16

Analyzing RNA–Protein Crosslinking Sites in Unlabeled Ribonucleoprotein Complexes by Mass Spectrometry

Henning Urlaub, Eva Kühn-Hölsken, and Reinhard Lührmann

Summary

Mass spectrometry is a powerful tool for the analysis of biomolecules, proteins, nucleic acids, carbohydrates, lipids. In combination with genome sequences that are available in the databases, it has proven to be the most straightforward and sensitive technique for the sequence analysis and hence the identification of protein components in the cells, their (post)translational modifications, and their relative and absolute abundance. In addition, mass spectrometric methods are successfully applied for the structural analysis of biomolecules (i.e., deciphering molecule–ligand interactions and spatial quartenary arrangements of molecule complexes). We describe a methodology for the mass spectrometric analysis of protein–RNA contact sites in purified ribonucleoprotein (RNP) particles. The method comprises ultraviolet (UV) crosslinking of proteins to RNA, hydrolysis of the protein and RNA moieties, isolation of cross-linked peptide-RNA oligonucleotides, MALDI (matrix-assisted laser desorption/ionization) mass spectrometry of the isolated conjugates to determine the sequence of the crosslinked peptide and RNA part. The utility of this methodology is demonstrated on crosslinks isolated from UV-irradiated spliceosomal particles; these were [15.5 K-61 K-U4atac] small nuclear ribonucleoprotein (snRNP) particles prepared by reconstitution in vitro and U1 snRNP particles purified from HeLa cells.

Key Words: Crosslinking; mass spectrometry; protein; RNA.

1. Introduction

RNA molecules play a fundamental role in cellular processes such as gene expression (transcription and pre-mRNA [messenger RNA] processing), post-transcriptional control (mRNA stability), RNA export, ribosomal RNA (rRNA) maturation, translation, and translational control. RNA molecules that are involved in these processes are rarely active in the absence of proteins but are found as components of stable ribonucleoprotein (RNP) particles. Since protein–RNA interactions lie at the structural and functional heart of the RNP

From: *Methods in Molecular Biology, vol. 488: RNA-Protein Interaction Protocols*
Edited by: Ren-Jang Lin © Humana Press Inc, Totowa, NJ

particles, much attention is currently being devoted to questions of protein and RNA tertiary structures and to the quaternary arrangements of the individual macromolecules in RNP particles.

We present a mass spectrometric (MS) method that we have established that allows us to characterize sites of direct protein–RNA contact in RNP particles, either native or reconstituted in vitro, after the contacts have been made permanent by ultraviolet (UV) crosslinking *(1–4)*. This method enables us to identify directly distinct regions of proteins that interact with RNA in RNP particles and thus allows the definition of novel putative RNA binding domains (RBDs) of individual proteins.

Furthermore, in the absence of highly resolved three-dimensional RNP structures, our approach yields information about the orientation and the overall arrangement of proteins and RNA within RNP particles.

In this manner, we have identified the RBDs in spliceosomal RNP particles, for example, the highly conserved Sm proteins in the U1 small nuclear RNP (snRNP) and 25 S [U4/U6.U5] tri-snRNP particles. In this case, we were able to demonstrate that the evolutionarily conserved Sm site of the snRNA is in contact with the inner surface of the heptameric Sm ring *(2)*. In addition, we found two amino acid residues within the RBD of the U1 70 K protein that are in direct contact with stem I of the U1 snRNA *(1)*. Comparison with structures of RBD-containing proteins already crystallized (such as U1 A and Sxl; *5,6*) revealed a similar amino acid–nucleotide interaction and provides further evidence for the highly conserved nature of the RBD–RNA interactions. We demonstrated that the novel U4/U6-specific protein 61 K *(4)* binds RNA directly, and that its evolutionarily conserved Nop domain *(7)* is in direct contact with the 5′ loop of the 5′ stem-loop of U4 snRNA in the presence of the 15.5 K protein.

In the protocols provided in this chapter, we refer to isolated U snRNP particles from HeLa cells *(8–10)* involved in pre-mRNA processing *(11,12)* and to [U4atac-15.5 K-61 K] protein–RNA complexes reconstituted in vitro *(4)*. Importantly, we note that the approach can be regarded as a general one so that the protocols can easily be adapted to investigations of other native RNP particles or of RNP particles reconstituted in vitro.

2. Materials

1. Purified native RNP particles (e.g., U1 snRNP particles, 25 S [U4/U6.U5] tri-snRNP particles).
2. Overexpressed RNA binding protein (e.g., U4/U6-specific 15.5 K protein, U4/U6-specific 61 K protein).
3. RNA prepared by transcription in vitro (e.g., U4atac snRNA).
4. UV crosslinking equipment (**Fig. 1**) or UV Stratalinker 2400 (Stratagene, La Jolla, CA).

