A General Approach for Identification of RNA-Protein Cross-linking Sites within Native Human Spliceosomal Small Nuclear Ribonucleoproteins (snRNPs)

ANALYSIS OF RNA-PROTEIN CONTACTS IN NATIVE U1 AND U4/U6.U5 snRNPs*

Henning Urlaub‡, Klaus Hartmuth‡, Susanne Kostka§, Gerlinde Grelle§, and Reinhard Lührmann‡¶

From the ‡Abteilung Zelluläre Biochemie, Max-Planck-Institut für Biopysikalische Chemie, Am Faßberg 11, D-37077 Göttingen, Germany and the §Abteilung Proteinchemie, Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Straße 10, D-13125 Berlin, Germany

We describe a novel approach to identify RNA-protein cross-linking sites within native small nuclear ribonucleoprotein (snRNP) particles from HeLa cells. It combines immunoprecipitation of the UV-irradiated particles under semi-denaturing conditions with primer extension analysis of the cross-linked RNA moiety. In a feasibility study, we initially identified the exact cross-linking sites of the U1 70-kDa (70K) protein in stem-loop I of U1 small nuclear RNA (snRNA) within purified U1 snRNPs and then confirmed the results by a large-scale preparation that allowed N-terminal sequencing and matrix-assisted laser desorption ionization mass spectrometry of purified cross-linked peptide-oligonucleotide complexes. We identified Tyr^{112} and Leu^{175} within the RNA-binding domain of the U1 70K protein to be cross-linked to G^{28} and U^{30} in stem-loop I, respectively. We further applied our immunoprecipitation approach to HeLa U5 snRNP, as part of purified 25 S U4/U6.U5 tri-snRNPs. Cross-linking sites between the U5-specific 220-kDa protein (human homologue of Prp8p) and the U5 snRNA were located at multiple nucleotides within the highly conserved loop 1 and at one site in internal loop 1 of U5 snRNA. The cross-linking of four adjacent nucleotides indicates an extended interaction surface between loop 1 and the 220-kDa protein. In summary, our approach provides a rapid method for identification of RNA-protein contact sites within native snRNP particles as well as other ribonucleoprotein particles.

The spliceosome catalyzes the two-step *trans*-esterification reaction that occurs during splicing of nuclear pre-mRNA, *i.e.* excision of the introns and ligation of the exons. Assembly of the spliceosome is an ordered process and involves the interaction of U1, U2, and U4/U6.U5 small nuclear ribonucleoprotein (snRNP)[1] particles with the intron-containing pre-mRNA

and a multitude of additional spliceosomal factors (for review, see Refs. 1–5). Although catalysis in the spliceosome appears to be RNA-based, both protein-protein and RNA-protein interactions contribute to the formation of its catalytic core (1–5).

The structures of the spliceosomal components are largely solved at the levels of primary structure and RNA secondary structures (3, 5). Much attention is currently being devoted to the issues of protein and RNP tertiary structures (6) and to the quaternary arrangement both of the individual macromolecules in the U snRNP particles and of these particles in the functioning spliceosome (7).

The U1 snRNP particle from HeLa cells has been extensively investigated in terms of RNA-protein and protein-protein interactions, as it is only of moderate complexity. Thus, a relatively detailed picture of its morphology and tertiary structure (in comparison with other spliceosomal RNPs) has emerged (3, 7, 8). The seven common spliceosomal Sm proteins (G, E, F, D1, D2, D3, and B) assemble on the Sm site of the U1 snRNA to form a doughnut-like structure (9, 10). The U1-specific proteins U1 70K and A specifically interact via their N-terminal RNA-binding domains and flanking amino acids with stem-loops I and II, respectively, of U1 snRNA (11–16). The U1 C protein does not bind directly to U1 snRNA; its interaction with the RNP is probably mediated by protein-protein interactions (17). Except for the U1 A protein (18), the snRNA contact sites at the molecular level of all the other U1 snRNPs remain elusive.

Much less is known about RNA-protein interactions within the other spliceosomal snRNPs. Studies have so far focused on the interaction of only a few U snRNP-specific proteins with their cognate U snRNAs, *e.g.* the A′ and B″ proteins with U2 snRNA (19–21), the 15.5-kDa protein with U4 snRNA (22), and the like-Sm (LSm) proteins with the 3′-end of U6 snRNA (23–25). In the absence of co-crystals or binding studies with purified components, cross-linking studies add valuable information about contacts between specific snRNPs and snRNAs in assembled particles. To date, most cross-linking studies have involved the ^{32}P labeling of the corresponding RNA, either in a random or site-specific manner. For example, the U1 70K and Sm G proteins were first identified to contact directly U1 snRNA by UV cross-linking of ^{32}P-labeled U1 snRNP from HeLa cells (26, 27). Within other snRNP particles, site-specific cross-linking also provided the first insights into the snRNA-protein contacts. In this manner, the highly conserved U5 snRNP-specific protein Prp8p (yeast homologue of the human U5 220-kDa protein) has been shown to contact several nucleotides of U5 snRNA and at least one position of U6 snRNA in reconstituted yeast U5 and tri-snRNP particles (28, 29).

¶ To whom correspondence should be addressed. Tel.: 49-551-201-1407; Fax: 49-551-1196; E-mail: reinhard.luehrmann@mpi-bpc.mpg.de.

[1] The abbreviations used are: snRNP, small nuclear ribonucleoprotein; snRNA, small nuclear RNA; 70K, 70-kDa; IL, internal loop; RP-HPLC, reversed phase high performance liquid chromatography; MALDI-MS, matrix-assisted laser desorption ionization mass spectrometry; RBD, RNA-binding domain.

This paper is available on line at http://www.jbc.org

However, no method has yet been described that allows the localization of direct RNA-protein contact sites in purified native particles. Here, we present a general approach that allows the rapid identification of RNA-protein contact sites in purified native human U snRNP particles subsequent to UV cross-linking. It combines immunoprecipitation of cross-linked proteins under semi-denaturing conditions, followed by primer extension analysis of the cross-linked RNA moiety. We initially chose native 12 S U1 snRNP isolated from HeLa cells to test the feasibility of our method since this particle is well characterized (see above). In this manner, we identified the exact cross-linking sites between the U1 70K protein and stem-loop I of U1 snRNA at the molecular level. The approach was then extended to the less well characterized 25 S U4/U6.U5 tri-snRNP particles purified from HeLa cells. We show a direct interaction of the 220-kDa protein with the four adjacent uridines of the conserved loop 1 of U5 snRNA, suggesting an extended interaction surface between these two components. An additional cross-linking site of the 220-kDa protein in the 3′-half of internal loop (IL) 1 indicates that this protein spans the entire 5′-stem of U5 snRNA.

EXPERIMENTAL PROCEDURES

Sample Preparation—U1, U2, and U5 snRNPs and U4/U6.U5 tri-snRNP particles were prepared from HeLa cell nuclear extracts (Computer Cell Culture Co., Mons, Belgium) by immunoaffinity chromatography using the H20 monoclonal antibody as described previously (30). 12 S U1 snRNPs were separated from contaminating U2 snRNPs by Mono Q anion-exchange chromatography (31) and stored at a concentration of 2 mg/ml in Mono Q elution buffer. 25 S U4/U6.U5 tri-snRNP particles were obtained from the total snRNP mixture by gradient centrifugation (30), and peak fractions were pelleted for 6 h at 70,000 rpm (TLA 100.3 rotor, Beckman Instruments). The tri-snRNP pellet was initially redissolved in 20 mM HEPES-KOH (pH 7.9), 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM dithioerythritol, and 250 mM NaCl (buffer A) at a concentration of 1 mg/ml.

UV Cross-linking of Native U snRNPs and Isolated U snRNAs—Two 25-μl droplets each containing 2.5 μg of 12 S U1 snRNP or 25 S U4/U6.U5 tri-snRNP particles were irradiated on glass dishes at 254 nm with four 8-watt germicidal lamps (G8T5, Herolab, Wiesloch, Germany) in parallel at a distance of 4 cm for 2 min on ice. 50 μl cross-linked samples were then incubated with 1 mg/ml proteinase K in the presence of 1% (w/v) SDS for 30 min at 37 °C. The snRNAs were extracted twice with phenol/chloroform, precipitated with ethanol in the presence of 20 μg of glycogen, and finally dissolved in 6.5 μl of CE buffer (20 mM cacodylate-KOH (pH 7.0) and 0.2 mM EDTA). For cross-linking experiments using isolated naked snRNAs derived from native 12 S U1 snRNP or 25 S U4/U6.U5 tri-snRNP particles, 5 μg of the corresponding particles were treated with proteinase K, and the snRNAs were extracted and precipitated as described above. The snR-NAs were dissolved in the same volume of buffer as described for the intact particles. UV cross-linking was performed as described above, and cross-linked snRNAs were again precipitated in the presence of 20 μg of glycogen and dissolved in 6.5 μl of CE buffer.

Large-scale Cross-linking of U1 snRNP—Prior to UV cross-linking, 0.8–1 mg of Mono Q-purified U1 snRNPs was diluted with 12.5 ml of C buffer (32) without glycerol. UV cross-linking was carried out in glass dishes with an inner diameter of 12.5 cm on ice for 2 min at 254 nm at a distance of 3–4 cm from the lamps. The snRNP solutions had a depth of ~1 mm in the glass dishes.

Immunoprecipitation of RNA-Protein Cross-links—5 μg of cross-linked or non-cross-linked 12 S U1 snRNPs and 25 S U4/U6.U5 tri-snRNP particles were incubated with 1% (w/v) SDS for 5 min at 70 °C. After cooling to room temperature, Triton X-100 was added to a final concentration of 5% (v/v), and the volume was adjusted to 350 μl with phosphate-buffered saline (NaCl (pH 8.0)). The samples were then incubated for 1 h at 4 °C with different antibodies against U1 or U5 snRNP-specific proteins bound to 10 μl of protein A-Sepharose (Amersham Pharmacia Biotech). After extensively washing with phosphate-buffered saline, the beads were incubated with proteinase K at a concentration of 1 mg/ml in 100 μl of buffer A containing 1% (w/v) SDS for 1 h at 37 °C. Cross-linked snRNA was extracted with phenol/chloroform, precipitated with ethanol as described above, and finally dissolved in 3.5 μl of CE buffer.

