Intron sequences and the length of the downstream second exon affect the binding of hnRNP C proteins in an *in vitro* splicing reaction

Prabhat Goswami and Carlos J. Goldenberg

Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, FL 33101, USA

Received August 8, 1987; Revised and Accepted April 29, 1988

ABSTRACT

The proteins that are in direct contact with the pre-mRNA in an *in vitro* splicing reaction were analyzed by UV cross-linking experiments. Six major proteins (120, 55, 44, 42, 39 and 38 KD) and three minor polypeptides (84, 72 and 63 KD) were detected. The predominant proteins 44, 42 KD belong to the class of hnRNP C proteins since they were immunoprecipitated by monoclonal antibodies directed against hnRNP C proteins. The cross-linked proteins were not detected in the absence of Mg^{2+} , ATP or when RNA lacking introns were used as substrates in the splicing reactions. The effect of exon sequences on the binding efficiency for the photocrosslinked proteins was investigated. Transcripts containing a second exon of 24 nucleotides for the β -globin or 107 nucleotides for the mouse insulin, yielded a reduced amount of cross-linked proteins when compared with "full length" pre-mRNAs. Sequences within the first exon of the β -globin pre-mRNA did not affect the binding efficiency of these proteins. The reduced binding efficiency of the cross-linked proteins for the truncated β -globin or mouse insulin pre-mRNAs correlated with the lower efficiency for *in vitro* splicing. Substitutions with unrelated sequences in the β -globin second exon restore the binding of the cross-linked proteins indicating that the length of the second exon and not specific sequences are relevant for the binding efficiency of these proteins. The SP6/mouse insulin oligonucleotides cross-linked to the hnRNP C proteins were isolated and sequenced. A 17-mer was located in the second exon (134 nucleotides downstream from the 3' splice site) and a 14-mer in the intron region (25 nucleotides downstream the 5' splice site). The β -globin oligonucleotides cross-linked to the hnRNP C proteins were a 13-mer in the second exon (28 nucleotides downstream the 3' splice site) and an 8-mer in the first exon (81 nucleotides downstream the 5' end of the pre-mRNA). Our results indicate that the hnRNP C proteins interact with those oligonucleotides located in different regions of the pre-mRNA. The binding efficiency of those proteins, however, depends on the length of the second exon and the presence of intron sequences (secondary and/or tertiary pre-mRNA structure).

INTRODUCTION

In eukaryotic cells, most protein-coding genes contain stretches of non-coding nucleotides (introns) which are first transcribed as parts of the messenger RNA precursor (pre-mRNA's) and subsequently removed by RNA splicing (for a review, see 9 and 33). Sedimentation experiments have indicated that the splicing of pre-mRNA involves the formation of a large ribonucleoprotein-complexes sedimenting at 50-60S in sucrose or glycerol gradients. Assembly of these complexes, known as spliceosomes, requires ATP (5,17,21,22,34), Sm small nuclear ribonucleoproteins (SnRNPs) (17,21,22) and heterogeneous

nuclear ribonucleoproteins (hnRNPs)(13). At least the Sm SnRNPs containing U₁, U₄/U₆ and U₅ RNAs are components of the spliceosome complex (20,22). Experiments involving either anti-snRNP antibodies (16,26,32,38) or oligonucleotide-targeted degradation of individual snRNAs in splicing extracts have shown that U₁, U₂ and U₄/U₆ snRNPs (6,7,25,26) are essential splicing components, required for the cleavage reaction. A protein associated with a fourth Sm snRNP (probably containing U₅ RNA) have been shown to specifically recognize the 3' splice site of a pre-mRNA (10,20,36).

The eukaryotic mRNA precursors are found in the nucleus associated with specific proteins forming the hnRNP complexes (for a review, see 24 and 28). These structures are thought to be involved in the packaging and processing of the hnRNA. Biochemical evidence of the hnRNP complex, demonstrates particles that sediment at about 40S (2,3 and reviewed in 28). The appearance by electron microscopy of the isolated 40S particles and of particles seen on nascent transcripts in spread chromatin preparations suggest a "beads on a string" structure for the hnRNPs (2-4,28,30). Large hnRNP complexes are formed by multiples of 40S particles connected by highly nuclease-sensitive regions of hnRNA (11). The major components of the 40S hnRNP particle include the A₁-A₂ (34 and 36 KD), B₁-B₂(38 and 37 KD), C₁-C₂ (41 and 42 KD), a 53 KD protein and doublets at 68 and 120 KD (2,12,15). These proteins are known to be present in the nuclear extracts that support *in vitro* splicing (13). The splicing reaction was inhibited by monoclonal antibodies which recognize the C proteins. The hnRNP C proteins were found in the 60S splicing complex, and depletion of these proteins from splicing extracts abolishes the splicing activity and the capacity to form a 60S splicing complex (13). Recently, we have shown (37) that unwinding of the pre-mRNA secondary structure is necessary for *in vitro* splicing. The RNA helicase requires ATP and is inhibited by monoclonal antibodies to the hnRNP C proteins (C. Goldenberg, unpublished results), indicating that those proteins either are components of or are required for assembly of the unwinding activity.