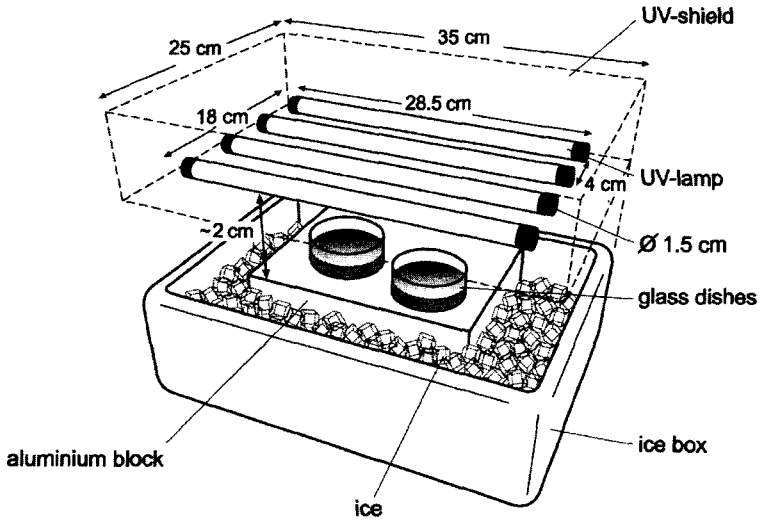


Fig. 1. Apparatus for ultraviolet (UV) crosslinking. The four 8-W germicidal lamps (e.g., G8T5, Herolab, Germany) have a dimension of 1.5 × 28.5 cm each. We used glass dishes with an inner diameter of 3.5 cm. The solution depth should be 1 mm. Dishes of different sizes can be used according to the sample volume. For details, see text.

5. Glass dishes with a planar surface and an inner diameter, for example, of 3.5 or 12.5 cm (according to sample volume).
6. Ethanol p.a. (per analysis) grade (Merck, Darmstadt, Germany).
7. Depending on scale: Standard Eppendorf tubes (Eppendorf, Hamburg, Germany) or Corex centrifugation tubes (30 mL); Sorvall rotor HB-4/HB-6 (Kendro Laboratory Products, Asheville, NC) or equivalent.
8. Depending on scale: Sorvall Evolution RC centrifuge (Kendro Laboratory Products) or cooled tabletop centrifuge (e.g., Heraeus Biofuge fresco; Kendro Laboratory Products) or equivalent.
9. Buffer 1: 6 *M* guanidinium hydrochloride (GHCl), 50 m*M* Tris/HCl, pH 8.0.
10. Buffer 2: 50 m*M* Tris/HCl, pH 7.5, 150 m*M* NaCl, 5 m*M* ethylenediaminetetraacetic acid (EDTA).
11. Endoproteinases trypsin (Promega, Madison, WI), Lys-C (Roche, Mannheim, Germany), Glu-C (Roche), chymotrypsin (Roche).
12. RNasin (Promega).
13. Equipment for size-exclusion (SE) chromatography (e.g., SMART System or FPLC System/Äkta Purifier system, Amersham Biotech/GE Healthcare, Uppsala, Sweden).
14. S75 HR SE column (3.1 × 300 m*M* or 10 × 300 m*M*, Amersham Biotech/GE Healthcare).

15. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) equipment and chemicals for silver staining according to **ref. *13***.
16. SDS sample buffer: 60 m*M* Tris-HCl, pH 6.8, 1 m*M* EDTA, 16% glycerol, 2% SDS, 0.1% bromophenol blue (BPB), 50 m*M* dithiothreitol (DTT).
17. Buffer 4: 50 m*M* Tris-HCl, pH 7.5, 2 m*M* EDTA.
18. Ribonucleases A and T1 (Ambion, Austin, TX).
19. Equipment for high-performance liquid chromatography (HPLC; e.g., SMART System; Microgradient system 140C, Applied Biosystems, Foster City, CA).
20. C18 reversed-phase (RP) column (2.1 × 150 m*M*; GraceVydac, Hesperia, CA) or 0.32 × 150 m*M* (Micro-Tech Scientific, Vista, CA).
21. HPLC solvent A: Water containing 0.1% (v/v) trifluoroacetic acid (TFA; Fluka, Buchs, Switzerland). Solvent B: Acetonitrile (ACN, Merck Darmstadt, Germany, LiChrosolv grade) containing 0.085% (v/v) TFA.
22. MALDI (matrix-assisted laser desorption/ionization) MS equipment (e.g., Bruker Reflex series, Ultraflex series, Applied Biosystems Voyager series, API 4700/4800 series).
23. DHB (2,5-dihydroxybenzoic acid; Sigma-Aldrich, Steinheim, Germany).
24. Matrix solution 1: 10 mg/mL solution of DHB in 50% ACN containing 0.1% TFA.
25. THAP (2′,4′,6′,-trihydroxyacetophenone; Fluka, Buchs, Switzerland).
26. Matrix solution 2: 10 mg/mL solution of THAP matrix in 50% ACN containing 0.5% TFA,
27. TFA (Fluka, Buchs, Switzerland).
28. Peptide standard calibration mixture for MALDI MS (e.g., Bruker Daltonics, Bremen, Germany).

3. Methods

The strategy we use for the identification of contact sites in native or in vitro reconstituted protein–RNA complexes is comprised of the following:

1. UV crosslinking of the complexes at 254 nm to generate a covalent bond between the protein and its cognate RNA at their site of interaction.
2. Purification of covalently linked peptide–RNA oligonucleotide heteroconjugates from the UV-irradiated complexes: SE chromatography; RP-HPLC.
3. Analysis of the purified heteroconjugates to identify the crosslinked protein region (including the actual crosslinked amino acid) and the crosslinked RNA region (including the actual crosslinked nucleotide): MALDI time-of-flight (TOF) MS.

3.1. General Remarks

Certain critical points have to be considered when one performs UV crosslinking and the subsequent identification of protein–RNA contact sites at the molecular level.