The successful identification of specific cross-linking sites (see below) was strictly dependent on the denaturing conditions prior to immunoprecipitation. Changing the immunoprecipitation conditions, *e.g.* dissociation of the particles in the presence of 1% SDS and then dilution to a final concentration of 0.05% SDS, does not allow the detection of specific RNA-protein interactions in cases where the protein itself is involved in strong protein-protein interactions (for example, the U5 220- and 116-kDa proteins; see "Results"). Accordingly, immunoprecipitation with an antibody against the U5 116-kDa protein under such conditions coprecipitates the U5 220-kDa proteins (data not shown).

Primer Extension Analysis of Cross-linked U1 and U5 snRNAs—Primer extension analysis of U1 and U5 snRNAs derived from cross-linked 12 S U1 snRNP and 25 S U4/U6.U5 tri-snRNP particles, respectively, or obtained after immunoprecipitation was performed with a 5′-³²P-labeled oligonucleotide complementary to nucleotides 63–77 of U1 snRNA and nucleotides 83–103 of U5 snRNA. Primer extension analysis using 2 units of avian myeloblastosis virus reverse transcriptase (Seikagaku) per reaction was carried out as essentially described (33). The reaction was stopped by adding loading buffer (50% (w/v) urea, 0.5× Tris borate/EDTA (pH 8.3), 0.025% (w/v) bromphenol blue, and 0.025% (w/v) xylene cyanol) and boiling for 3 min. Sequencing ladders were generated from *in vitro* U1 and U5 snRNA transcripts (Ribomax, Promega) using 0.5 mM dideoxy-NTPs under identical conditions. Transcripts were separated on a 50% (w/v) urea and 9.6% polyacrylamide (20:1) sequencing gel in 1× Tris borate/EDTA (pH 8.3) and exposed to Biomax film (Eastman Kodak Co.) with two intensifying screens for 1 day (for Figs. 2*A* and 8*A*) or 6 days (for Figs. 2*B* and 8*B*).

Isolation and Sequence Determination of Cross-linked Peptide-Oligonucleotide Complexes—UV-irradiated U1 snRNPs were precipitated with 3 volumes of ethanol in Corex tubes (Sorvall) and washed with 80% ethanol, and the pellet was briefly air-dried. U1 snRNPs were dissolved in 8 M urea, 50 mM Tris-HCl (pH 7.5), and 5 mM dithioerythritol; heated for 5 min at 90 °C; and allowed to cool to room temperature. The U1 snRNP solution was diluted with 50 mM Tris-HCl (pH 7.5) and 1 mM CaCl₂ until the urea concentration was below 1 M. Endoproteolytic cleavage of spliceosomal proteins within the U1 snRNP solution was carried out with trypsin (EC 3.4.21.4; Promega) or chymotrypsin (EC 3.4.21.1; Roche Molecular Biochemicals) with an enzyme/substrate ratio of 1:20 for 16 h at 37 °C. The U1 snRNP particles were then precipitated with 3 volumes of ethanol; washed with 80% ethanol; air-dried; and dissolved in an appropriate volume of 20 mM Tris-HCl (pH 7.8), 150 mM NaCl, and 5 mM EDTA containing 8 M urea. The samples were injected onto a Superose 12 size-exclusion column (10/30, Amersham Pharmacia Biotech) running with the same buffer conditions at a flow rate of 400 μl/min. The absorbance was monitored at 254 nm. An aliquot of each fraction was precipitated with 5 volumes of acetone and analyzed by 13% SDS-polyacrylamide gel electrophoresis and subsequent silver staining. snRNA-containing fractions were pooled and precipitated with 3 volumes of ethanol, washed with 80% ethanol, and dissolved in 25 mM Tris-HCl (pH 7.8) and 2.5 mM EDTA. snRNAs were digested with 10 μg of RNase A or 10 μg of RNase T1 (Ambion, Inc.) for 2 h at 50 °C. The snRNA oligonucleotide pool was subjected to a second incubation with 1 μg of the respective endoproteinases for 16 h at 37 °C. Digestion was stopped by injection onto an RP-HPLC C₁₈ column (Vydac, Hesperia, CA) running at 5% solvent B. Solvent A was water and 0.1% trifluoroacetic acid, and solvent B was acetonitrile and 0.085% trifluoroacetic acid. Isolation of the cross-linked peptide-oligonucleotide heteromers was achieved at a flow rate of 100 μl/min with the following gradients: first, isocratic elution at 5% solvent B for 30 min, then 5–45% solvent B for 120 min, and finally 45–90% solvent B for 20 min. An additional isocratic step that varied in time was performed at 10% solvent B until the base line was reached. The absorbance was monitored at 220 and 260 nm. Fractions with absorbances at both 220 and 260 nm were dried under vacuum and subjected to automated N-terminal sequence analysis in a PROCISE™ protein sequencer (Applied Biosystems Inc., Foster City, CA) and to MALDI-MS analysis. MALDI-MS of the cross-linked peptide-oligonucleotide complexes was performed essentially under conditions described previously (34, 35), but without an ion-exchange procedure before measurement.

RESULTS

Cross-link Identification Strategy—Our method for the identification of RNA-protein contact sites in native snRNP particles from HeLa cell nuclear extracts combines (i) UV cross-linking, (ii) immunoprecipitation of cross-linked proteins under semi-denaturing conditions, and (iii) primer extension analysis

FIG. 1. *A*, primer extension analysis for detection of RNA-protein cross-links in RNP particles as illustrated for U1 snRNPs. UV-cross-linked snRNPs were dissociated; the protein moiety was digested with proteinase K; and snRNAs were extracted. RNA-protein cross-linking sites were then identified by primer extension, as a few amino acids remain covalently attached to the cross-linked bases on the RNA. The actual cross-linking site is located one nucleotide upstream from the site of the stop of the reverse transcriptase. See "Experimental Procedures" and "Results" for further details. *B*, overall strategy for the identification of a single RNA-protein cross-linking site in UV-irradiated RNP particles as illustrated for U1 snRNPs. Immunoprecipitation with antibodies against U1 snRNP-specific proteins (*e.g.* the U1 70K protein) performed under such conditions where the UV-cross-linked particle was completely dissociated (1% SDS and 5% Triton X-100; see "Experimental Procedures") led to the isolation of a single protein together with its cross-linked RNA component. Following digestion of the precipitated protein moiety with proteinase K, the RNA was extracted; and subsequently, a primer extension analysis should reveal the exact cross-link site on the RNA corresponding to the precipitated protein. See "Experimental Procedures" and "Results" for further details.
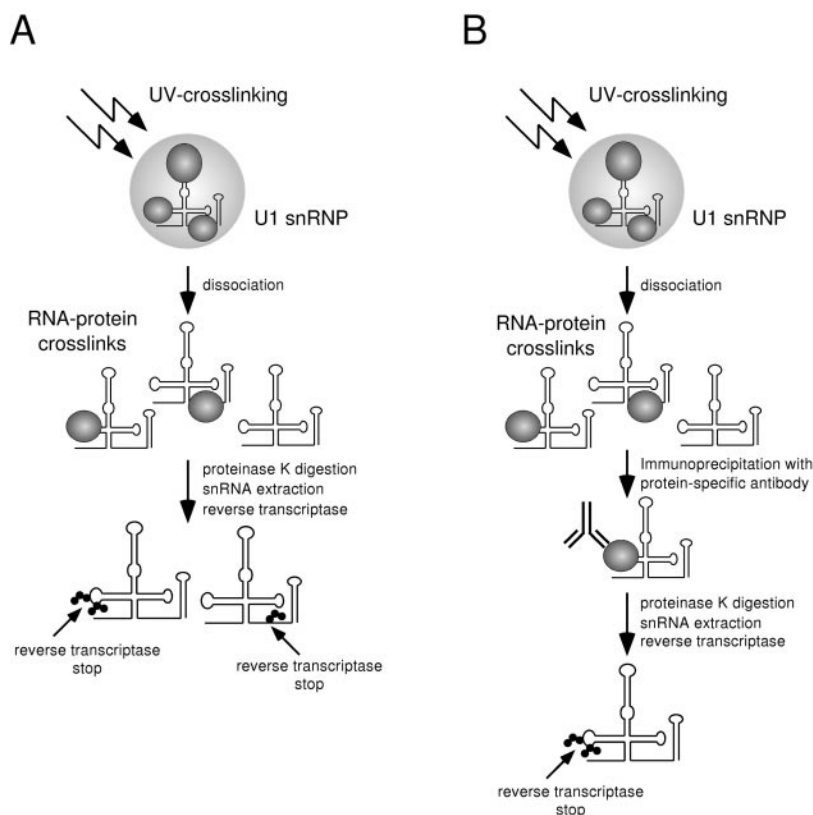
of the cross-linked snRNA moiety in the precipitate. The overall strategy is schematically outlined in Fig. 1 for U1 snRNP. UV cross-linking at 254 nm of an RNP particle generates a direct cross-link between a specific protein and an RNA base whenever the two components occupy favorable relative positions. In the first experiment (Fig. 1*A*), dissociation of the cross-linked particle and subsequent digestion with proteinase K are followed by identification of the exact cross-linking site by primer extension analysis. Cross-linked nucleotides are detected as discrete reverse transcriptase stops because a few amino acid residues of the cross-linked protein always remain covalently attached to the base of the RNA. The actual cross-linking site is interpreted as occurring one nucleotide upstream from the stop site of the reverse transcriptase. As an initial approach to distinguish between stops due to intra- and inter-RNA cross-links or UV-induced RNA strand breaks and those stops that are due to a cross-linked protein, cross-linking experiments were carried out with naked RNAs isolated from purified RNP particles. Comparison of the reverse transcriptase patterns of the two experiments allowed the identification of putative RNA-protein cross-linking sites on the RNA.

To identify the snRNP-specific proteins cross-linked to the nucleotides identified in the first experiment, we combined immunoprecipitation of a single denatured cross-linked RNA-protein product with primer extension analysis of the cross-linked RNA moiety. The approach is outlined schematically in Fig. 1*B*. After UV irradiation at 254 nm, the snRNP particles were dissociated in the presence of 1% SDS at 70 °C and cooled to room temperature, and then Triton X-100 was added to a final concentration of 5% (36). Under such conditions, where the protein-protein interactions are completely disrupted, the antibody should precipitate only a single snRNP-specific protein with the cross-linked snRNA. After digestion of the immunoprecipitated protein with proteinase K, the RNA was analyzed by primer extension analysis as described above to identify the exact cross-linking sites on the RNA corresponding

to the precipitated protein. Non-cross-linked RNA as well as other proteins should not be coprecipitated (Fig. 1). Ideally, in the primer extension analysis of the immunoprecipitated sample, no full-length RNA should be detected. However, in practice, the bands due to the stops at the cross-linking sites of the precipitated protein have intensities significantly greater than that of the background.