Here we have analyzed, by UV cross-linking experiments, the proteins that are in direct contact with the pre-mRNA during an *in vitro* splicing reaction. The predominant proteins (44, 42 KD) that we have detected, belong to the class of hnRNP proteins since they were immunoprecipitated by monoclonal antibodies directed against the hnRNP C proteins. The UV cross-linked proteins were not detected either in the absence of Mg²⁺, ATP or when RNA lacking introns were used as substrates in the splicing reactions. The effect of exon sequences on the binding efficiency of the photocrosslinked proteins was investigated. We found that transcripts truncated within the second (3') exon of the mouse insulin or the β -globin gene yielded a reduced amount of cross-linked proteins. We show that substitutions with unrelated sequences in the second exon of the β -globin pre-mRNA restore the binding of the cross-linked proteins. Deletions within the first

(5') exon of the β -globin pre-mRNA did not affect the binding efficiency of these proteins. The RNA segments of the mouse insulin and β -globin pre-mRNAs that were UV cross-linked to the C hnRNP proteins were isolated and sequenced. These experiments demonstrate that the hnRNP C proteins interact with specific oligonucleotides located in different regions of the pre-mRNA. Moreover, the data show that the binding efficiency of the cross-linked proteins depends on the length of the second exon and the presence of intron sequences and not on specific sequences.

MATERIALS AND METHODS

SP6 Transcriptions and in vitro splicing.

The pre-mRNAs were prepared in the presence of all four [α - 32 P] ribonucleotides, as described (14,22). The SP64/ β -globin plasmid DNAs, HBPL which have a substitution in the second exon 14 nucleotides downstream the 3' splice site with SP64 plasmid sequences and the HBHaeIII which have the first exon deleted up to 21 nucleotides from the 5' splice site (see diagram Fig. 4B), were a gift from Drs. P. Furdon and R. Kole, University of North Carolina. The transcripts were gel purified, eluted and used as substrates for splicing reactions (14,22). The pre-mRNAs (4×10^5 cpm in 25 μ l reactions) were incubated at 30°C for different times in splicing buffer (12 mM Hepes, pH 7.9, 60 mM KCl, 3.2 mM MgCl₂, 1 mM DTT, 12% glycerol, 2.1% w/v polyvinyl alcohol, 1000 U/ml RNasin, 0.5 mM ATP and 20 mM creatine phosphate) plus 12 μ l (60 μ g protein) of HeLa nuclear extract (22).

UV-crosslinking of splicing reactions.

The splicing reactions (100 μ l), after incubations were transferred to a microcuvette and UV irradiated with constant stirring at 4°C for 10 min. The UV irradiation source (254 nm wavelength), placed at 3 cm from the samples, was a pair of 2.5 x 42 cm 15-W germicidal lamps (Sylvania G15-T8), 3.5×10^5 ergs/mm². Equal aliquots were normalized to their protein concentrations (8) and digested for 60 min. at 37°C with pancreatic RNase at 25 μ g/ml and micrococcal nuclease 400 U/ml in the presence of CaCl₂ 1 mM. The reactions were precipitated with 10% TCA (trichloroacetic acid) and resuspended in 0.0625 M Tris-HCl, pH 6.8; 10% glycerol, 2% SDS, 5% mercaptoethanol (protein loading buffer). The proteins were analyzed on 12% polyacrylamide-SDS gels. Electrophoresis of the gel (22 cm x 16 cm) was for 16 hr. at 25 mA. The autoradiograms were scanned using a soft laser scanning densitometer.

RNA Sequence Analysis.

The oligonucleotides cross-linked to the hnRNP C proteins were purified as described in the legend to Fig. 5. RNA was sequenced as described (1,27). 32 P-3'-end labeled RNA in a solution containing 250 μ g/ml tRNA was partially digested with 1 unit of RNase PhyM, B or T1 (pH5) and U2 (pH 3.5) at 50°C for 15 min. Partial alkaline

hydrolysis in 50 mM $\text{NaHCO}_3/\text{Na}_2\text{CO}_3$ at pH 9.0 was carried out at 90°C for 5 min. Samples were loaded on 8M urea - 20% polyacrylamide gel and electrophoresed at 1600 V for 8 hrs.

RESULTS

To examine the proteins that bind tightly to the pre-mRNA during *in vitro* splicing we have performed RNA/protein UV cross-linking of splicing reactions. This technique makes it possible to isolate under denaturing conditions proteins that must have been tightly associated with the pre-mRNA *in vitro* to become cross-linked to it by UV light.

Figure 1: Time course of proteins cross-linked to pre-mRNAs during *in vitro* splicing. Splicing reactions (100 μl) containing either SP6/ β -globin (Panel A) or SP6/mouse insulin pre-mRNAs (Panel B) were incubated at 30°C for the indicated times. Lane 1: 5 min., Lane 2: 20 min., Lane 3: 90 min., Lane 4: 180 min. and Lane 5: 240 min. The reactions were UV irradiated and the cross-linked proteins analyzed as described in Materials and Methods. The samples were normalized to equal aliquots, according to their protein concentrations (8) before digestions with RNase and micrococcal nuclease, and loaded on the gel (Materials and Methods). The numbers to the left of lane 1 are $\text{Mr} \times 10^{-3}$ of pre-stained protein standards. The diagram below panels A and B represent the structure of the SP6/ β -globin and the SP6/mouse insulin pre-mRNAs. Boxes: exons; lines: introns; numbers above: length in nucleotides of the exons and introns.

Splicing reactions, incubated for different times at 30°C, with either the SP6/mouse insulin or the SP6/ β -globin pre-mRNAs were UV irradiated and digested with pancreatic ribonuclease and micrococcal nuclease as described in Materials and Methods. The proteins were then analyzed on SDS-PAGE and visualized by autoradiography. Figure 1 shows that a similar pattern of UV cross-linked proteins was obtained with either the β -globin (Fig. 1A) or the mouse insulin pre-mRNA (Fig. 1B). Six major proteins migrating as 120, 55, 44, 42, 39 and 38 KD and three minor polypeptides migrating as 84, 72 and 63 KD were detected. The cross-linking is strictly dependent on UV light, and splicing reactions that were not UV irradiated but digested with pancreatic ribonuclease and micrococcal nuclease did not show any protein on the autoradiogram (data not shown).