3.1.1. Amount and Concentration of the Particles

3.1.1.1. Native RNP Particles

Owing to the relatively low yield of UV crosslinking at 254 nm (*see* **Subheading 3.1.3.**, we recommend using a starting amount of RNP particles that is sufficient to allow purification and subsequent analysis of peptide–RNA oligonucleotide heteroconjugates. When one is using a capillary RP-HPLC system (*see* **Subheading 3.4.2., step 5**) with an RP column with an inner diameter of 300 µm for the final step in the purification of covalently linked peptide–RNA oligonucleotides, 10–50 µg of purified or in vitro reconstituted RNP complexes should be used as starting material. The native RNP particles are typically adjusted to a concentration of 0.1 mg/mL before crosslinking.

3.1.1.2. In Vitro Reconstituted Particles

Our purification procedure can also be applied to the analysis of RNP particles assembled in vitro from a known number of defined complexes. RNA can be most easily transcribed in vitro by phage RNA polymerase ***(14,15)*** or by chemical synthesis. Proteins can be produced by recombinant techniques ***(16)*** in *Escherichia coli*, yeast, or insect cells, or they can be purified from any other available biological source. The starting concentration of reconstituted particles is in the same range as for native ones. However, a number of considerations have to be taken into account in work with reconstituted particles. First, reconstitution in vitro can result in artificial crosslinking events when the assembly is incomplete or nonspecific. Incomplete assembly on the RNA can be minimized by using an excess of protein over RNA. In addition, nonspecific crosslinking can be due to the lack of additional specific protein components. An example for this is the crosslinking of the highly conserved Sm proteins, which are assembled in vitro on their cognate U snRNA Sm site ***(17,18)***. UV irradiation studies with radioactively labeled Sm site RNA and only three of the seven proteins (Sm G, E, F) revealed numerous crosslinks between these proteins and the RNA ***(2)***. However, only when all seven Sm proteins are fully assembled on the Sm site RNA is the crosslinking pattern of the proteins similar to that observed in native particles.

3.1.2. Buffer Systems

In general, any buffer systems are suitable, with the proviso that the buffer should not contain substantial concentrations of reagents that are known to scavenge radicals, for example, glycerol with a concentration of >15% (v/v). UV crosslinking is a radical reaction, and radical scavengers therefore drastically reduce its yield.

3.1.3. UV Crosslinking

UV crosslinking at 254 nm is a straightforward technique to detect direct protein–RNA interactions. It generates a covalent bond between an amino acid side chain of the protein and a base of the RNA when both are in a favorable position.

Putative UV-crosslinkable amino acids are tyrosine, histidine, phenylalanine, leucine, and cysteine *(3)*. The most UV-reactive base is uridine *(19)*. Note that not all proteins that are tightly associated with RNA can be UV crosslinked. An example of this is the human spliceosomal protein 15.5 K bound to U4/U4atac snRNA *(20)*. The crystal structure of protein 15.5 K in complex with the 5′ stem-loop of U4 snRNA shows that this protein interacts almost exclusively with a purine-rich internal loop (AAU) within the 5′ stem-loop of U4 snRNA *(21)*. Although a variety of amino acids within the protein interact through hydrogen bonds and hydrophobic interactions with the bases of the asymmetric internal loop of the RNA, neither of these amino acids contains a crosslinkable side chain. In addition, corresponding regions within the protein and the RNA must have enough flexibility to allow the formation of a new covalent bond between the components. At the RNA level, this is only the case when the base of the RNA is not involved in basepairing. Correspondingly, the most flexible—and thus the most easily crosslinked—areas within proteins are the loop regions *(2,22)*.

Despite the fact that the yield of UV crosslinking at 254 nm is relatively low when compared with that of the crosslinking of in vitro reconstituted particles bearing an artificially introduced crosslinking label on the RNA, it has the advantage that it can be applied to native purified RNP particles that have been isolated directly from cellular compartments. It further avoids generation of the "false-positive" results that are frequently associated with the heterogeneous populations generated as a result of incomplete assembly in the reconstitution reaction with labeled RNA. Nonetheless, the approach of UV crosslinking at 254 nm followed by identification of the protein–RNA crosslinking sites has been also successfully applied to protein–RNA complexes reconstituted in vitro *(20)*.

3.1.4. Crosslinking Conditions

UV crosslinking is performed in a flat glass dish under a suitable light source. The result is strongly influenced by the choice of lamp, the distance between the lamp and the sample, and the irradiation time. Our laboratory uses a custom-made UV irradiation device (*see* **Fig. 1** and **Subheading 3.2.**). Alternatively, one can use a commercially available UV irradiation apparatus (e.g., UV Stratalinker 2400). Importantly, the conditions of UV irradiation—in particular the irradiation time and the distance of the sample from the lamp—have to be

adjusted. During our studies with snRNP particles, we have observed that 2 min of irradiation is sufficient to achieve a maximum crosslinking yield. In contrast, for more rigid RNP particles such as ribosomes, the irradiation time can be prolonged. The samples are irradiated in custom-made glass dishes with a planar surface. Precooling of the glassware is essential. The depth of the sample solution is ideally 1 mm, and the size of the glass dish or the sample volume should be adjusted accordingly.

3.2. UV Crosslinking Protocol

The first step of purification of covalently linked peptide–RNA oligonucleotides is the UV irradiation of the samples to form a covalent linkage between a protein and the RNA.