Identification of Cross-linking Sites in Stem-loop I of U1 snRNA for the U1 70K Protein by Immunoprecipitation Combined with Primer Extension Analysis—We used HeLa U1 snRNPs as a test system for investigating the RNA-protein interactions in native snRNPs. This HeLa snRNP was chosen for the following reasons. (i) U1 snRNP is the best characterized particle in terms of RNA-protein interactions among all human snRNPs. For example, the U1 70K protein can be cross-linked to U1 snRNA with high yield (26, 27). (ii) Although binding studies have shown that the U1 70K protein specifically interacts with stem-loop I of U1 snRNA (11, 12), the exact cross-linking sites have not yet been identified. Our approach thus provides an opportunity to identify these sites exactly and to test the results by seeing whether the U1 70K protein cross-linking site(s) are congruent with the interaction site in stem-loop I. Therefore, we chose a cDNA primer complementary to positions on the 3'-side of a putative cross-linking site, namely positions 63–87 of stem-loop II.

Fig. 2*A* shows the primer extension analysis of UV-irradiated naked U1 snRNA isolated from U1 snRNP by phenol extraction (*lanes 1* and *2*) in comparison with U1 snRNA isolated from UV-cross-linked U1 snRNP after treatment with proteinase K (*lanes 3* and *4*). Natural, strong reverse transcriptase stops on U1 snRNA occur at U^{45} to G^{38} within stem-loop I, as these stops are present irrespective of irradiation (compare *lanes 1–4*). These stops were not investigated in detail. Strong reverse transcriptase stops at C^{31} to A^{26} in stem-loop I were observed only on U1 snRNA isolated from UV-cross-linked U1 snRNPs, but not on irradiated naked U1
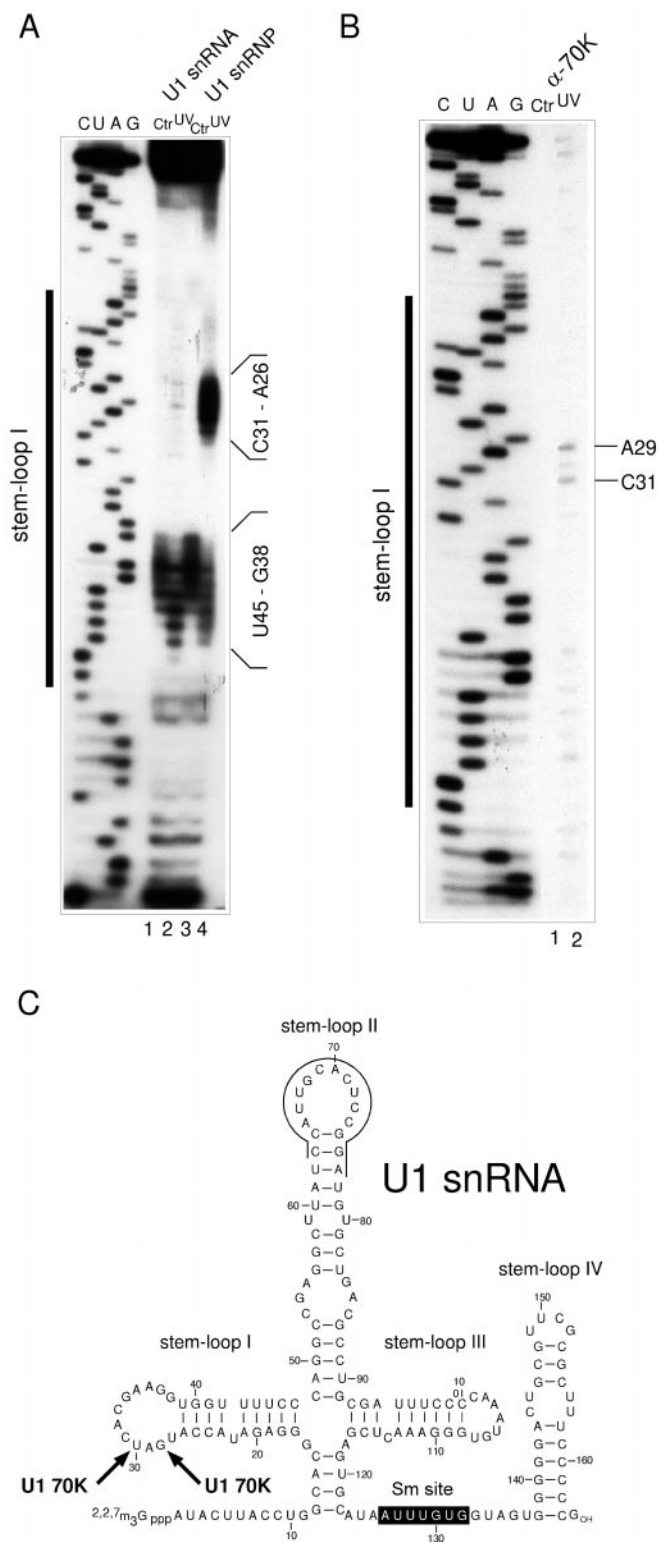
FIG. 2. *A*, primer extension analysis of U1 snRNA derived from UV-cross-linked U1 snRNPs (*lanes 3* and *4*) compared with UV-irradiated naked U1 snRNA (*lanes 1* and *2*). The cDNA primer used was complementary to nucleotides 63–77 in stem-loop II of U1 snRNA (indicated in *C*). *Lanes 1* and *3*, no UV irradiation (control (*Ctr*)); *lanes 2* and *4*, 2 min of UV irradiation at 254 nm. *C, U, A,* and *G* indicate dideoxy sequence markers. Stops of the reverse transcriptase during primer extension analysis are denoted on the right (U^{45} to G^{38} and C^{31} to A^{26}). The *black bar* on the left indicates stem-loop I of U1 snRNA. *B*, primer extension analysis of cross-linked U1 snRNA after immunoprecipitation of cross-linked U1 snRNPs with anti-70K protein antibody (*α-70K*). Prior to immunoprecipitation, cross-linked U1 snRNPs were dissociated with the SDS/Triton X-100 procedure (see "Experimental Procedures"). *Lane 1*, no UV irradiation (*Ctr*); *lane 2*, 2 min of UV irradiation at 254

snRNA (Fig. 2*A*, compare *lane 4* with *lane 2*), indicating that nucleotides one position 5′ to the stop sites (see above) must be cross-linked to a U1 snRNP in native particles.

To determine whether the reverse transcriptase stops observed in the initial experiment (Fig. 2*A*, *lane 4*) were indeed due to a cross-link with the U1 70K protein (as expected for those reasons stated above), we employed the immunoprecipitation procedure outlined above using a monoclonal antibody (H111) (8) against the U1 70K protein. Fig. 2*B* shows the primer extension analysis of U1 snRNA after UV cross-linking of the U1 snRNP particles and immunoprecipitation with the anti-70K protein antibody in the presence of SDS/Triton X-100. Strikingly, strong reverse transcriptase stops were observed only at A^{29} and C^{31} of stem-loop I of U1 snRNA in the UV-irradiated sample. Both stops are significantly enriched (Fig. 2, compare *A* (*lane 4*) with *B* (*lane 2*)) and are located at positions of U1 snRNA at which RNA-protein cross-links were detected in the initial experiment (C^{31} to A^{26}) (Fig. 2*A*). In control experiments carried out with antibodies specific for other U1 snRNPs such as the U1 C protein or the Sm proteins, no reverse transcriptase stops in stem-loop I were detected (data not shown). We thus conclude that G^{28} and U^{30}, located one nucleotide on the 5′-side of the stops at A^{29} and C^{31}, respectively, are two independent cross-linking sites for the U1 70K protein (Fig. 2*C*).

N-terminal Sequencing and MALDI-MS Analysis of U1 70K Peptides Cross-linked to Stem-loop I of U1 snRNA—The relatively large amount of U1 snRNP purified by anion-exchange chromatography (31) enabled us to verify our immunoprecipitation results by an independent method, *i.e.* by isolation and sequencing of cross-linked U1 70K peptide-oligonucleotide complexes. A similar approach was recently described for several ribosomal proteins isolated from UV-irradiated ribosomal subunits (34, 35, 37). Fig. 3 shows the purification strategy for the isolation of cross-linked peptides from U1 snRNPs. Purified 12 S U1 snRNPs were UV-irradiated at 254 nm, dissociated in the presence of 8 M urea, and digested with various endoproteinases (see "Experimental Procedures"). Cross-linked peptide-snRNA complexes were enriched by size-exclusion chromatography, and the snRNA thus isolated was digested with ribonucleases T1 and/or A. Cross-linked peptide-oligonucleotide complexes were then separated by RP-HPLC. Peak fractions eluting from the RP-HPLC column that showed a strong absorbance at both 220 and 260 nm are good candidates for peptides (220 nm) cross-linked to oligonucleotides (260 nm) (37). Each peak was collected and subjected to automated N-terminal sequencing and MALDI-MS. N-terminal sequencing revealed the cross-linked amino acid residue of the peptide moiety because a gap is expected to occur at the position of the cross-linked amino acid during analysis of the Edman degradation products (37–39). In addition, MALDI-MS analysis of the cross-linked peptide-oligonucleotide complex allows the identification of the cross-linked oligonucleotide (34, 35).