Figure 2: Requirements for proteins cross-linked to pre-mRNAs during in vitro splicing reactions.

A: The splicing reactions were incubated as above in the presence or absence of Mg²⁺ and ATP, as indicated. The reactions were UV irradiated and the cross-linked proteins analyzed as described in Materials and Methods. The numbers to the left are Mr x 10⁻³ of pre-stained standards.

B: Time course of UV irradiation: The splicing reactions were incubated for 20 min. at 30°C, UV-irradiated for 5 min. (lane 1), 10 min. (lane 2) or 30 min. (lane 3) and the cross-linked proteins analyzed as described in Materials and Methods.

The maximum amount of UV cross-linked proteins were detected from splicing reactions incubated for 20 min. at 30°C. With increasing length of incubation (180, 240 min.) the total amount of cross-linked proteins decreased over 30-fold. It should be noted that instability of the 60S (spliceosome) complex, after 90 min of incubation, was previously reported (17,21,22).

Analysis of the mouse insulin oligonucleotides cross-linked to the 44-42 KD proteins, from 10 min. or 90 min. reactions, revealed a set of oligonucleotides migrating as a 17-mer and a 14-mer (Fig. 5A, lanes 2 and 3). Thus, most likely the 44, 42 KD proteins contact the same regions of the pre-mRNA, at least for the mouse insulin, during the 90 min. of incubation.

The cross-linking of the proteins to the pre-mRNA requires Mg^{2+} and ATP (Fig. 2A). The pre-mRNA transcripts for the experiment in figure 2A were stable after incubation in splicing reactions and UV cross-linking (data not shown). The time of exposure to UV light that we have used in our experiments (see Materials and Methods) was 10 min. (Fig. 2B). The possibility exists that some of the bands are products of protein-protein cross-linking, however, the same pattern of polypeptides was observed with a shorter exposure to UV light (Fig. 2B, lane 1).

Characterization of the cross-linked proteins with specific antibodies

To characterize the cross-linked proteins we have performed immunoselections with monoclonal antibodies directed against the hnRNP C proteins or monoclonal antibodies directed against the Sm determinant of SnRNPs. The immunoprecipitations were carried out at 420 mM NaCl (Fig. 3). Under these conditions the hnRNP complex dissociates (2,13,15) and most of the hnRNP proteins are not precipitated with the C proteins-complex. The immunoprecipitates were selected on a protein-A agarose column and analyzed on SDS-PAGE (see Materials and Methods). Monoclonal antibodies (4F4) directed against the antigenically related C₁ and C₂ hnRNP protein immunoselected only the 44-42 KD cross-linked proteins (Fig. 3A, lane 7). All of the other major cross-linked proteins (120,55,39 and 38 KD) were detected in the flow-through fraction of the column (Fig. 3A, lane 6). These results indicate that the cross-linked proteins 44-42 KD belong to the class of hnRNP C proteins.

The monoclonal antibodies control (Fig. 3A lanes 4 and 5) and anti-Sm (Fig. 3A lanes 2 and 3) were not able to select the cross-linked proteins.

To further characterize the cross-linked hnRNP C proteins we have directly compared the uncross-linked C hnRNP proteins present in the splicing extract with the [³²P]-cross-linked proteins. The proteins present in the splicing extract, and the proteins cross-linked to [³²P]-labeled mouse insulin pre-mRNA were first immunoselected with 4F4 monoclonal antibodies or control antibodies and then electrophoresed side by side on SDS-PAGE. The proteins were then electroblotted onto nitrocellulose and probed

Figure 3: Antibody selection of the cross-linked proteins. Splicing reactions containing the SP6/mouse insulin pre-mRNA were incubated for 20 min at 30°C, UV irradiated, digested with RNase and micrococcal nuclease (Materials and Methods) and adjusted to 420 mM NaCl. Monoclonal antibodies (4F4) which recognize the C₁ and C₂ hnRNP proteins or control antibodies (SP₂/0), were kindly provided by Dr. G. Dreyfuss (Northwestern University, Ill.). The antibodies were purified from ascites fluid (14). The protein concentration was adjusted to 5 mg/ml and 40 µl of antibodies were incubated on ice for 60 min. with 100 µl splicing reactions (400 µg protein) to immunoprecipitate the cross-linked proteins. The samples were then applied to a protein-A agarose column and the flow-through fraction was collected. The column was then washed extensively with splicing buffer, eluted with protein loading buffer and proteins analyzed on SDS-PAGE (Materials and Methods).

A: Total cross-linked proteins without antibodies (lane 1). Monoclonal antibodies anti-Sm were used to immunoprecipitate the cross-linked proteins. Flow-through fraction from the protein-A column (lane 2). Eluted fraction (lane 3). Monoclonal antibodies control SP₂/0. Flow-through fraction (lane 4). Eluted fraction (lane 5). Monoclonal anti-hnRNP C proteins 4F4. Flow-through fraction (lane 6). Eluted fraction (lane 7).