1. Starting materials are native purified snRNPs (for example, 17 S U1 snRNP, or 25 S [U4/U6.U5] tri-snRNP) or RNP particles reconstituted in vitro, such as [U4/U6-15.5 K-61 K]. The purification of native U snRNPs and the reconstitution in vitro of [U4atac-15.5 K-61 K] protein–RNA complexes are described in detail elsewhere ***(4,8–10)***. The concentration of the RNP particles is adjusted to approximately 0.05–0.1 mg/mL with the buffer in which they were purified or reconstituted (*see* **Note 1**).
2. The samples are pipeted into precooled glass dishes with a typical diameter of 3.5 cm. The glass dishes must be flat (to provide for a homogeneous layer of liquid), and they should be placed on an aluminum block in ice (**Fig. 1**; *see* **Note 2**).
3. The samples are irradiated for 2 min at a distance of 2 cm from the UV source. We use a custom-made UV irradiation device with four 8-W germicidal lamps (G8T5, Herolab, Germany) mounted in parallel (**Fig. 1**).
4. The samples are pooled and precipitated with 3 volumes of ice-cold ethanol (p.a. grade) in the presence of 1/10 volume of 3 *M* sodium acetate, pH 5.3, for at least 2 h at −20 °C. They are then centrifuged for 30 min at 4 °C in a tabletop centrifuge (13,000 rpm [16,060 *g*]; Eppendorf tubes) or in a HB-4/HB-6 rotor (10,000 rpm [16,340 *g*]; Corex tubes). The pellets are washed with an appropriate volume of ice-cold 80% (v/v) ethanol/water and spun down as before. If in Eppendorf tubes, they are then dried briefly *in vacuo* (for not more than 5 min); if in Corex tubes, they are dried in air on the laboratory bench for 10–20 min (*see* **Note 3**).

3.3. First Separation Step: Size-Exclusion Chromatography

Size-exclusion chromatography ("gel filtration" SEC) is the first step in the purification of crosslinked peptide–RNA oligomers. The principle is illustrated in **Fig. 2**. UV-irradiated RNP complexes are allowed to dissociate, and the protein moiety is hydrolyzed with endoproteinases to yield intact RNA that still carries specific peptides covalently attached to the RNA at the sites of crosslinking. Noncrosslinked RNA and RNA with crosslinked peptides can be separated from the noncrosslinked peptides by SEC. This step is critical since any residual noncrosslinked peptides in the RNA-containing fractions will lead to a

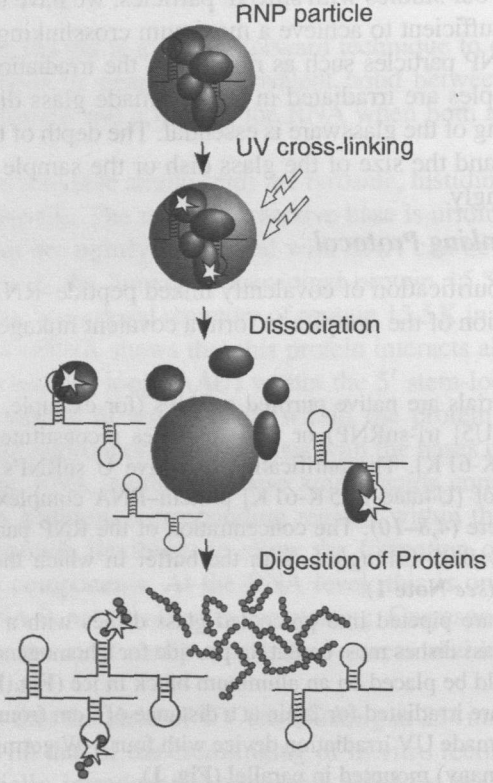


Fig. 2. Schematic representation of the initial step in the purification of crosslinked peptide–RNA oligonucleotides. After ultraviolet (UV) irradiation of ribonucleoprotein (RNP) particles, the particles are dissociated, and the protein moiety is digested with endoproteinases to obtain intact RNA that carries crosslinked peptides (white stars indicate the site of crosslinking) and noncrosslinked peptides. See text for details.

complex elution pattern in the final separation by RP-HLPC (*see* **Subheading 3.4.1.**) and to additional complexity of the mass spectra obtained in the analysis of the crosslinks.

3.3.1. General Remarks

3.3.1.1. Size-Exclusion Columns

Size-exclusion chromatography can only be applied in a first separation step when the RNA is significantly larger than the average size of the peptides generated (10–30 amino acids, i.e., 1000–3000 Da). For example, RNA oligomers (≤30 nucleotides) cannot be separated from peptides. The choice of the SE column is

important as separation is influenced by its size (length and diameter) as well as by the matrix. The smallest column for our purpose is a Superose 75 HR column, measuring 3.2 × 300 mm, mounted in a SMART System. The maximum sample volume that can be applied to these is 50 µL. Smaller SE columns will usually have a smaller maximum sample volume and will therefore not be usable as it is difficult to obtain samples below 50 µL by the procedures described above. For large-scale preparations, we use a Superose 74 HR measuring 10 × 300 mm, mounted in a standard FPLC system or in the Äkta Purifier system. The maximum sample volume that can be applied on this column is 200 µL.

3.3.1.2. Dissociation Conditions

Dissociation and the subsequent digestion of the RNP particles with endoproteinases must be complete. Incomplete digestion will result in either larger peptide fragments or in only partially hydrolyzed proteins, which will comigrate with the RNA and will thus interfere with the final detection of peptide–RNA oligonucleotide crosslinks during RP-HPLC (*see* **Subheading 3.4.2.**). Incomplete digestion is a particular danger when proteins are tightly associated with their cognate RNAs (e.g., Sm proteins bound to the Sm site RNA in U snRNAs; ***16,17***).