In this manner, we identified two different peptide stretches of the U1 70K protein cross-linked to U1 snRNA oligonucleotides after digestion of the native UV-irradiated U1 snRNPs with trypsin and ribonuclease T1. The two peptides coeluted within the same fractions upon RP-HPLC (data not shown).
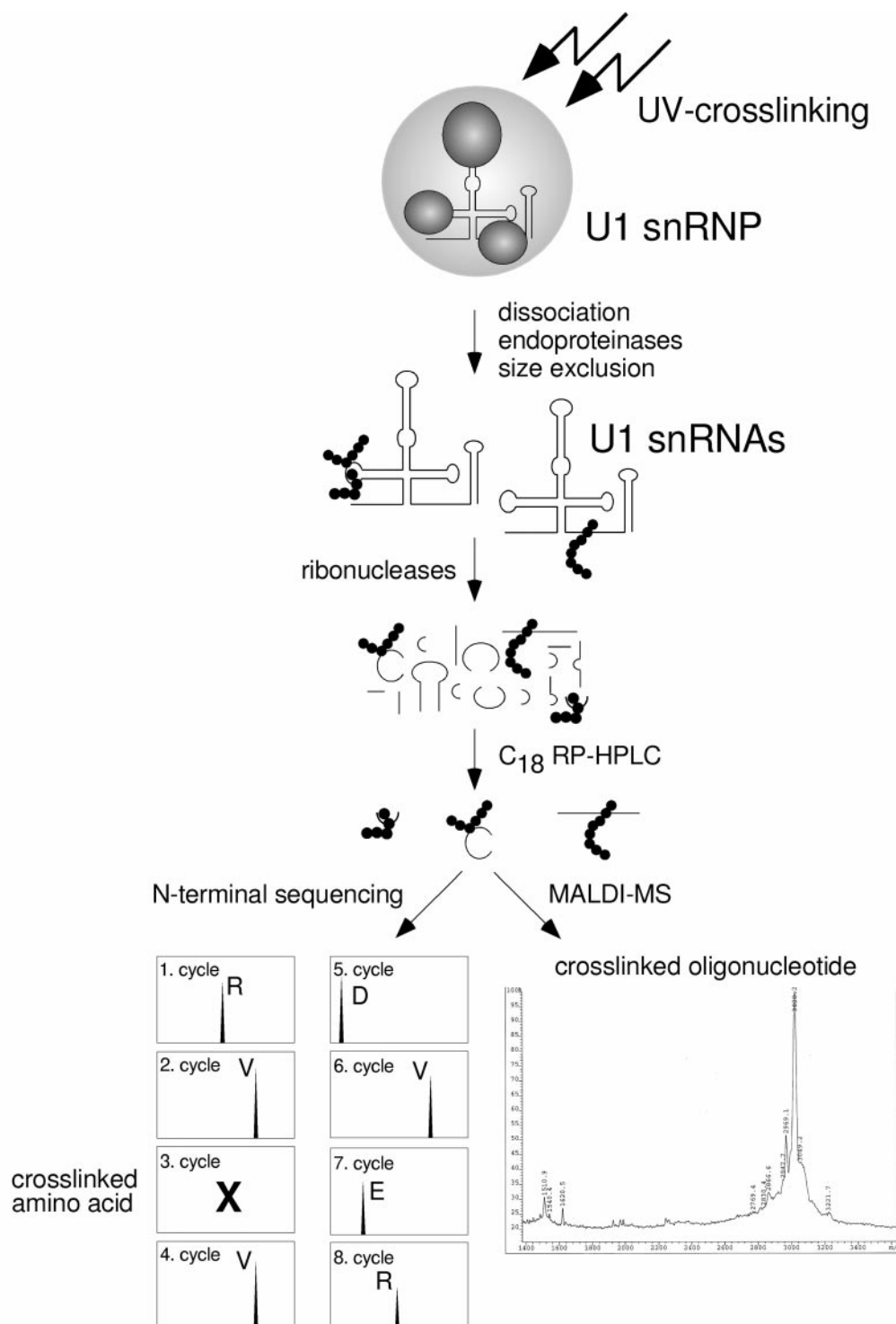
nm. Nucleotide positions of reverse transcriptase stops (A^{29} and C^{31}) are given on the right. *C*, secondary structure of the U1 snRNA. *Arrows* depict nucleotides that are identified to be cross-linked to the U1 70K protein (U^{30} and G^{28}). Note that the positions of cross-linked nucleotides are one position upstream of the stops caused by the reverse transcriptase (A^{29} and C^{31}; see *B*). The position of the cDNA primer used for primer extension analysis is indicated by the *solid line* in stem-loop II.

FIG. 3. **General strategy for N-terminal sequencing and MALDI-MS analysis of cross-linked peptide-oligonucleotide complexes isolated from UV-irradiated U1 snRNPs.** 0.8–1.0 mg of UV-cross-linked U1 snRNPs was dissociated in the presence of 8 M urea and diluted to a final urea concentration of ≤1 M, and the protein moiety was digested with various endoproteinases (see "Experimental Procedures"). Size-exclusion chromatography separated snRNAs and snRNAs that carry cross-linked peptides from the non-cross-linked peptide moiety. Cross-link-enriched snRNAs were digested with endoproteinase T1 or A, and cross-linked peptide-oligonucleotide complexes were then separated by RP-HPLC. RP-HPLC fractions with absorbances at 220 and 260 nm were collected and analyzed by automated N-terminal sequencing and MALDI-MS analysis. N-terminal sequence analysis revealed the cross-linked amino acid because a gap in the analysis of the Edman degradation products occurs at the site of the cross-linked amino acid (37–39). MALDI-MS of the cross-linked peptide-oligonucleotide complex identified the oligonucleotide part. The mass difference between the complex and the sequenced peptide reveals the composition of the cross-linked oligonucleotide (34, 35).

The analysis of the Edman degradation products of the tryptic fragments is shown in Fig. 4. The major sequence was identified as RVXVDVER (where X is an unknown amino acid) and corresponds to a tryptic fragment of the U1 70K protein spanning positions 173–180, RVLVDVER. The minor sequence is read as VNXDTTESKLR, corresponding to a tryptic fragment of the U1 70K protein from positions 110 to 120, *i.e.* VNYDTTESKLR. Importantly, Leu[175] of the major sequence and Tyr[112] of the minor sequence were absent in cycle 3 of the analysis (denoted as X in Fig. 4), thus confirming that both
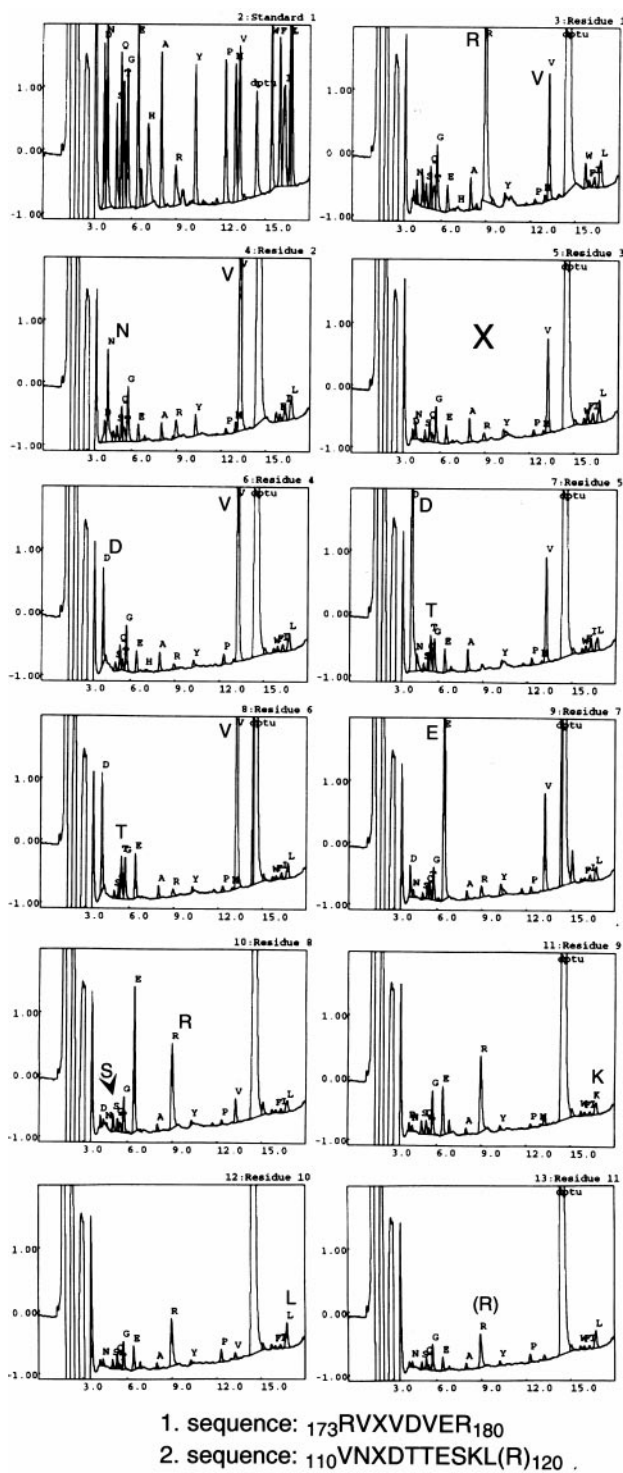
FIG. 4. N-terminal sequence analysis of the isolated U1 70K tryptic fragments cross-linked to U1 snRNA oligonucleotides. Both tryptic fragments eluted within the same fraction from the RP-HPLC column and hence were sequenced concomitantly. The analysis and identification of the Edman degradation products of cycles 1–11 of both peptides are shown. The *first panel* shows the elution profile of the phenylthiohydantoin-derivative standards (each 10 pmol) given in one-letter amino acid code. *dptu* is diphenylthiourea, which is an Edman degradation by-product. Amino acids corresponding to residues 1–11 are in *boldface*. The C-terminal arginine of the minor sequence could not be unambiguously identified and is therefore shown in *parentheses*. The N-terminal sequences of both peptides and their positions within the U1 70K protein sequence are listed. The phenylthiohydantoin-derivatives of residue 3 (Leu[175] and Tyr[112], respectively) are missing (designated as *X*), confirming these residues as amino acids cross-linked to the U1 snRNA (see "Results" for details).

amino acid residues are cross-linked to the U1 snRNA within the U1 70K tryptic peptides.

An aliquot of this fraction was subjected to MALDI-MS (Fig. 5) to determine the mass of the cross-linked peptide-oligonucleotide complex. The difference between the mass of the cross-linked peptide-oligonucleotide complex and that of the peptide alone (RVLVDVER, $[M + H]+ = 985.6$, and VNYDTTESKLR, $[M + H]+ = 1325.7$; cross-linked amino acids are underlined) allows the determination of the composition of the cross-linked oligonucleotide (34, 35). MALDI-MS analysis of the fraction showed a mass peak of 3020.2 (Fig. 5*A*), but a mass difference of either 2034.6 or 1694.5, respectively, does not correspond to any T1 fragment of U1 snRNA sequence. However, it is well known from MALDI-MS analysis of oligonucleotides that a variety of metal counterions interact with the phosphate backbone, causing multiple or "false" mass peaks of cross-linked complexes with increasing numbers of nucleotides (40). In fact, the exact mass of a large cross-linked complex could previously be determined only after performing an ion-exchange procedure (34, 40). Taking this into account, the mass of 3020.2 corresponds to a T1 oligonucleotide of U1 snRNA with the nucleotide composition $G_1A_2Y_3$ (where Y is pyrimidine since C and U differ by only 1 mass unit) with four Mg^{2+} ions attached, cross-linked to the major tryptic fragment of the U1 70K protein, RVLVDVER (Fig. 5*C*; see figure legend for further details). Only one T1 fragment matches this calculated composition, namely 5′-AUCACG-3′ from positions 29 to 34 of stem-loop I. This was verified by further MALDI-MS analysis of the fraction after partial hydrolysis (Fig. 5*B*). The spectrum shows multiple mass peaks (designated *b–f*), which are analyzed in Fig. 5*C*. They correspond to the hydrolysis products of the U1 snRNA oligonucleotide still cross-linked to the [173]RVLVD-VER[180] U1 70K tryptic peptide. The cross-linked oligonucleotide composition perfectly matches the composition of the U1 snRNA T1 fragment from positions 29 to 34 of stem-loop I (5′-AUCACG-3′) (Fig. 5*D*). Furthermore, the mass analysis revealed that the actual cross-linking site must be located at the 5′-end of the fragment (5′-AUCACG-3′) (Fig. 5, *B* and *C*, *mass peak a/b*). This is consistent with the identification of U^{30} being one of the cross-linking sites for the U1 70K protein as detected by immunoprecipitation combined with primer extension analysis (Fig. 2, *B* and *C*; see above). Although the fraction analyzed contained the minor second tryptic peptide of the U1 70K protein ([110]VNYDTTESKLR[120]) cross-linked to U1 snRNA (see above), the corresponding cross-linked oligonucleotide could not be identified in this experiment because the major cross-linked complex in this fraction (RVLVDVER cross-linked to 5′-AUCACG-3′) obscured the minor complex in the mass spectrum.

