrophoresis, several ^{32}P-nucleotide-linked proteins were resolved including one of ~62,000 mol. wt. (data not shown); the pattern was similar to the one reported by Garcia-Blanco *et al.* (see Figure 4 in ref. 5.).

To more specifically ask if p62 was present in active splicing complexes, we KCl-dissociated A and B splicing complexes in the gel slices (see Methods and Materials) and then challenged these fractions with the ^{32}P-labeled 61 nt RNA, followed by crosslinking, nuclease digestion and gel electrophoresis. As shown in Figure 8B, lanes 3 and 4, p62 was present in both A and B splicing complexes. No p62 crosslinkable proteins were present in the non-denaturing gel fractions corresponding to the positions of A and B splicing complexes when no adeno pre-mRNA was added to the nuclear extract (Figure 8B, lane 5; see legend for details).

Site of p62 crosslinking

We next undertook experiments to define in more detail the binding site of p62. A somewhat reduced level of p62 crosslinking

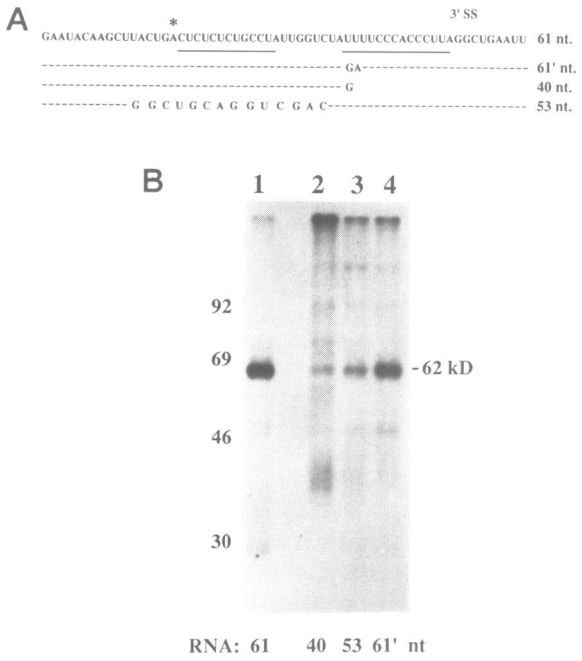

Figure 9. p62 crosslinking to mutant RNAs. ^{32}P-labeled RNAs shown in *A* were generated as detailed in Materials and Methods and incubated in nuclear extract at 4°C, followed by crosslinking, nuclease digestion and electrophoresis. (The asterisk indicates the lariat branch site.) *B*. Lane 1, the wild type 61 nt lariat branch site/polypyrimidine tract/3′ splice site RNA; lane 2, 40 nt RNA lacking the 3′ splice site and the polypyrimidine tract; lane 3, 53 nt RNA having the lariat branch site and flanking sequence replaced by a 57% A+G pGEM sequence; lane 3, 61 nt RNA having a UU → GA replacement in the the polypyrimidine tract.

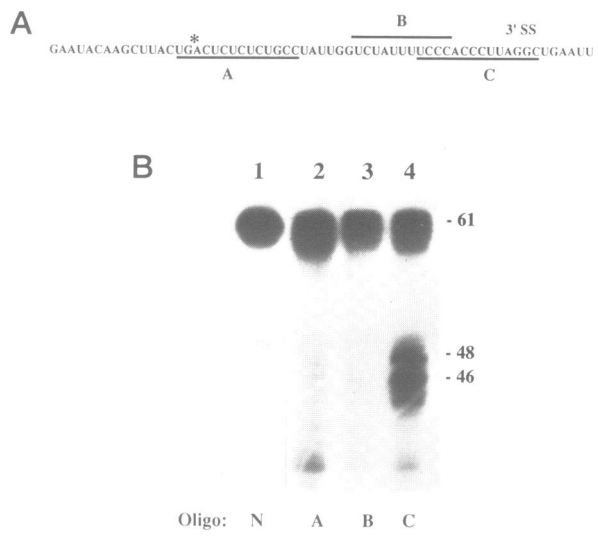

Figure 10. RNase H protection. ^{32}P-labeled 61 nt RNA was incubated in the Q-Sepharose unbound fraction (see Figure 7) for 10 min., prior to addition of one of the oligodeoxynucleotides complementary to the RNA sites shown in *A*. (The asterisk indicates the lariat branch site.) After 30 min., the RNA was isolated by phenol-chloroform extraction and analyzed by electrophoresis in a 10% polyacrylamide-7M urea gel. *B*. Lane 1, no oligo; lane 2, oligo A; lane 3, oligo B; lane 4, oligo C.

was observed when the UU dinucleotide at positions 22 and 23 after the lariat branch site was changed to GA (Figure 9, lane 4). A greater reduction in the level of p62 crosslinking was observed when the lariat branch site region in the 61 nt RNA was replaced with a 57% purine pGEM1 polylinker sequence (Figure 9, lane 3), or when the polypyrimidine tract was deleted (Figure 9, lane 2). Mutation to U of the lariat adenosine itself in the normal 61 nt RNA had no effect on p62 crosslinking (not shown).