B: Cross-linked vs. uncross-linked hnRNP C proteins. The extract used for splicing reactions (50 µl, 250 µg protein) or splicing reactions (100 µl) containing [³²P]-labeled SP6/mouse insulin pre-mRNA were UV irradiated as above, adjusted to 420 mM NaCl and immunoprecipitated with either 4F4 monoclonal anti-hnRNP C proteins or monoclonal control SP₂/0. The samples were applied to a protein-A column and the eluted fractions analyzed on SDS-PAGE. The proteins were transferred to nitrocellulose as described (15). The blot was then incubated first with 4F4 monoclonal antibodies and then with ¹²⁵I-anti-mouse antibodies. Uncross-linked proteins immunoprecipitated with 4F4 monoclonal antibodies (lane 1). [³²P]-cross-linked proteins immunoprecipitated with 4F4 (lane 2) or control SP₂/0 (lane 3).

Figure 4: Role of exon sequences on the binding efficiency of U.V. cross-linked proteins.

A: The SP6/ β -globin DNA was linearized with different restriction enzymes and used for in vitro transcriptions with SP6 polymerase. The transcribed RNAs, which are truncated at the appropriate restriction sites, are diagrammed below. The β -globin DNA clone was restricted with Bam HI (Lane 1), Sau3A (lane 2), AvaII (lane 3), AccI (lane 4) and MboII (lane 5). The numbers on the left of the diagram correspond to the transcript used for the experiments shown on each lane. Open boxes: exons, lines: introns. The numbers to the right of each transcript indicate the nucleotide length of the transcript, the numbers on top of the boxes and lines indicate the nucleotide length of exons and introns, respectively. The different pre-mRNAs were gel purified and incubated in splicing reactions for 20 min. at 30°C. The reactions were UV-irradiated, the samples normalized to equal aliquots according to their protein concentration (8), and the cross-linked proteins analyzed on SDS-PAGE (Materials and Methods).

B: Transcripts were prepared from a SP6/ β -globin clone DNA which have a substitution in the second exon, 14 nucleotides downstream the 3' splice site, with SP64 sequences (Plasmid HBPL gift from Dr. R. Kole, University of North Carolina). The "full length" pre-mRNA is represented in lane 1 (see row 1 in the diagram below the panel). HBPL linearized with Hpa I (lane 2) and EcoRI (lane 3). Transcripts were also prepared from a SP6/ β -globin clone DNA which have as a first exon only 21 nucleotides upstream the 5' splice site (Plasmid HBHaeIII, gift from Dr. Kole). HBHaeIII linearized with Sau3A (lane 4), AvaII (lane 5), and MboII (lane 6). The cross-linked proteins were analyzed as above. The autoradiogram on the right was from another gel than in lanes 1-3. Marks on the right indicate molecular weight protein standards (111, 68, 45 and 24 KD).

C: The SP6/mouse insulin DNA clone was linearized with EcoRI (lane 1), MboII (lane 2), AvaII (lane 3) and HaeIII (lane 4). SP6/mouse insulin transcript lacking introns (lane 5). The transcripts were gel purified and the cross-linked proteins analyzed as above.

with the 4F4 monoclonal antibodies (Western Blot). The cross-linked hnRNP C proteins migrated as 44-42 KD, about 1 KD increase in relation to the uncross-linked C proteins (Fig. 3B, lanes 1 and 2). It is likely that some of the C proteins migrating as 42 KD in the [32 P]-cross-linked sample (Fig. 3B, lane 2) represents unreacted and uncross-linked C proteins that light up on the Western blot probed with 4F4 antibodies. The unreacted uncross-linked 41 KD C protein was barely detected in those samples. We do not know if this represents a different efficiency of C1 and C2 proteins to cross-linked pre-mRNAs or it is just a difference in stabilities of both proteins to the manipulations performed in the cross-linked samples. Thus, we conclude that a slight change in the migration of the hnRNP C protein occurs upon UV cross-linking.

The Effect of Exon and Intron sequences on the binding of cross-linked proteins.

We have examined a series of SP6 RNA precursors truncated in exon 2, ranging in size from 14 to 209 nucleotides for the β -globin pre-mRNA and 107 to 226 nucleotides for the mouse insulin pre-mRNA (see schematic diagram, Fig. 4A and C). For convenience, we refer to the 209 nucleotide exon of the β -globin and the 226 nucleotide exon of the mouse insulin as "full-length", even though there are actually shorter than the bona fide exons 2. The different RNA precursors were gel purified and the splicing reactions were UV irradiated as described in Materials and Methods. When the β -globin pre-mRNA second exon was 24 or 14 nucleotides we detected at least a 20-fold decrease in binding of the cross-linked proteins when compared with the "full-length" pre-mRNA (Fig. 4A, lanes 1 to 4). For the mouse insulin pre-mRNA, truncation of the second exon to 107 nucleotides yielded a 20-fold decrease in the efficiency of transfer of label to proteins by UV light (Fig. 4C, lanes 1 to 3). Both pre-mRNAs, truncated in intron sequences were totally inefficient for the binding of those proteins (Fig. 4A, lane 5 and 4C, lane 4). These results show that the second exon regions, located between nucleotides 24 to 53 for the β -globin pre-mRNA and between nucleotides 107 to 148 for the mouse insulin pre-mRNA, are necessary for an efficient UV cross-linking.

To determine if the length of the second exon affect the efficiency of UV cross-linking, pre-mRNAs were transcribed from a β -globin DNA which have the second exon substituted at 14 nucleotides downstream the 3' splice site, with sequences from the plasmid SP64. These pre-mRNAs were able to restore the binding of the cross-linked proteins (Fig. 4B, lanes 1-3). These results indicate that rather than specific sequences, the length of the β -globin second exon influences the efficiency of transfer of label to proteins by UV light.