In a similar experiment using chymotrypsin and RNase A for the generation of cross-linked peptide-oligonucleotide complexes, we could isolate and sequence an RP-HPLC fraction containing a predominant chymotryptic fragment of the U1 70K protein, [107]VARVNYDTTESKL[119] (data not shown). In absolute agreement with the data derived from the tryptic fragment, Tyr[112] was found by Edman degradation to be the actual site of cross-linking to the U1 snRNA in the chymotryptic fragment (data not shown). Fig. 6*A* shows the MALDI-MS analysis of the fraction. The mass peak designated as *a* (2492.7) corresponds to the mass of the peptide (VARVNYDTTESKL, $[M + H]+ = 1495.8$) cross-linked to a 3-mer oligonucleotide with the composition $G_1A_1Y_1$ (Fig. 6*B*). The other mass peaks could not be assigned, but most probably correspond to contaminating minor peptides within the fraction already apparent in the Edman degradation. Mass peaks of cross-linked peptide-oligonucleotide complexes shows a reduced intensity in com-

FIG. 5. **MALDI-MS analysis of the U1 70K tryptic peptide ([173]RVLVD-VER[180]) cross-linked via Leu[175] to an RNase T1 fragment ([29]AUCACG[34]) in stem-loop I of U1 snRNA.** *A*, mass spectrum of the RP-HPLC-purified fraction containing two U1 70K tryptic fragments ([173]RVLVDVER[180] and [107]VARVNY-DTTESKL[119]) cross-linked to U1 snRNA T1 oligonucleotides. *B*, mass spectrum of the fraction after partial hydrolysis of the cross-linked oligonucleotide. *C*, nucleotide composition of the U1 snRNA T1 oligonucleotide cross-linked to the U1 70K tryptic fragment ([173]RVLVDVER[180]). The nucleotide composition (*fourth column*) was determined from *mass peaks a–f* as shown in *A* and *B*. The cross-linked U1 70K peptide sequence (fragment 173–180) and its mass ($[M + H]^+$) are also listed (*third column*). The cross-link site within the peptide (Leu[175]) as identified by Edman degradation (see Fig. 3) is *underlined*. *Y* denotes pyrimidines C and U. Note that C and U differ by only 1 mass unit (323 and 324, respectively). Since MALDI-MS in the linear mode does not allow the unambiguous differentiation between C and U, different compositions of the cross-linked oligonucleotide were considered. Hence, the calculated mass (*second column*) is given as the average of all possible combinations of C and U residues within the oligonucleotide concerned. The mass of the cross-linked T1 oligonucleotide was calculated as 1937.5 ± 1.5. The mass difference of 97.1 ± 1.5 between the measured mass of *peak a* (*A*; 3020.2) and the calculated mass of 2923.1 ± 1.5 (985.6 + (1937.5 ± 1.5) = 2923.1 ± 1.5) is due to the interaction of four Mg^{2+} ions with the complex (*fourth column*). The calculated mass of the total complex (*second column*) therefore includes the masses of four Mg^{2+} ions ((2923.1 ± 1.5) + 97.2 = 3020.3 ± 1.5). Furthermore, the partial hydrolysis of the cross-linked complex resulted in fragments that have 2′,3′-cyclic phosphate termini (*mass peaks b, d,* and *f*), which is indicated as $-H_2O$ in the *fourth column*. *D*, sequence of the U1 snRNA T1 oligonucleotide cross-linked to Leu[175] in the U1 70K protein. Positions on the U1 snRNA are shown as *subscript numbers*. *Brackets* indicate either the mass of the total complex (*mass peak a*) or of fragments obtained after partial hydrolysis (*mass peaks b–f*). Designation of the *bars* correspond to the peaks in *A* and *B* and the *first column* in *C*.
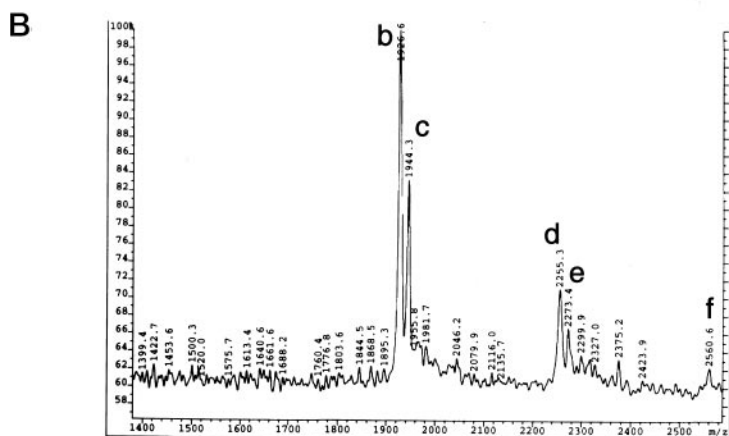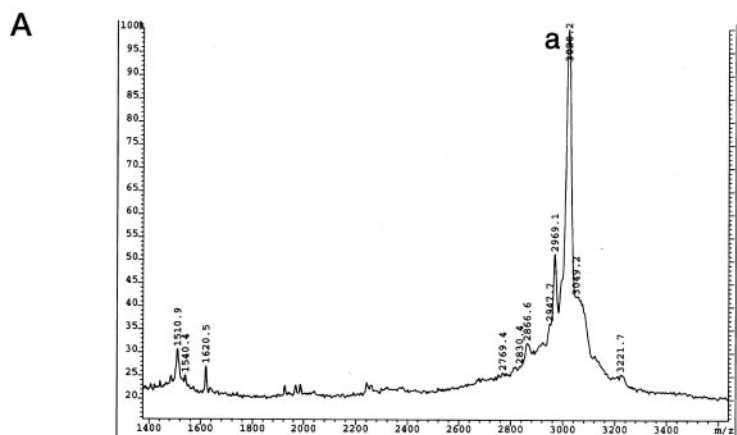
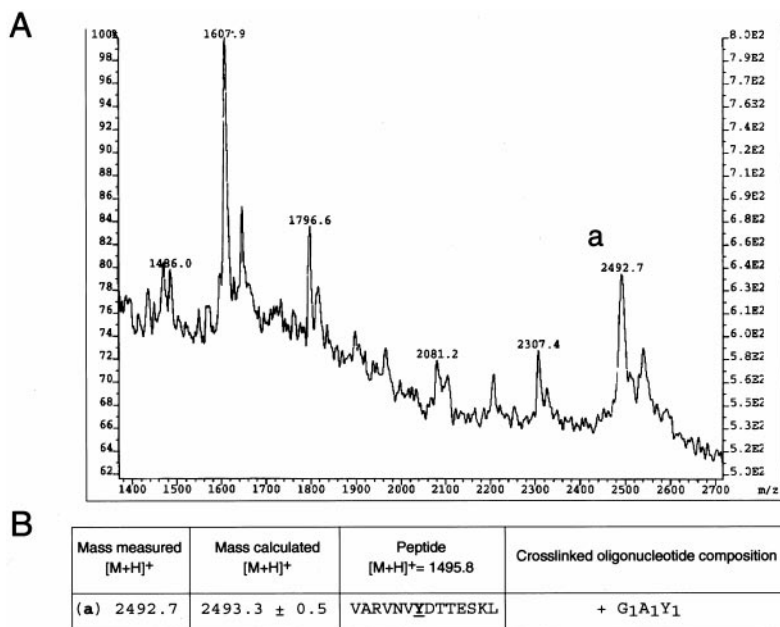|	Mass measured $[M+H]^+$	Mass calculated $[M+H]^+$	Peptide $[M+H]^+= 985.6$	Crosslinked oligonucleotide composition
(a)	3020.2	3020.3 ± 1.5	RVL̲VDVER	+ $G_1A_2Y_3$, + 4 Mg^{2+}
(b)	1926.6	1925.6 ± 1.0	RVL̲VDVER	+ A_1Y_2, − H_2O
(c)	1944.3	1943.6 ± 1.0	RVL̲VDVER	+ A_1Y_2
(d)	2255.3	2254.6 ± 1.0	RVL̲VDVER	+ A_2Y_2, − H_2O
(e)	2273.3	2272.6 ± 1.0	RVL̲VDVER	+ A_2Y_2
(f)	2560.6	2560.1 ± 1.5	RVL̲VDVER	+ A_2Y_3, − H_2O
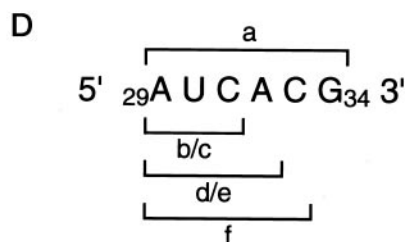
parison with non-cross-linked peptides when measured under standard conditions for peptides.[2] HeLa U1 snRNA contains four RNase A fragments with the determined nucleotide composition ([28]GAU[30], [84]GAC[86], [93]GAU[95], and [135]AGU[137]). The GAU 3-mer from positions 28 to 30 is located in stem-loop I of the U1 snRNA and encompasses the second cross-linking site for the U1 70K protein (G[28]) as identified independently by our

immunoprecipitation and primer extension analysis (Fig. 2*B*; see above). Our sequencing results with the cross-linked U1 70K peptide-oligonucleotide complexes clearly confirm our interpretation of the immunoprecipitation experiment, *i.e.* that the U1 70K protein is cross-linked via two independent sites to U[30] and G[28] in stem-loop I. We conclude that Tyr[112] of the U1 70K protein is cross-linked to G[28] and Leu[175] to U[30]. In summary, immunoprecipitation combined with primer extension analysis is shown to be a reliable method for the detection of

---

[2] B. Thiede and H. Urlaub, unpublished data.