To define more precisely the site of protein interaction within the 61 nt RNA, we carried out RNase H protection experiments using the endogenous RNase H activity in the nuclear extract QU fraction. ^{32}P-labeled 61 nt RNA was incubated in the QU fraction (Figure 7), followed by the addition of oligodeoxynucleotides complementary to the regions of the 61 nt RNA shown in Figure 10A (These oligos all induced cleavage of the 61 nt RNA by *E. coli* RNase H in buffer; data not shown.). RNase H cleavage of the 61 nt RNA by oligos A and B was blocked by prior incubation of the substrate RNA in the QU fraction, indicating the interaction of proteins with these sites (Figure 10B, lane 2 and 3). In contrast, an oligo complementary to the 3′ end of the polypyrimidine tract and spanning the 3′ splice site (oligo C) induced RNase H cleavage, indicating that this region is less protected by associated proteins (Figure 10, lane 4). We also characterized the protected regions of the 61 nt RNA by RNase A digestion and secondary RNase T1 analysis, which confirmed the interaction of proteins with the regions complementary to oligos A and B (data not shown).

DISCUSSION

The polypyrimidine stretch located in the 3′ region of mammalian pre-mRNA introns (13) is a hot spot (probably the hottest spot) during the early stages of spliceosome formation. This region is a binding site for several proteins including the A1, C and D hnRNP proteins (1), the reported 100 kD (3) and 70 kD (4) intron binding proteins, and U2AF (6), and p62 (5; also this report). It is improbable that all of these proteins make direct contact with the polypyrimidine tract at the same time. Some may interact at this site via other proteins that directly recognize the polypyrimidine tract. It is also possible that there is sequential binding and/or repositioning of these various proteins on the polypyrimidine tract as this dynamic center undergoes spliceosome assembly.

A 62,000 mol. wt. HeLa nuclear extract protein has previously been shown to recognize the polypyrimidine tract of adenovirus and β-globin pre-mRNA introns by UV crosslinking (5). Moreover, purified U2AF—which contains two polypeptides of approximately 65,000 and 35,000 mol. wt.—has been shown to bind to the polypyrimidine tract/3′ splice site region (6). We have no reason to doubt that our p62 is the same protein as described in these two previous studies (5,6). Its molecular weight (in the form of carrying crosslinked nucleotides), its site of crosslinking, its extremely rapid crosslinking at 0°C, the lack of an ATP requirement, and its presence in active splicing complexes all argue that our p62 is the same protein as previously described (5,6).

In the present study we used a 61 nucleotide long RNA spanning the lariat branch site/polypyrimidine tract/3′ splice site. While a complex pattern of proteins is crosslinked to a β-globin pre-mRNA containing the first two exons and intron (ref. 2; see also Figure 1), only p62 is crosslinked when the 61 nt RNA is used. Thus, our results establish that the crosslinking of p62 to

the polypyrimidine tract does not require upstream intron sequences or flanking exon elements, nor factors that associate with such sequences. This is in agreement with one of the previous studies in which it was reported (albeit as data not shown) that p62 was efficiently crosslinked to a 63 nt RNA spanning the lariat branch site/polypyrimidine tract of an adenovirus intron (5).

By the criterion of protein-RNA crosslinking, p62 appears to possess a very selective affinity for the 61 nt RNA, as compared to other proteins present in HeLa nuclear extract. This conclusion emerges from the analysis of protein crosslinking to the 61 nt RNA in nuclear protein fractions obtained by Q-Sepharose chromatography. In the non-bound fraction, p62 was the only protein that could be crosslinked to the 61 nt RNA. However, the bound fraction contained several other proteins that were readily crosslinked. These latter proteins clearly recognize either the 61 nt nucleotide sequence, or its pyrimidine richness, since they did not crosslink efficiently to a RNA antisense to the 61 nt RNA (Figure 7). Yet, when the two Q-Sepharose fractions were combined, only p62 was crosslinked to the 61 nt RNA, indicating that it out-competes the other aformentioned proteins for binding and/or crosslinking to the polypyrimidine tract.