The role of the β -globin RNA exon 1 (5') sequences was investigated with pre-mRNAs containing only 21 nucleotides upstream the 5' splice site. Figure 4B (lane 4) shows that these pre-mRNAs were able to cross-linked proteins with one to two-fold the efficiency of the "full-length" pre-mRNA (compare lanes 1 and 4). We have observed

Figure 5: Analysis of the oligonucleotides from the mouse insulin pre-mRNA cross-linked to proteins in an *in vitro* splicing reaction.

A: Splicing reactions containing mouse insulin pre-mRNAs were incubated for 10 min. or 90 min. at 30°C, UV-irradiated and immunoselected with 4F4 monoclonal antibodies. The immunoselected cross-linked proteins were analyzed on SDS-PAGE as described in Fig. 3 and Materials and Methods. The 44-42 Kd proteins were excised and eluted from the gel by diffusion overnight at room temperature in 10 mM sodium phosphate pH 6.8, 0.2% SDS. The proteins were then digested for 30 min. at 37°C with Proteinase K (200 µg/ml) in the presence of 0.5% SDS and 10 mM EDTA. The samples were then phenol-chloroform extracted and the nucleic acids precipitated with ethanol. The resistant oligonucleotides were resolved on 8M urea-20% polyacrylamide gels. Lane 1: [³²P] synthetic 17-mer deoxy-oligonucleotide used as a marker. Lane 2: ribo-oligonucleotides cross-linked to proteins from reactions incubated for 10 min. at 30°C. Lane 3: ribo-oligonucleotides cross-linked to proteins from reactions incubated for 90 min. at 30°C. Lane 4: ³²P-labeled SP6/mouse insulin pre-mRNA digested with ribonuclease T1 used as markers.

B: The 17-mer oligonucleotides shown in panel A lane 2 was eluted from the gel, dephosphorylated and ³²P 3'-end labeled with T₄ RNA ligase. Equal aliquots were treated with alkali (lane OH⁻, partial alkaline hydrolysis produced a "ladder" showing the position of all the oligonucleotides), ribonuclease PHyM (lane PHY, cleaves after A and U), U2 (lane U2, cleaves after A), T1 (lane T1, cleaves after G) and B (lane B, cleaves after C and U).

C: The 14-mer oligonucleotide shown in panel A, lane 2 was sequenced as described in B.

D: A schematic representation of the SP6/mouse insulin pre-mRNA and the positions for the 17-mer and 14-mer oligonucleotides cross-linked to proteins in an *in vitro* splicing reaction. The 14-mer is located in the intron at 25 nucleotides from the donor splice site and the 17-mer in the exon 2 at 134 nucleotides from the acceptor splice site.

slight variations in the efficiency of cross-linking to the β -globin pre-mRNA for the 120 KD protein, (Fig. 4B right, lane 1). It should be noted that the autoradiogram shown in Figure 4B (lanes 1-6) was from another gel than the experiment in lanes 1-3. When these pre-mRNAs were truncated at 24 nucleotides in the second exon (Fig. 4B, lane 5) or in intron sequences (Fig. 4B, lane 6) the cross-linked proteins were not detected, like with "wild type" pre-mRNAs (Fig. 4A, lanes 3 and 5). Thus, sequences of the β -globin exon 1, located in the first 130 nucleotides moderately affect the efficiency of UV cross-linking.

To investigate the effect of intron sequences on the binding efficiency of the cross-linked proteins a mouse insulin pre-mRNA, containing "full length" exon 1 and 2 but lacking introns, was incubated in a splicing reaction (see Materials and Methods). The RNA transcript lacking introns failed to show any cross-linked protein (Fig. 4C, lane 5). Taken together, these results indicate that the efficiency of UV cross-linking depends on the length of the second exon and the presence of intron sequences, and not on specific sequences. All the RNA transcripts for the experiment in figure 4 were stable after incubation in splicing reactions and UV cross-linking (data not shown).

Nucleotide Sequence of the pre-mRNA regions that cross-linked to the hnRNP C proteins.

To determine the contact regions of the mouse insulin pre-mRNA with the hnRNP C proteins, the UV cross-linked proteins were immunoprecipitated with 4F4 monoclonal antibodies. The proteins migrating as 44 and 42 KD (see Fig. 3A, lane 7) were eluted from the gel, digested with Proteinase K and the resistant oligonucleotides resolved on sequencing gels (Fig. 5). A 17-mer and a 14-mer oligonucleotides were detected from splicing reactions incubated either for 10 or 90 min (Fig. 5A, lanes 2 and 3). The oligonucleotides were eluted, 3'-end labeled with ^{32}P and subjected to partial digestions with ribonucleases (PhyM, U_2 , T_1 and B). Figure 5B and C show the nucleotide sequence of the cross-linked oligonucleotides. Figure 5D shows the two regions protected, a 17-mer in the exon 2 (located 134 nucleotides downstream the 3' splice site) and a 14-mer in the intron (located 25 nucleotides downstream the 5' splice site).

The same nucleotide sequence analysis was performed for the oligonucleotides from the β -globin pre-mRNA cross-linked to the immunoselected hnRNP C proteins (44-42 KD). A 13-mer and an 8-mer oligonucleotides were detected (Fig. 6A and B). Figure 6C shows the two regions protected in the β -globin pre-mRNA, a 13-mer in the exon 2 (located 28 nucleotides downstream the 3' splice site) and an 8-mer in the exon 1 (located 81 nucleotides downstream the 5'-end of the pre-mRNA).