For dissociation/denaturation and subsequent digestion of the irradiated particles, SDS, urea, or GHCl can be used according to the conditions specified for each endoproteinase. In our hands, best results are obtained by dissociation/denaturation in 6 *M* GHCl at room temperature for reconstituted particles and at elevated temperature (90 °C) for native particles.

SDS has the disadvantage that it prevents endoproteinases, such as chymotrypsin or trypsin, from working properly, even at very low SDS concentrations. For example, when we used trypsin in the presence of 0.1% (w/v) SDS to generate peptides, we noticed that a considerable number of cleavage sites were missed.

The use of 6–8 *M* urea for dissociation/denaturation can cause carbamylation of the lysine residues due to the presence of traces of cyanate in the urea; carbamylated lysine is no longer a substrate for trypsin, and most importantly, carbamylation causes a mass shift of 43.006 (monoisotopic mass) in the mass spectrometer. Such modification has to be considered in calculating the cross-linked peptide and RNA moiety from a measured mass (*see* **Subheading 3.5.3.**).

3.3.1.3. Digestion Conditions

It must be emphasized that the choice of the endoproteinase is also critical for the identification of the cross-linked protein region. If the sequence of the crosslinked protein is known, then one should perform a theoretical digestion of the protein with various endoproteinases and check that reasonable peptides

(i.e., 10–20 amino acids) are expected; the endoproteinase should then be chosen accordingly. When one is working with noncharacterized particles, it is difficult to predict which endoproteinase will generate crosslinked peptides of reasonable size. It can happen that a crosslinked peptide cannot be identified because it is too small, is too large, does not elute from the HPLC column, and so on. Thus, a negative result does not necessarily mean that no crosslinking occurred. Changing the endoproteinase can often lead to positive results.

The following endoproteinases are routinely used for digestion of crosslinked proteins: trypsin (sequencing grade, Promega), chymotrypsin (sequencing grade, Roche), Lys-C (Roche), and Glu-C (sequencing grade, Roche). Lys-C has the advantage that it generates larger peptides than does trypsin. Larger, crosslinked peptides elute at a higher percentage of solvent B (water, ACN, 0.085% TFA) from the RP-HPLC column in the final purification step (*see* **Subheading 3.4.2., step 5**), whereas tryptic, crosslinked fragments coelute at lower percentages of buffer B together with larger, noncrosslinked RNA oligomers. As a rule of thumb, we recommend starting the first experiment with trypsin. If no crosslinks can be purified, then one should resort to one of the other endoproteinases.

3.3.1.4. SDS-PAGE Analysis of Crosslinks

We strongly recommend analyzing aliquots of the fractions after SE chromatography by SDS-PAGE *(23)* and visualization of their components by silver staining *(13)*. **Figure 3A** shows a typical SE elution profile of UV-irradiated RNPs treated with endoproteinase (in this case, irradiated U1 snRNPs digested with trypsin in the presence of GHCl; *see* **Subheading 3.3.2., step 1**), and **Fig. 3B** shows the corresponding SDS-PAGE analysis. The RNA-containing fractions contain only U1 snRNA and crosslinked U1 snRNA. The endoproteinase trypsin is detected in fractions that are eluted later, and—importantly—no larger protein fragments due to incomplete digestion are detectable. A similar pattern is observed in reconstituted [15.5 K-61 K-U4atac] RNPs after digestion with trypsin in the presence of GHCl (**Fig. 3C**). In contrast, digestion of the same RNP with chymotrypsin in the presence of 0.1% (v/v) SDS leads to incomplete digestion, as revealed by the presence of numerous larger protein fragments that coelute with U4atac snRNA (**Fig. 3D**).

3.3.2. SE Chromatography Protocol

1. The precipitated UV-irradiated samples are dissolved in an appropriate volume (typically 50 µL) of buffer 1 and heated to 90 °C for 5 min. They are then diluted to give a final concentration of 1 *M* GHCl with buffer 1 without GHCl. Endoproteinase is added to a final concentration of 1:20 (w/w) enzyme/substrate ratio, and the mixtures are incubated overnight at 37 °C in the presence of 40 U RNasin (Promega).