FIG. 6. *A*, MALDI-MS spectrum of the RP-HPLC-purified fraction containing mainly the U1 70K chymotryptic fragment ([107]VARVNYDTTESKL[119]) cross-linked via Tyr[112] to an RNase A fragment of U1 snRNA. *B*, nucleotide composition of the RNase A fragment of U1 snRNA cross-linked to the U1 70K chymotryptic fragment ([107]VARVNYDTTESKL[119]). The nucleotide composition is listed in the *forth column* and was determined from *mass peak a* in *A*. The cross-linked U1 70K peptide sequence (fragment 107–119) and its mass ($[M + H]^+$) are also listed (*third column*). The cross-linked amino acid within the peptide (Tyr[112]) as identified by Edman degradation is *underlined* (see Fig. 3). See "Results" and the legend to Fig. 5 for further details.

Mass measured $[M+H]^+$	Mass calculated $[M+H]^+$	Peptide $[M+H]^+= 1495.8$	Crosslinked oligonucleotide composition
(a) 2492.7	2493.3 ± 0.5	VARVNV**Y**DTTESKL	+ $G_1A_1Y_1$

RNA-protein cross-linking sites on the RNA in native UV-irradiated snRNP particles.

U5 snRNA-Protein Cross-linking Sites within Native 25 S U4/U6.U5 Tri-snRNPs—We used our primer extension approach to investigate the U5 snRNA-protein interaction within native 25 S U4/U6.U5 tri-snRNP particles isolated from HeLa cells. Fig. 7A shows the primer extension analysis of U5 snRNA derived from cross-linked tri-snRNP particles (*lanes 3* and *4*) and of UV-irradiated naked U5 snRNA (*lanes 1* and *2*). In comparison with the irradiated naked U snRNAs, primer extension analysis of U5 snRNA derived from UV-irradiated tri-snRNP particles shows additional strong reverse transcriptase stops at U^{41} to A^{44} (Fig. 7A, *lane 4*), corresponding to U^{40} to U^{43} within the highly conserved loop 1 being cross-linked (Fig. 7B). Although stops at these nucleotides, in particular at U^{40} and U^{41}, were also present in the irradiated naked U5 snRNA, they were significantly increased (~50-fold) in the irradiated tri-snRNP sample (Fig. 7A, compare *lanes 2* and *4*). This suggests that U^{40} to U^{43} are targets for RNA-protein cross-links within loop 1 of U5 snRNA. In addition, less strong RNA-protein cross-linking sites were observed at A^{70} and U^{72} in the 3'-half of IL1 (Fig. 7A; see also Fig. 7C for U5 snRNA sequence), corresponding to reverse transcriptase stops at C^{71} and C^{73} (Fig. 7A). Weak stops were also observed at A^{47} and in the 5'-half of IL2 (Fig. 7A). At these sites, no stops were detectable on the irradiated naked U5 snRNA (Fig. 7A, compare *lanes 2* and *4*), demonstrating that they are also due to RNA-protein cross-linking events.

To identify the U5-specific proteins cross-linked to these nucleotides within U5 snRNA, we applied our procedure of immunoprecipitation and primer extension to the UV-cross-linked tri-snRNP particles using antibodies specific for the 220-, 200-, 116-, and 40-kDa proteins (41–43). These four proteins form a remarkably stable heteromeric protein complex in the absence of U5 snRNA (43). Fig. 7B shows the results of the primer extension analysis of U5 snRNA after cross-linking of the tri-snRNP particles and immunoprecipitation of the 220- and 116-kDa proteins. Reverse transcriptase stops were exclusively detected in the UV-light irradiated sample that was subjected to immunoprecipitation with anti-220-kDa protein antibody (Fig. 7B, *lane 2*). The stops observed after immunoprecipitation with the anti-220-kDa protein antibody are located at U^{41} to A^{44} (Fig. 7B), corresponding to cross-links to U^{40}

to U^{43} within the highly conserved loop 1 of U5 snRNA. In addition, a weaker reverse transcriptase stop was observed at C^{73}, corresponding to a cross-link to U^{72} within the 3'-half of IL1 of U5 snRNA. No stops were detected within loop 1 and IL1 of U5 snRNA with anti-116-kDa protein antibody (Fig. 7B, *lane 4*) or anti-200- and anti-40-kDa protein antibodies (data not shown). This clearly demonstrates that within native tri-snRNP particles, only the U5-specific 220-kDa protein is cross-linked to the four adjacent nucleotides within loop 1 as well as to U^{72} within the 3'-half if IL1 of U5 snRNA (Fig. 7C). The absence of any detectable full-length U5 snRNA transcript in either the irradiated (Fig. 7B, *lanes 2* and *4*) or non-cross-linked (*lanes 1* and *3*) samples shows that non-cross-linked U5 snRNA was not coprecipitated. After UV irradiation, only those U5 snRNAs were coprecipitated that were covalently attached to the 220-kDa protein via cross-links to nucleotides in either loop 1 or IL1 (Fig. 7C). The low overall level of cross-linking yield argues against multiple cross-links occurring in one U5 snRNA molecule. Therefore, every nucleotide that causes a reverse transcriptase stop in this experiment represents an authentic cross-linking site for the 220-kDa protein. This situation is similar to that observed in the case of the U1 70K protein (Fig. 2; see above), where independent sites within the protein became cross-linked to neighboring nucleotides of stem-loop I of U1 snRNA, thus causing apparently multiple reverse transcriptase stops.

DISCUSSION

In this study, we have employed a novel method involving UV cross-linking to investigate direct snRNA-protein interactions within native HeLa snRNP. To identify the exact sites of cross-linking of the proteins to the RNA, we developed an approach that combines immunoprecipitation of cross-linked proteins with primer extension analysis of the cross-linked RNA moiety.

To test the feasibility of our approach, we chose native U1 snRNP and tri-snRNP particles purified from HeLa cells. The primer extension analysis subsequent to immunoprecipitation identified multiple cross-linking sites of the U1 70K protein in U1 snRNA (G^{28} and U^{30}) and of the U5 snRNP-specific 220-kDa protein in the U5 snRNA (U^{40} to U^{43} and U^{72}). The fact that the U1 70K protein amino acids Tyr[112] and Leu[175] were found to be cross-linked to G^{28} and U^{30}, respectively, by N-
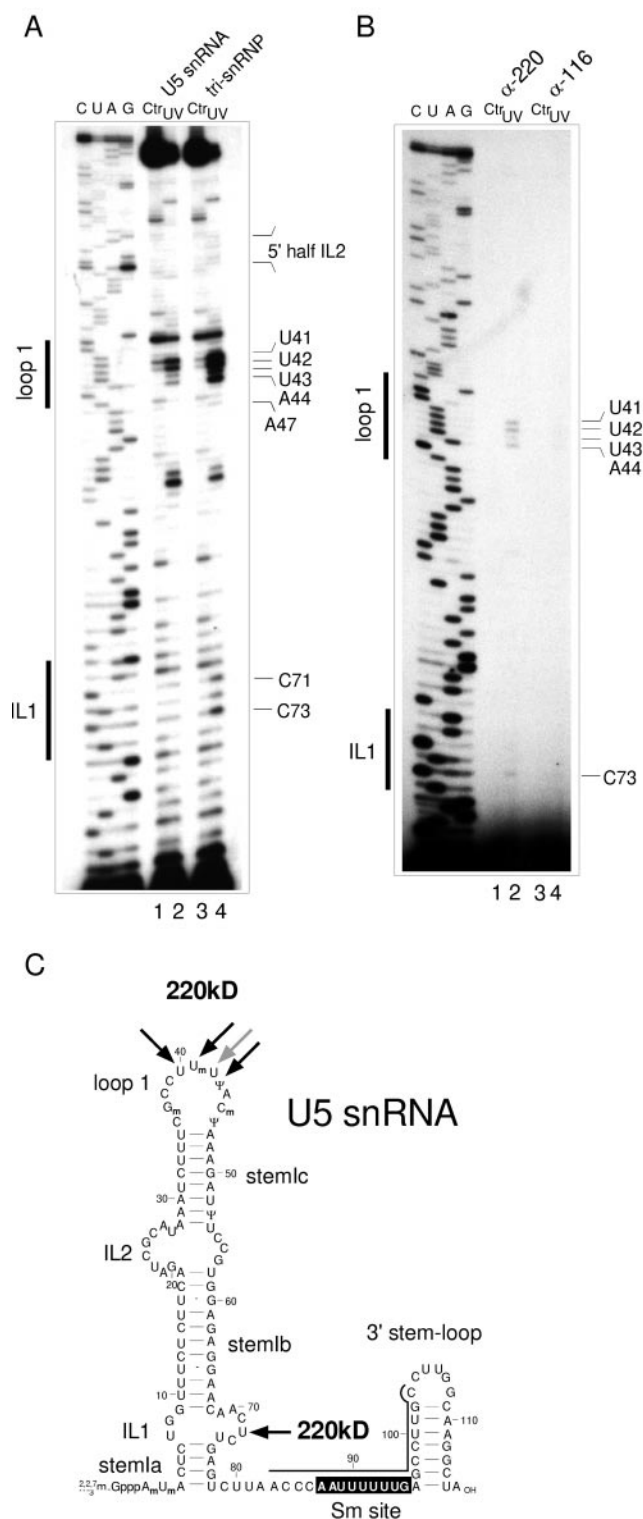
FIG. 7. *A,* primer extension analysis of U5 snRNA derived from UV-cross-linked 25 S U4/U6.U5 tri-snRNP (*lanes 3* and *4*) compared with UV-irradiated naked U5 snRNA (*lanes 1* and *2*). *Lanes 1* and *3,* no UV irradiation (control (*Ctr*)); *lanes 2* and *4,* 2 min of UV irradiation at 254 nm (*UV*). *C, U, A,* and *G* indicate dideoxy sequence markers. The cDNA primer used in this experiment is complementary to positions 83–103 (indicated as a *solid line* in *C*). The *in vitro* U5 snRNA transcript used for generation of marker lanes has five additional nucleotides on the 5′-end. Reverse transcriptase stops at U^{41} to A^{44}, C^{71}, and C^{73} and weak stops at nucleotides within the 5′-half of IL2 are denoted on the right. *Black bars* on the left indicate loop 1 and the 3′-half of IL1 of U5 snRNA. *B,* primer extension analysis of cross-linked U5 snRNA after immuno-precipitation of cross-linked 25 S U4/U6.U5 tri-snRNPs with anti-220-kDa protein (*α-220*) or anti-116-kDa protein (*α-116*) antibody. Prior to immunoprecipitation, cross-linked 25 S U4/U6.U5 tri-snRNPs were

terminal sequencing and MALDI-MS of purified U1 70K peptide-oligonucleotide cross-links (Figs. 4–6) provides an independent confirmation that each strong reverse transcriptase stop observed after immunoprecipitation is an authentic cross-linking site. Our approach can therefore be considered as a general approach suitable for the detection of single and/or multiple RNA-protein contact sites in a variety of different native UV-irradiated RNP particles.