Splicing efficiency of the β -globin and mouse insulin pre-mRNA truncated in the second exon

The results presented in Figure 4 indicate that the binding efficiency of the cross-linked proteins decreased 20-fold when the β -globin or mouse insulin pre-mRNAs were

Figure 6: Analysis of the oligonucleotides from the β -globin pre-mRNA cross-linked to proteins in an *in vitro* splicing reaction. The preparation of the cross-linked samples and purification and sequence of the oligonucleotides were as described in Figure 5, except that β -globin pre-mRNA was used during the splicing reactions.

A. Nucleotide sequence of the 8-mer oligonucleotide.
B. Nucleotide sequence of the 13-mer oligonucleotide.
C. A schematic representation of the β -globin pre-mRNA and the locations of the 13-mer and 8-mer oligonucleotides. The 13-mer is located in the second exon, 28 nucleotides downstream the 3' splice site and the 8-mer in the first exon, 81 nucleotides from the 5'-end of the pre-mRNA.

truncated in the second exon at 24 or 107 nucleotides, respectively. It was then of interest to investigate the splicing efficiency of those pre-mRNAs. To calculate the splicing efficiency, the different RNA bands shown in Figure 7 were excised from the gel and cerenkov counted. The cpm of the spliced RNAs (band 3 in Fig. 7A, B and C), were divided over the input cpm of the reactions. Figure 7 shows that the β -globin and mouse insulin transcripts terminating at 24 and 107 nucleotides, respectively, downstream from the 3' splice site were spliced with a lower efficiency (20-fold decreased) than the "full-length" pre-mRNA's (for the mouse insulin compare Fig. 7A lane 2 and 7B lane 1, and, for the β -globin compare Fig. 7C lanes 1 and 5). The same result with truncated β -globin pre-mRNAs was reported by Furdon and Kole (18). Thus, the lower splicing efficiency of the truncated β -globin or mouse insulin pre-mRNAs correlated with the

Figure 7: *In vitro* splicing of the truncated SP6/mouse insulin and β -globin pre-mRNAs. 32 P-labeled RNA was transcribed and processed as described in Materials and Methods and analyzed on a 5% polyacrylamide sequencing gel.

A. The SP6/mouse insulin DNA clone was linearized with Eco RI (412 nucleotides, see Fig. 4C, lane 1 of schematic diagram), transcribed and incubated in splicing reactions for 60 min. (lane 2) or 90 min. (lane 3). Lane 1, pre-mRNA incubated for 90 min. minus extract.

B. Pre-mRNAs transcribed from the SP6/mouse insulin plasmid DNA restricted at the *Ava*II site (293 nucleotides, see Fig. 4C, lane 3 of schematic diagram) incubated in splicing reactions for 60 min. (lane 1) or 90 min. (lane 2). Pre-mRNAs transcribed from a SP6/mouse insulin DNA restricted at the *Mbo*II site (334 nucleotides, see Fig. 4C, lane 2 of schematic diagram) were incubated for 60 min. (lane 4) or 90 min. (lane 3). Lane 5, the 334 nucleotide pre-mRNA incubated for 90 min. minus extract.

C. Pre-mRNAs transcribed from the SP6/ β -globin DNA clone restricted at the *Bam*HI site (493 nucleotides, see Fig. 4A, lane 1 of schematic diagram) were incubated in splicing reactions for 60 min. (lane 1) and 90 min. (lane 2). The SP6/ β -globin DNA was restricted at the *Sau*3A site (337 nucleotides, see Fig. 4A, lane 2 of schematic diagram) and the pre-mRNAs incubated in splicing reactions for 60 min. (lane 8) and 90 min. (lane 7). Lane 6, pre-mRNA incubated for 90 min. minus extract. The SP6/globin DNA restricted at the *Ava*II site (308 nucleotides, see Fig. 4C, lane 3 of schematic diagram) was incubated in splicing reactions for 60 min. (lane 4) or 90 min. (lane 5). Lane 3; pre-mRNA (308 nucleotides) incubated for 90 min. minus extract.

The structure of the RNA in bands 1-5 are indicated on the right of panel A. The numbers on the right and left of panels B and C indicate the expected products and intermediates of the splicing reaction. To calculate the splicing efficiency the different bands were excised from the gel and Cerenkov counted.

lower binding efficiency of the cross-linked proteins. Furthermore, the same β -globin pre-mRNAs used in this study containing substitutions with SP64 sequences in the second exon were reported to restore the splicing efficiency (19) and the binding of the cross-linked proteins (Fig. 4), when compared with truncated pre-mRNAs. Deletions in the first β -globin exon have no effect on either the splicing efficiency (19) or the binding of cross-linked proteins (Fig. 4). Taken together these data suggests a correlation between the binding efficiency of cross-linked proteins and splicing efficiency.

DISCUSSION

The photocrosslinking experiments presented here identify and characterize some of the RNP proteins which are in direct contact with the pre-mRNA during in vitro splicing. Other RNP proteins, however, might exist in direct contact with the pre-mRNA that were not detected in these studies. The major cross-linked proteins, 44, 42 KD belong to the class of hnRNP C proteins since they were immunoprecipitated by monoclonal antibodies which recognizes antigenic sites on the hnRNP C proteins. These proteins have been shown to be involved in pre-mRNA splicing (13). Although we have not characterized the other major cross-linked proteins (120,55,39 and 38 KD), it should be noted that their molecular weights corresponded to the reported for the other hnRNP proteins (2,15). Preliminary experiments with polyclonal antibodies against hnRNP proteins (gift from W. Szer, 35) showed that those antibodies were able to immunoprecipitate all of the major cross-linked proteins (data not shown). It is likely then that the major proteins we have detected tightly associated with the pre-mRNA in in vitro splicing reactions belong to the class of hnRNP proteins.