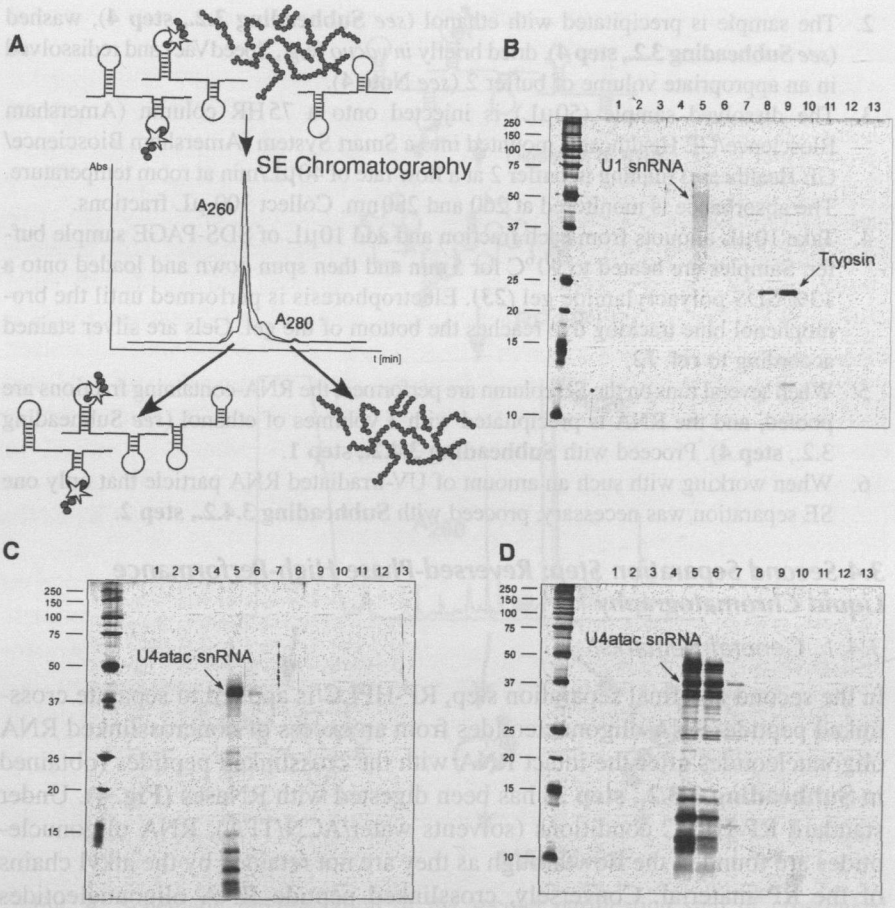


Fig. 3. (**A**) Schematic representation of the first purification step of crosslinked peptide–RNA oligonucleotides by size-exclusion (SE) chromatography. SE chromatography separates RNA molecules—both with and without crosslinked peptides—from the excess of noncrosslinked peptides. (**B**) Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of fractions derived from SE chromatography of ultraviolet (UV)-irradiated U1 small nuclear RNPs (snRNPs) dissociated in the presence of 6*M* guanidinium hydrochloride (GHCl) and digested with trypsin. U1 snRNA and trypsin are visualized by silver staining ***(12)***. Note that except for trypsin no additional proteins (or large protein fragments) are identified. (**C**) SDS-PAGE analysis of fractions derived from SE chromatography of UV-irradiated [U4atac-15.5 K-61 K] RNP dissociated in the presence of 6 M GHCl and treated with chymotrypsin. (**D**) SDS-PAGE analysis of fractions derived from SE chromatography of UV-irradiated [U4atac-15.5 K-61 K] RNP dissociated in the presence of 1% SDS and treated with chymotrypsin after adjustment of the SDS concentration to 0.1%. The position of U4atac is indicated. Silver staining of SE fractions shows that the RNA-containing fractions still include numerous protein fragments compared with the products of digestion in the presence of GHCl.

2. The sample is precipitated with ethanol (*see* **Subheading 3.2., step 4**), washed (*see* **Subheading 3.2., step 4**), dried briefly *in vacuo* (e.g., SpeedVac) and redissolved in an appropriate volume of buffer 2 (*see* **Note 4**).
3. The dissolved sample (50 µL) is injected onto a 75 HR column (Amersham Bioscience/GE Healthcare) mounted into a Smart System (Amersham Bioscience/GE Healthcare) running in buffer 2 at a flow rate of 40 µL/min at room temperature. The absorbance is monitored at 260 and 280 nm. Collect 100-µL fractions.
4. Take 10 µL aliquots from each fraction and add 10 µL of SDS-PAGE sample buffer. Samples are heated to 90 °C for 5 min and then spun down and loaded onto a 13% SDS polyacrylamide gel *(23)*. Electrophoresis is performed until the bromophenol blue tracking dye reaches the bottom of the gel. Gels are silver stained according to **ref.** *12*.
5. When several runs on the SE column are performed, the RNA-containing fractions are pooled, and the RNA is precipitated with 3 volumes of ethanol (*see* Subheading 3.2., **step 4**). Proceed with **Subheading 3.4.2., step 1**.
6. When working with such an amount of UV-irradiated RNA particle that only one SE separation was necessary, proceed with **Subheading 3.4.2., step 2**.

3.4 Second Separation Step: Reversed-Phase High-Performance Liquid Chromatography

3.4.1. General Remarks

In the second and final separation step, RP-HPLC is applied to separate crosslinked peptide–RNA oligonucleotides from an excess of noncrosslinked RNA oligonucleotides after the intact RNA with the crosslinked peptides (obtained in **Subheading 3.3.2., step 3**) has been digested with RNases (**Fig. 4**). Under standard RP-HPLC conditions (solvents water/ACN/TFA), RNA oligonucleotides are found in the flowthrough as they are not retained by the alkyl chains of the RP material. Conversely, crosslinked peptide–RNA oligonucleotides are retained because of their peptide moiety and are eluted at a higher percentage of solvent B (80% ACN, 0.085% TFA in water). Peptide–RNA crosslinks are detected by monitoring the absorbance at 220 nm (peptide) and 260 nm (nucleotides). Fractions that show an absorbance at 220 and 260 nm (due to the crosslinked RNA moiety) are collected and are further analyzed by MALDI-TOF MS. **Figure 4** shows a schematic representation of the final purification of peptide–RNA oligonucleotide crosslinks with an RP-HPLC example derived from [U4atac-15.5 K-61 K] RNPs after digestion of the complexes with chymotrypsin and RNases A and T1. Within the chromatogram, two fractions that show an absorbance at 220 and 260 nm could be detected and were thus chosen for further analysis in the MALDI-TOF mass spectrometer (Bruker Daltonics).