Since no label for either the RNA or the protein moiety can be used in native particles, it is clear that our approach is not as sensitive as when labeled components are used; and therefore, more material is required. Despite the lower sensitivity, our approach has advantages over those that use reconstituted particles. First, purified native particles are stable and fully assembled and thus more homogeneous. Reconstitution of particles has been successfully used to increase the cross-linking yield by incorporation of site-specific cross-links. However, cross-linking of such particles depends on the efficiency of reconstitution, and incomplete or incorrect assembly can result in different subpopulations or false positives, which in turn complicate the interpretation of the complex cross-linking pattern. Second, primer extension analysis of the cross-linked RNAs subsequent to immunoprecipitation can reveal multiple contact sites between one protein and the RNA within one experiment. Thus, our approach allows the rapid and exact identification of RNA-protein cross-links.

In addition to its methodological importance, this work also contributes valuable information that helps in our understanding of the molecular organization of the U1 particle as well as that of U5 snRNP within the context of tri-snRNP. Previous deletion and mutation analyses demonstrated that the U1 70K protein directly and specifically interacts with stem-loop I of U1 snRNA via its RNA-binding domain (RBD) (11) (Fig. 8*A*), requiring 8 of 10 bases (positions 28–37) of the loop for binding (12, 13). Our U1 70K protein cross-linking results (cross-linking of Tyr^{112} and Leu^{175} to G^{28} and U^{30} within stem-loop I, respectively) complement the previous studies and allow us to formulate a structural model to explain the U1 70K RBD interaction with stem-loop I of U1 snRNA. Based on a combination of a secondary structure prediction of the U1 70K RBD (11, 44) (Fig. 8*A*) and crystallographic data of other RBD-containing proteins complexed with RNA (U1 A (18), U2 A′/U2 B″ (21), and Sxl (45)), we modeled a three-dimensional structure of the U1 70K RBD (SWISS-MODEL) (46). In this model, the cross-linked amino acids are located in loop 1 (Tyr^{112}) and β-strand 4 (Leu^{175}) adjacent to the octamer consensus sequence motif present in β-strand 3 (11) (Fig. 8, *A* and *B*). The side chains of the amino acids directly interact with nucleotides in U1 stem-loop I separated by 1 base (Fig. 8*B*). This site-specific interaction found in the U1 70K RNP is highly reminiscent of that observed in crystal structures of other RBD-containing proteins complexed with RNA. Thus, amino acids located in loop 1 and β-strand 4 of RBD-1 from the U1 A (18) and Sxl (45) proteins contact nucleotides separated by 1 base in their cognate RNAs (of U1 snRNA stem-loop II and the transformer

dissociated with the SDS/Triton X-100 procedure (see "Experimental Procedures"). *Lanes 1* and *3,* no UV irradiation; *lanes 2* and *4,* 2 min of UV irradiation at 254 nm. See Fig. 1 for details. Nucleotide positions of reverse transcriptase stops (C^{73} and U^{41} to A^{44}) are given on the right. *Black bars* on the left indicate loop 1 and the 3′-half of IL1 of U5 snRNA. *C,* secondary structure of the U5 snRNA. *Arrows* depict nucleotides that were identified to be cross-linked to the 220-kDa protein (U^{40} to U^{43} and U^{72}). Note that the positions of the cross-linked nucleotides are one position upstream of the stops caused by the reverse transcriptase (C^{73} and U^{41} to A^{44}; see *A* and *B*). The *gray arrow* indicates the weaker cross-link at U^{42}. The position of the cDNA primer used for primer extension analysis is indicated by the *solid line.*
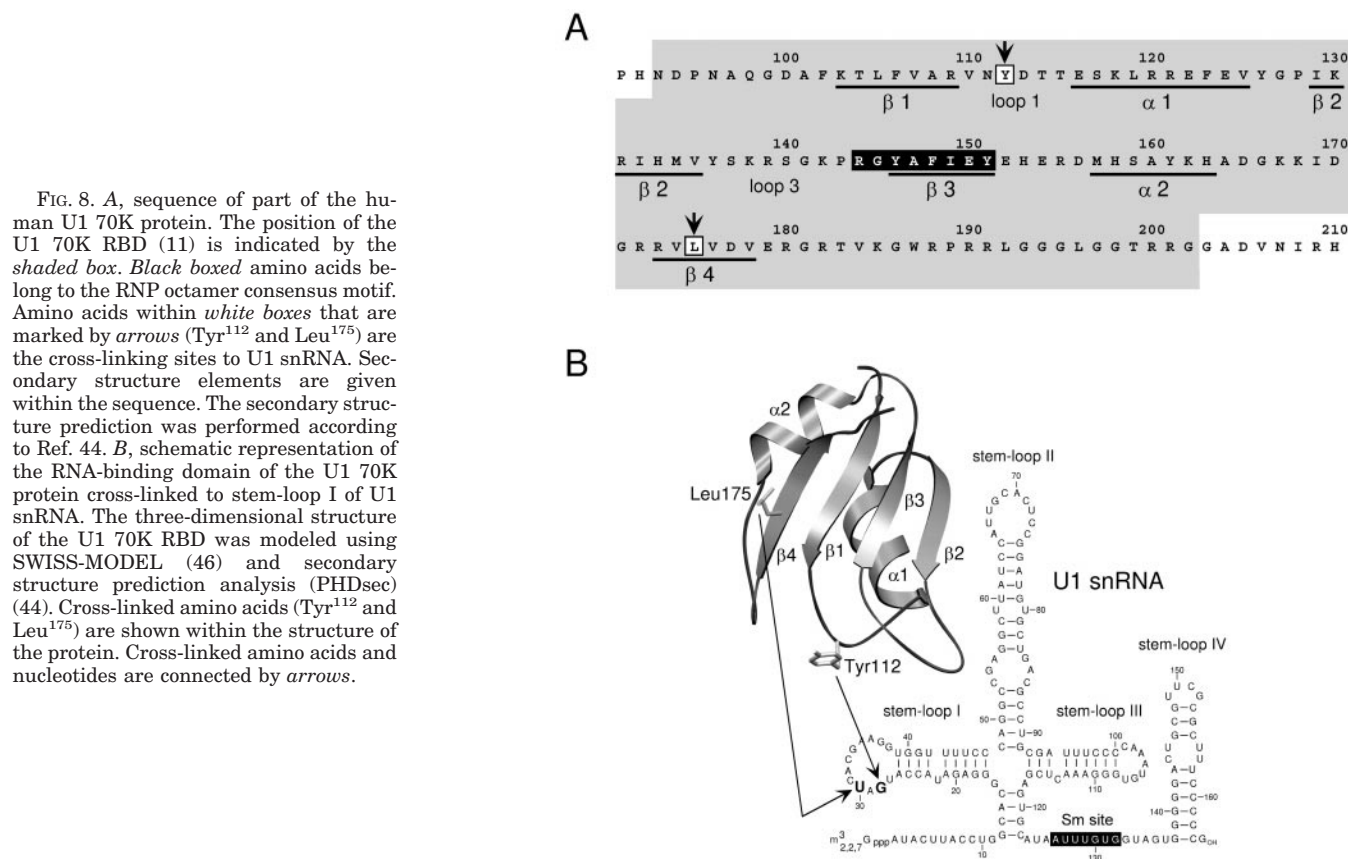
Fig. 8. *A*, sequence of part of the human U1 70K protein. The position of the U1 70K RBD (11) is indicated by the *shaded box*. *Black boxed* amino acids belong to the RNP octamer consensus motif. Amino acids within *white boxes* that are marked by *arrows* (Tyr[112] and Leu[175]) are the cross-linking sites to U1 snRNA. Secondary structure elements are given within the sequence. The secondary structure prediction was performed according to Ref. 44. *B*, schematic representation of the RNA-binding domain of the U1 70K protein cross-linked to stem-loop I of U1 snRNA. The three-dimensional structure of the U1 70K RBD was modeled using SWISS-MODEL (46) and secondary structure prediction analysis (PHDsec) (44). Cross-linked amino acids (Tyr[112] and Leu[175]) are shown within the structure of the protein. Cross-linked amino acids and nucleotides are connected by *arrows*.

polypyrimidine tract, respectively) (18, 45). Our results therefore provide further evidence for the highly conserved nature of the RBD-RNA interactions.

In UV-irradiated tri-snRNP particles, we demonstrated that the predominant UV-induced cross-links occur between the U5 snRNP-specific 220-kDa protein and several nucleotides within the highly conserved loop 1 of U5 snRNA. In addition, we observed a weaker site of cross-linking of the 220-kDa protein to U[72] in the 3′-half of IL1 (Fig. 7).

Our cross-linking results from HeLa tri-snRNP complement and extend previous cross-linking studies of U5 snRNPs reconstituted *in vitro* within yeast nuclear extracts (28). In these studies, the 220-kDa yeast homologue Prp8p was found to be cross-linked mainly to loop 1 of U5 snRNA. Additional site-specific cross-linking sites for Prp8p were also identified in the 3′-half of IL1 (corresponding to IL2 in the yeast nomenclature) and for nucleotides in the 5′- and 3′-halves of IL2 (corresponding to yeast IL1). In our studies, we further observed RNA-protein cross-linking sites within the 5′-half of IL2 of U5 snRNA. Due to their low abundance, we were not able to identify the U5 or tri-snRNP(s) involved in these or the other weak RNA-protein cross-linking sites observed (to U[46] and A[70]; see "Results"). In yeast, the U5 snRNP-specific Snu114p was found to be cross-linked to the 5′-half of IL2 (yeast IL1) (28). We did not identify the Snu114p human homologue, the U5 116-kDa protein, within any cross-linked products. A possible explanation for the different 116-kDa protein/Snu114p cross-linking pattern may be that yeast U5 snRNP assumes a different RNP conformation than HeLa U5 snRNP, possibly due to the additional variable stem-loop in yeast that is located close to the cross-linking site of Snu114p.