The molecular weights of the C hnRNP proteins have been reported as 41 KD for the C₁ and 42 KD for the C₂ (2,15). Our results show that the cross-linked C hnRNP proteins migrated slower in SDS-PAGE (44, 42 KD) when compared with the uncross-linked C hnRNP proteins (41-42 KD). This is in agreement with the reported molecular weights for the in vivo cross-linked hnRNP proteins (15).

Mayrand et al. (29) recently reported UV cross-linked proteins in the range of 70 KD to 13 KD, during in vitro splicing reactions, that were selected with monoclonal antibodies against hnRNP core proteins. Individual hnRNP proteins, however, were not identified. They also showed that anti-Sm monoclonal antibodies directed against Sm snRNP proteins immunoselected proteins in the range of 46 KD to 10 KD (29). We have not detected UV cross-linked proteins that reacted with monoclonal anti-Sm antibodies. This discrepancy could be due to differences in UV irradiation protocols and/or that in our experiments the non-crosslinked pre-mRNA was digested with pancreatic RNase and micrococcal nuclease before the immunoprecipitations. By UV-irradiation of intact cells Dreyfus et al. (15) reported that only hnRNP proteins cross-linked to polyadenylated

hnRNA, Sm proteins were not detected in that study. Those conclusions are in agreement with the results presented here.

The time course of pre-mRNA/protein cross-linking during the *in vitro* splicing reaction indicates that a maximum amount was detected at 20 min. of incubation and it decreased with longer incubation times (see Fig. 1). The size of the cross-linked oligonucleotides, however, did not change with incubation times as long as 90 min. (see Fig. 5), suggesting that the protein-RNA contacts remained the same throughout the splicing reaction. The binding of the cross-linked protein decreased at 180 min of incubation, correlating with the reported (17,21,22) instability of the 60S (spliceosome) complex at longer incubation times. These results are consistent with the notion that the hnRNP proteins bind to the pre-mRNA and then they slowly dissociate from the pre-mRNA. The movement of hnRNP proteins has been previously suggested by electron microscopy studies of hnRNPs isolated from nuclei (3,4,31).

The transfer of labeling from RNA to proteins by UV light required Mg^{2+} and ATP (see Fig. 2). It is possible that ATP binding sites are present in the proteins itself or alternatively interaction of the pre-mRNA with other factors which requires ATP, are necessary for binding of the hnRNP proteins. Interestingly, an RNA helicase activity involved in the splicing reaction was inhibited by the monoclonal anti-hnRNP C antibodies and requires ATP (37). The unwinding of the pre-mRNA secondary structure could be required for or coupled to an efficient binding of the hnRNP proteins to the RNA.

We have purified and sequenced two oligonucleotides, a 17-mer and a 14-mer, from the 44-42 KD hnRNP C proteins cross-linked to the mouse insulin pre-mRNA. The 17-mer was in the second exon (134 nucleotides downstream the 3' splice site) and the 14-mer in the intron (25 nucleotides downstream the 5' splice site). The β -globin pre-mRNA produced also two oligonucleotides cross-linked to the hnRNP C proteins. A 13-mer located in the second exon (28 nucleotides downstream the 3' splice site) and an 8-mer located in the first exon (81 nucleotides downstream the 5'-end of the pre-mRNA). We cannot rule out, however, that other oligonucleotides from other regions in the pre-mRNA might exist and were not detected in our analysis. Since we have analyzed the oligonucleotides cross-linked to both hnRNP C proteins, we do not know if each of the C₁ and C₂ proteins cross-linked either to both or only one of the oligonucleotides. The increase in molecular weight observed for the cross-linked C protein (1KD) however, suggests that each of the C proteins might be in contact with only one oligonucleotide.

We have shown that the hnRNP proteins interact with specific sequences of the pre-mRNA. The UV cross-linking efficiency, however, was not determined by those sequences; it was rather dependent on the length of the pre-mRNA second exon and the presence of intron sequences. Thus, a distinct possibility is that these proteins recognize

first secondary structural features of the introns and afterwards scan the length of the pre-mRNA toward exon sequences. Our results demonstrated that the oligonucleotides cross-linked to the β -globin pre-mRNA located in the first or second exon could be either deleted or substituted with unrelated sequences without affecting the efficiency of UV cross-linking. Furthermore, comparisons of the oligonucleotide sequences from the same or different pre-mRNAs did not show any homology. These results suggest that the secondary and/or tertiary structure of the pre-mRNA and not specific sequences might be the main feature for the binding efficiency of the hnRNP proteins.

We have shown that transcripts truncated in the second exon, 24 nucleotides for the β -globin or 107 nucleotides for the mouse insulin, which bound cross-linked proteins very inefficiently, were spliced in vitro with a lower efficiency than the "full length" pre-mRNAs. Although we can not directly relate the lower binding efficiency for hnRNP proteins of the truncated β -globin or mouse insulin pre-mRNAs as the cause for the decrease in splicing efficiency, it is possible that it might be one of the major contributing factors. Furdon and Kole (19) using the same β -globin transcripts used here, containing substitutions in the second exon or deletions in the first exon, showed that the splicing efficiency in vitro was the same as for the "full-length" pre-mRNA. Our results indicate that those transcripts were also able to bind hnRNP proteins as well as the "full-length" pre-mRNA. These results suggest a strong correlation between the binding of the hnRNP proteins and the efficiency of in vitro splicing.