Our results also provide new information on the RNA sequence recognition properties of p62. Crosslinking of p62 was competed by an oligodeoxynucleotide having sequence homology with the 61 nt RNA, showing that p62 does not require ribose 2'-OH groups or uracil for its binding interaction with RNA. While the selectivity of p62 for pyrimidines is clearly shown by the hierarchy of crosslinking to the various mutant and control RNAs we tested, and by the relative crosslinking competition strengths of poly(U) and poly(C) versus poly(A), we also found that mere replacement of a UU dinucleotide in the polypyrimidine tract by a GA dinucleotide was sufficient to reduce crosslinking. This result suggests that a certain length of consecutive pyrimidines, or a proximity of polypyrimidine motifs, is necessary for efficient p62 crosslinking. In the β-globin intron region we have studied the longest consecutive pyrimidine tracts are, counting the lariat branch site as 'position 0', +1 to +8 and +22 to +28 (14). The aforementioned UU → GA mutant that displays diminished p62 crosslinking has the latter $(Py)_7$ tract reduced to $(Py)_5$, indicating that the length of this polypyrimidine tract is critical. However, our results with the 53 nt RNA containing a 57% A+G pGEM polylinker replacement at the lariat branch site region (Figure 9) establish that this region is also important for p62 crosslinking. Since the lariat branch site adenosine itself is not important for p62 crosslinking, we conclude that surrounding sequences are necessary (Figure 9). This extends the known domain of p62 interaction to the sequences immediately adjacent to the lariat branch site. The p62 binding domain demonstrated by these results is consistent with the oligo/RNase H protection experiments (Figure 10) although it must be borne in mind that the latter results do not necessarily reflect protection only by p62.

A final point worth noting is that p62 was not crosslinkable to the polypyrimidine tract when this sequence was in the form of double-stranded RNA (Figure 3). From this we infer that p62 does not possess duplex RNA unwinding activity, nor does the HeLa nuclear extract contain unwinding activity for this particular 61 base-pair RNA. Yet, in the secondary structure of a human β-globin pre-mRNA determined by chemical probes and RNA-RNA crosslinking methods (15,16) the polypyrimidine tract of the first intron is base-paired with a purine-rich stretch located at nucleotides 83−110 of exon 2. Given the extremely rapid and high affinity binding of p62 to the exon-lacking 61 nt polypyrimidine tract RNA we observed *in vitro*, it is quite possible that, *in vivo*, p62 binds to the polypyrimidine tract on nascent β-globin pre-mRNA before exon 2 is even transcribed (see Discussion in ref. 17).

ACKNOWLEDGMENTS

Supported by NIH grant GM-21595-16 to T.P. We thank Maria Konarska, Adrian Krainer, Tom Maniatis, Douglas Melton and Phillip Sharp for plasmids, and Vipin Kohli for synthesizing oligodeoxynucleotides in the Worcester Foundation's central facility (supported by Cancer Center core grant P30-12708-19). We are grateful to Ann Kleinschmidt, John Maddalena, Sandra Mayrand, Colleen McKiernan, Molly Rhoadhouse and Jamal Temsamani for advice and assistance, and we thank Sandra Mayrand and Jamal Temsamani for perspicacious comments on the manuscript. We also gratefully acknowledge Carol Tuttle for word processing.

REFERENCES

1. Swanson, M.S. and Dreyfuss, G. (1988) *EMBO J.* **7**: 3519−3529.
2. Mayrand, S.H. and Pederson, T. (1990) *Nucleic Acids Res.* **18**: 3307−3318.
3. Tazi, J., Alibert, J., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C. and Jeanteur, P. (1986) *Cell* **47**: 755−766.
4. Gerke, V. and Steitz, J.A. (1986) *Cell* **47**: 973−984.
5. Garcia-Blanco, M.A., Jamison, S.F. and Sharp, P.A. (1989) *Genes & Dev.* **3**: 1874−1886.
6. Zamore, P.D. and Green, M.R. (1989) *Proc. Natl. Acad. Sci. USA* **86**: 9243−9247.
7. Ruskin, B., Zamore, P.D. and Green, M.R. (1988) *Cell* **52**: 207−219.
8. Dingnam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acids Res.* **11**: 1475−1489.
9. Konarska, M.M. and Sharp, P.A. (1987) *Cell* **49**: 763−774.
10. Economidis, I.V. and Pederson, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**: 4296−4300.
11. Mayrand, S.H., Pedersen, N. and Pederson, T. (1986) *Proc. Natl. Acad. Sci. USA* **83**: 3718−3722.
12. Marciniak, R.A., Garcia-Blanco, M.A. and Sharp, P.A. (1990) *Proc. Natl. Acad. Sci. USA* **87**: 3624−3628.
13. Ohshima, Y. and Gotoh, Y. (1987) *J. Mol. Biol.* **195**: 247−259.
14. Lawn, R.M., Efstratiadis, A., O'Connell, C. and Maniatis, T. (1980) *Cell* **21**: 647−651.
15. Teare, J. and Wollenzien, P.L. (1989) *Biochem.* **28**: 6208−6219.
16. Teare, J. and Wollenzien, P.L. (1990) *Nucleic Acids Res.* **18**: 855−864.
17. Calvet, J.P. and Pederson, T. (1977) *Proc. Natl. Acad. Sci. USA* **74**: 3705−3709.