Taken together, our 220-kDa protein cross-linking data suggest that, similar to the situation in yeast, multiple 220-kDa protein regions directly contact U5 snRNA and that the human 220-kDa protein spans the entire 5′-stem-loop of U5 snRNA.

Thus, not only is the 220-kDa protein evolutionarily highly conserved between yeast and man, but also its interactions with the U5 snRNA within U5 snRNPs.

The fact that the highly conserved loop 1 of U5 snRNA extensively contacts the 220-kDa protein (with U[40] to U[43] being cross-linked to this protein (see Fig. 7)) deserves special attention for the following reasons. First, it has been shown within both the yeast and mammalian *in vitro* splicing systems that deletion of loop 1 has no effect on the first catalytic step of splicing (47, 48). Additionally, loop 1 is dispensable for the second catalytic step of splicing within the mammalian splicing system (48). These results suggest that, under certain conditions, loop 1 is not absolutely required for splicing and that other spliceosomal factors can compensate for it when it is absent. The extended interaction surface between loop 1 and the 220-kDa protein, observed in our experiments, reinforces the idea that the 220-kDa protein substitutes for the function of loop 1 in exon alignment prior to the second catalytic step of splicing when loop 1 is deleted (48). Second, since loop 1 and the 220-kDa protein/Prp8p can be cross-linked to equivalent positions at the 5′- and 3′-splice sites in HeLa cells as well as in the yeast system (49–51), it has been hypothesized that one role of the protein is the stabilization of the loop 1-pre-mRNA interaction throughout the splicing reaction (50). The fact that at least two of the nucleotides that we identified here as cross-linking sites for the 220-kDa protein, namely U[40] and U[41], are identical to those found cross-linked to the 5′-splice site and adjacent to the 3′-splice site in both splicing systems (52, 53) lends support to the proposed function of the protein in exon alignment.

Cross-linking and genetic studies of the 220-kDa protein and its yeast homologue indicate a functional interaction of the carboxyl-terminal region of the protein and the 5′-splice site as well as the 3′-splice site in the catalytic center (54–56). Accordingly, the exact localization of the region of the 220-kDa protein

that is involved in the observed extended interaction with loop 1 of U5 snRNA (*e.g.* according to the methods outlined above for the U1 70K protein) will be important, as it allows us to gain further insight into the functional and/or structural domains of the protein that are close to or within the catalytic center of the spliceosome. As it has recently been shown that Prp8p also directly interacts with U6 snRNA (29) and seems to stabilize tertiary interactions of the 5′-splice site, the 3′-splice site, and U6 snRNA prior to the second step of splicing (57), it would thus be interesting to see whether distinct protein regions or domains might have distinct functional features.

REFERENCES

1. Krämer, A. (1996) *Annu. Rev. Biochem.* **65,** 367–409
2. Will, C. L., and Lührmann, R. (1997) *Curr. Opin. Cell Biol.* **9,** 320–328
3. Will, C. L., and Lührmann, R. (1997) in *Eukaryotic mRNA Processing* (Krainer A. R., ed) pp. 130–173, Oxford University Press, Oxford
4. Staley, J. P., and Guthrie, C. (1998) *Cell* **92,** 315–326
5. Burge, C. B., Tuschl T., and Sharp, P. A. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. E., eds) 2nd Ed., pp. 525–560, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
6. Kambach, C., Walke, S., and Nagai, K. (1999) *Curr. Opin. Struct. Biol.* **9,** 222–230
7. Kastner, B. (1998) in *RNP Particles, Splicing and Autoimmune Diseases* (Schenkel, J., ed) pp. 95–140, Springer-Verlag, Berlin
8. Kastner, B., Kornstadt, U., Bach, M., and Lührmann, R. (1992) *J. Cell Biol.* **116,** 839–849
9. Raker, V. A., Plessel, G., and Lührmann, R. (1996) *EMBO J.* **15,** 2256–2269
10. Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Lührmann, R., Li, J., and Nagai, K. (1999) *Cell* **96,** 375–387
11. Query, C. C., Bentley, R. C., and Keene, J. D. (1989) *Cell* **57,** 89–101
12. Surowy, C. S., van Santen, V. L., Scheib-Wixted, S. M., and Spritz, R. A. (1989) *Mol. Cell. Biol.* **9,** 4179–4186
13. Query, C. C., Bentley, R. C., and Keene, J. D. (1989) *Mol. Cell. Biol.* **9,** 4872–4881
14. Lutz-Freyermuth, C., and Keene, J. D. (1989) *Mol. Cell. Biol.* **9,** 2975–2982
15. Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J., and Mattaj, I. W. (1989) *EMBO J.* **8,** 4163–4170
16. Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., and Nagai, K. (1991) *EMBO J.* **10,** 3447–3456
17. Nelissen, R. L., Will, C. L., van Venrooij, W. J., and Lührmann, R. (1994) *EMBO J.* **13,** 4113–4125
18. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) *Nature* **372,** 432–438
19. Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., and Mattaj, I. W. (1990) *Nature* **345,** 502–506
20. Scherly, D., Dathan, N. A., Boelens, W., van Venrooij, W. J., and Mattaj, I. W. (1990) *EMBO J.* **9,** 3675–3681
21. Price, S. R., Evans, P. R., and Nagai, K. (1998) *Nature* **394,** 645–650
22. Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Lührmann, R. (1999) *EMBO J.* **18,** 6119–6133
23. Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999) *EMBO J.* **18,** 5789–5802
24. Mayes, A. E., Verdone, L., Legrain, P., and Beggs, J. D. (1999) *EMBO J.* **18,** 4321–4331
25. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S., and Séraphin, B. (1999) *EMBO J.* **18,** 3451–3462
26. Woppmann, A., Rinke, J., and Lührmann, R. (1988) *Nucleic Acids Res.* **16,** 10985–11004
27. Heinrichs, V., Hackl, W., and Lührmann, R. (1992) *J. Mol. Biol.* **227,** 15–28
28. Dix, I., Russell, C. S., O'Keefe, R. T., Newman, A. J., and Beggs, J. D. (1998) *RNA* **4,** 1675–1686
29. Vidal, V. P., Verdone, L., Mayes, A. E., and Beggs, J. D. (1999) *RNA* **5,** 1470–1481
30. Laggerbauer, B., Lauber, J., and Lührmann, R. (1996) *Nucleic Acids Res.* **24,** 868–875
31. Bach, M., Krol, A., and Lührmann, R. (1990) *Nucleic Acids Res.* **18,** 449–457
32. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res.* **11,** 1475–1489
33. Hartmuth, K., Raker, V. A., Huber, J., Branlant, C., and Lührmann, R. (1999) *J. Mol. Biol.* **285,** 133–147
34. Urlaub, H., Thiede, B., Müller, E. C., Brimacombe, R., and Wittmann-Liebold, B. (1997) *J. Biol. Chem.* **272,** 14547–14555
35. Thiede, B., Urlaub, H., Neubauer, H., and Wittmann-Liebold, B. (1999) *Methods Mol. Biol.* **118,** 63–72
36. Chiara, M. D., Palandjian, L., Feld Kramer, R., and Reed, R. (1997) *EMBO J.* **16,** 4746–4759
37. Urlaub, H., Kruft, V., Bischof, O., Müller, E. C., and Wittmann-Liebold, B. (1995) *EMBO J.* **14,** 4578–4588
38. Metz-Boutigue, M. H., Reinbolt, J., Ebel, J. P., Ehresmann, C., and Ehresmann, B. (1989) *FEBS Lett.* **245,** 194–200
39. Willis, M. C., LeCuyer, K. A., Meisenheimer, K. M., Uhlenbeck, O. C., and Koch, T. H. (1994) *Nucleic Acids Res.* **22,** 4947–4952
40. Nordhoff, E., Cramer, R., Karas, M., Hillenkamp, F., Kirpekar, F., Kristiansen, K., and Roepstorff, P. (1993) *Nucleic Acids Res.* **21,** 3347–3357
41. Lauber, J., Fabrizio, P., Teigelkamp, S., Lane, W. S., Hartmann, E., and Lührmann, R. (1996) *EMBO J.* **15,** 4001–4015
42. Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W. S., and Lührmann, R. (1997) *EMBO J.* **16,** 4092–4106
43. Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Lührmann, R. (1998) *Mol. Cell. Biol.* **18,** 6756–6766
44. Rost, B., and Sander, C. (1993) *J. Mol. Biol.* **232,** 584–599
45. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) *Nature* **398,** 579–585
46. Guex, N., Diemand, A., and Peitsch, M. C. (1999) *Trends Biochem. Sci.* **24,** 364–367
47. O'Keefe, R. T., Norman, C., and Newman, A. J. (1996) *Cell* **86,** 679–689
48. Ségault, V., Will, C. L., Polycarpou-Schwarz, M., Mattaj, I. W., Branlant, C., and Lührmann, R. (1999) *Mol. Cell. Biol.* **19,** 2782–2790
49. Umen, J. G., and Guthrie, C. (1995) *Genes Dev.* **9,** 855–868
50. Teigelkamp, S., Newman, A. J., and Beggs, J. D. (1995) *EMBO J.* **14,** 2602–2612
51. Reyes, J. L., Kois, P., Konforti, B. B., and Konarska, M. M. (1996) *RNA* **2,** 213–225
52. Sontheimer, E. J., and Steitz, J. A. (1993) *Science* **262,** 1989–1996
53. Newman, A. J., Teigelkamp, S., and Beggs, J. D. (1995) *RNA* **1,** 968–980
54. Umen, J. G., and Guthrie, C. (1996) *Genetics* **143,** 723–739
55. Reyes, J. L., Gustafson, E. H., Luo, H. R., Moore, M. J., and Konarska, M. M. (1999) *RNA* **5,** 167–179
56. Siatecka, M., Reyes, J. L., and Konarska, M. M. (1999) *Genes Dev.* **13,** 1983–1993
57. Collins, C. A., and Guthrie, C. (1999) *Genes Dev.* **13,** 1970–1982

Henning Urlaub, Klaus Hartmuth, Susanne
Kostka, Gerlinde Grelle and Reinhard
Lührmann

Access the most updated version of this article at doi: 10.1074/jbc.M007434200

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 54 references, 29 of which can be accessed free at
http://www.jbc.org/content/275/52/41458.full.html#ref-list-1