The successful detection of peptide–RNA oligonucleotides in RP-HPLC depends not only on the items discussed (i.e., starting material, crosslinking

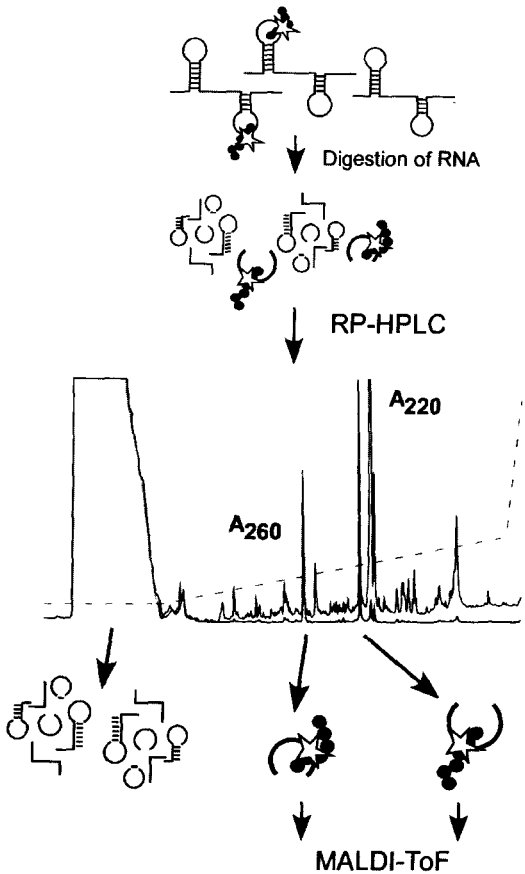


Fig. 4. Schematic representation of the second purification step of crosslinked peptide–RNA oligonucleotides by reversed-phase high-performance liquid chromatography (RP-HPLC). RNA-containing fractions from the size-exclusion (SE) chromatography are pooled, and the RNA is hydrolyzed with ribonucleases (RNase A or RNase T1). The mixture, which consists of noncrosslinked RNA oligonucleotides and crosslinked peptide–RNA oligonucleotides, is injected onto an RP-HPLC column. The noncrosslinked RNA oligonucleotides are found in the flowthrough, while crosslinked peptide–RNA oligonucleotides are retained and are only eluted at a higher concentration of solvent B (dashed line). Fractions that show an absorbance at 220 and 260 nm due to the crosslinked RNA oligonucleotide are analyzed in the MALDI-TOF (matrix-assisted laser desorption/ionization time-of-flight) mass spectrometer.

yield, choice of endoproteinase, etc.) but also on the hardware components used in this final purification step.

In principle, any HPLC system can be used for separation if it is equipped at least with a dual-wavelength recorder (to monitor for peptides at 214–220 nm and RNA at 254–260 nm). The HPLC system should be equipped with a standard peptide C18 column. In our hands, C18 material with a pore size of 300 Å and a particle size of 300 µm works best. The length of the column is not critical. The most important point is the sensitivity of the detection. This depends on the size of the RP column (e.g., for conventional HPLC 4.6-mm inner diameter, for analytical HPLC 2.1-mm inner diameter, for micro-HPLC 0.5- to 1.0-mm inner diameter, for capillary HPLC 0.1- to 0.5-mm inner diameter, and for nano-HPLC ≤ 0.1-mm inner diameter). For example, we have analyzed preparative amounts of UV-irradiated prokaryotic ribosomal subunits with a conventional HPLC *(24)*, preparative amounts of UV-irradiated U1 snRNP or 25 S [U4/U6.U5] tri-snRNPs on an analytical HPLC *(1)*, and small amounts (≤20 µg) of U1 snRNP1 and in vitro reconstituted [U4atac-15.5 K-61 K] RNPs on a capillary HPLC.

The purity and homogeneity of the sample also have a strong influence on the final separation and detection of the crosslinked peptide–RNA. Incomplete digestion of the protein moiety, as mentioned, will lead to a crowded chromatogram, in which the crosslinks coelute with noncrosslinked peptides and thus impede detection. Incomplete hydrolysis of the RNA prior to HPLC separation also drastically impairs the detection of crosslinks. Nonhydrolyzed RNA or larger RNA oligomers elute at similar concentration of solvent B (8–15% v/v ACN, 0.08% v/v TFA in water) as crosslinks (12–20% v/v ACN, 0.08% v/v TFA in water). Owing to the very strong absorbance of the RNA at 254 nm in this particular region of the chromatogram, the actual crosslink cannot be detected.