ACKNOWLEDGEMENTS

We thank Dr. R. Kole for gifts of the plasmids HBPL and HBHaeIII. We are also grateful to Dr. Gideon Dreyfuss for the monoclonal (4F4) anti-C antibodies. We also thank Delia Gutman and Argelia Gatell for excellent technical assistance. This work was supported by research grants from the National Institute of Health (NIH AI-24479).

REFERENCES

1. Ahlquist, P., Dasgupta, R., Shih, D.S., Zimmermann, D., and Koesberg, P. (1979) *Nature* 281, 277-282.
2. Beyer, A.L., Christensen, M.E., Walker, B.W., and LeSturgeon, W.M. (1977) *Cell* 11, 127-138.
3. Beyer, A.L., Miller, O.L., and McKnight, S.L. (1980) *Cell* 20, 75-84.
4. Beyer, A.L., Bouton, A.H., and Miller, Jr., O.I. (1981) *Cell* 26, 155-165.
5. Bindereif, A., and Green, M.R. (1986) *Mol. Cell. Biol.* 6, 2582-2592.
6. Black, D.L., Chabot B., and Steitz, J.A. (1985) *Cell* 42, 737-750.
7. Black, D.L., and Steitz, J.A. (1986) *Cell* 46, 697-704.
8. Bradford, M.M. (1976) *Anal. Biochem.* 72, 248-254.
9. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
10. Chabot, B., Black, D.L., LeMaster, D.M. and Steitz, J.A. (1985) *Science* 230, 1344-1349.
11. Choi, Y.D. and Dreyfuss, G. (1984) *Proc. Natl. Acad. Sci. USA* 81, 7471-7475.
12. Choi, Y.D. and Dreyfuss, G. (1984) *J. Cell Biol.* 99, 1997-2004.
13. Choi, Y.D., Grabowski, P.J. and Sharp, P.A. (1986) *Science* 231, 1534-1539.

14. DiMaria, P.R., Kaltwasser, G., and Goldenberg, C.J. (1985) *J. Biol. Chem.* 260, 1096-1102.
15. Dreyfuss, G., Choi, Y.D., and Adams, S. (1984) *Mol. Cell. Biol.* 4, 1104-1114.
16. Fradin, A., Jove, R., Hemenway, C., Keiser, H.D., Manley, J.C. and Prives, C. (1984) *Cell* 37, 927-936.
17. Friendewey, D., and Keller, W. (1985) *Cell* 42, 355-367.
18. Furdon, P.J., and Kole, R. (1986) *Proc. Natl. Acad. Sci. USA* 83, 927-931.
19. Furdon, P.J. and Kole, R. (1988) *Mol. Cell. Biol.* in press.
20. Gerke, V., and Steitz, J.A. (1986) *Cell* 47, 973-984.
21. Grabowski, P.J., Seiler, S.R., and Sharp, P.A. (1985) *Cell* 42, 345-353.
22. Kaltwasser, G., Spitzer, S.G., and Goldenberg, C.J. (1986) *Nucleic Acids Res.* 14, 3687-3701.
23. Konarska, M.M. and Sharp, P.A. (1987) *Cell* 49, 763-774.
24. Knowler, J.T. (1983) *Int. Rev. Cytol.* 84, 103-153.
25. Krainer, A.R., and Maniatis, T. (1985) *Cell* 42, 725-736.
26. Kramer, A., Keller, W., Appel, B. and Luhrman, R. (1984) *Cell* 38, 299-307.
27. Krupp, G., and Gross, H.J. (1979) *Nucleic Acids Res* 6, 3481-3490.
28. Martin, T.E., Pullman, J.M., and McMullen, M.E. (1980) In *Cell Biology: A comprehensive Treatise*, eds. Prescott, D.M. and Goldstein, L. (Academic Press, Inc., New York) 4, 137-174.
29. Mayrand, S.H., Pedersen, N. and Pederson, T. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3718-3722.
30. McKnight, S.L. and Miller, O.L. (1976) *Cell* 5, 155-167.
31. Osheim, Y.N., Miller, Jr., O.L., and Beyer, A.L. (1985) *Cell* 43, 143-151.
32. Padgett, R.A., Mount, S.M., Steitz, J.A., and Sharp, P.A. (1983) *Cell* 35, 101-107.
33. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, Sh. and Sharp, P.A. (1986). *Ann. Rev. Biochem.* 55, 1119-1150.
34. Perkins, K.K., Furneaux, H.M., and Hurwitz, J. (1986) *Proc. Natl. Acad. Sci. USA* 83, 887-891.
35. SieraKowska, H., Szer, W., Furdon, P.J., and Kole, R. (1986) *Nucleic Acids Res.* 14, 5241-5254.
36. Tazi, J., Alibert, Ch., Tamsamani, J., Reveillaud, I., Cathala, G., Brunel, C., and Jeanteur, P. (1986) *Cell* 47, 755-766.
37. Wollenzien, P.L., Goswami, P., Szeberenyi, J., and Goldenberg, C.J. (1987) *Nucleic Acids Res.* 15, 9279-9298.
38. Yang, V.W., Lerner, M.R., Steitz, J.A., and Flint, S.J. (1981) *Proc. Natl. Acad. Sci. USA* 78, 1371-1375.