3.4.2. RP-HPLC Protocol

1. When samples have been pooled and precipitated (*see* Subheading 3.3.2., **step 5**), after SE chromatography they are centrifuged, washed, and dried as described in **Subheading 3.2., step 4**. For RNA hydrolysis, these are dissolved in an appropriate volume of buffer 4, and 10 µg of RNase A or 15 µg of RNase T1 (Ambion) are added (*see* **Note 5**).
2. Alternatively, samples can be treated with 1 µg RNase A or 2 µg T1 in buffer 2 (*see* **Subheading 3.3.2., step 1**, and **Note 5**)
3. Incubation is carried out for 2 h at 52 °C.
4. Add 1 µg of endoproteinase (the same as the one used in **Subheading 3.3.2.**), and the mixture is incubated overnight at 37 °C (*see* **Note 6**).
5. The sample is injected onto an RP-HPLC column mounted in a corresponding HPLC system. We use the following systems routinely in our laboratory:

a. For UV-irradiated RNP particles 50 µg or larger: C18 RP column, 150 × 2.1 mm (218TP5215; GraceVydac) mounted in a SMART System (Amersham Biotech/GE Healthcare) equipped with a 10 µL flow cell running at a flow rate of 100 µL/min.
b. For UV-irradiated RNP particles smaller than 50 µg: C18 RP column, 150 × 0.3 mm (Micro-Tech Scientific) mounted in an Applied Biosystems Microgradient system 140C equipped with a 35 nL flow cell running at a flow rate of 2 µL/min.
c. The solvents are as follows: For the SMART System, solvent A is water containing 0.1% (v/v) TFA; solvent B is ACN containing 0.085% (v/v) TFA. For the Cap-LC system, solvent A is water containing 0.1% (v/v) TFA, and solvent B is 80% ACN containing 0.085% (v/v) TFA.
d. The following gradients are applied: For the SMART System, isocratic elution at 5% solvent B until after elution of the injection peak (i.e., until the UV trace returns to the baseline); 5% solvent B to 45% solvent B in 120 min; 45% solvent B to 90% solvent B in 10 min; 10 min at 90% solvent B. It may be necessary to include another isocratic step around 10% solvent B if a constant and strong increase in the absorbance at 260 nm is observed at that percentage (indicating incomplete RNA hydrolysis or larger RNA oligonucleotides). Importantly, it should be borne in mind that this value refers to the theoretical concentration of solvent B in our SMART System, which is equipped with an analytical column. The value might vary in different HPLC systems because of different delay volumes. For the Cap-LC system, isocratic elution at 5% solvent B until after elution of the injection peak (i.e., until the UV trace returns to the baseline); 5% solvent B to 60% solvent B in 60 min, 60% to 90% solvent B in 3 min, 7 min at solvent 90% B.

6. Fractions that show an absorbance at 220 and 260 nm are collected, dried completely *in vacuo*, and stored for further at analysis –20 °C.
7. Proceed with **Subheading 3.5.2., step 1**.

3.5. Mass Spectrometry

3.5.1. General Remarks: MALDI-TOF vs Electrospray Ionization

Mass spectrometry is by now the fastest and most sensitive method for sequence analysis of biomolecules ***(25–28)***. In particular, proteins and peptides including (posttranslational) modifications can be sequenced on an almost-routine basis. In addition, MS has been successfully applied for the sequencing of RNA/DNA molecules ***(29–31)***. MS uses two main methods to ionize biological macromolecules: electrospray ionization (ESI; ***32***) and MALDI ***(33,34)***. The principal difference between ESI and MALDI is that ESI is a continuous ionization method by which multiply charged biomolecules (e.g., $[M+2H]^{2+}$ to $[M+nH]^{n+}$) are produced from a capillary electrode placed at high voltage with respect to a grounded counterelectrode. In contrast, MALDI is a pulsed ionization technique

in which the biomolecules in a UV-absorbing matrix (e.g., nicotinic acid) are desorbed from/with the matrix molecules by a laser pulse, finally becoming mainly singly charged (e.g., $[M+H]^+$) by charge transfer from the matrix molecules to the analyte molecules.

The latter has the advantage that the resulting mass spectra are relatively easy to interpret as they do not show several mass peaks from one species. Moreover, in terms of sample preparation and instrumentation, MALDI MS is much easier to handle than the electrospray technique. On the other hand, sequence information, especially from modified peptides/biomolecules, is more reliable when derived from ESI MS combined with collision-induced decay *(35)*. In contrast, sequencing of unknown or modified peptides (i.e., *de novo* sequencing) in a MALDI-TOF mass spectrometer by postsource decay (PSD; *36,37*) has several disadvantages: (1) It requires sufficient amounts of sample, (2) sequencing is often restricted to a certain size of the fragment selected for sequencing (1000–2000 m/z), and (3) much experience in spectrum evaluation and interpretation—or in bioinformatics—is required to deduce the peptide sequence from the mass spectrum.

Notably, MS of peptide–RNA oligonucleotides differs from the routine MS analysis of modified peptides or of pure oligonucleotides as peptides and RNA (oligomers) behave differently in MS. RNA is analyzed under conditions by which it becomes negatively ionized by changing the polarity of the ESI and MALDI mass spectrometer in combination with the use of a neutral, volatile buffer (e.g., ammonium acetate) and an RNA-compatible matrix (e.g., 3-hydroxy-picolinic acid; Sigma-Aldrich), respectively. Conversely, peptides are analyzed under acidic conditions, at which they are positively charged.

Despite these disadvantages, we have chosen MALDI-TOF instead of ESI (on a conventional mass spectrometer with ESI source, as triple-quadrupole instruments, Q-ToF instruments, linear and three-dimensional ion traps) for the analysis of purified peptide–RNA heteroconjugates. This was for the following reasons: (1) MALDI-TOF is highly sensitive; (2) it has an extremely high mass accuracy ($\leq$ 10 ppm, which is important for calculation of possible composition and sequence of the peptide and RNA moieties from the measured mass; *see* **Subheading 3.5.3.**) with external calibration; (3) samples can be easily prepared; and most important, (4) it is much easier to find conditions under which a sufficient number of ionized crosslinked peptide–RNA species travels through the analyzer toward the detector, especially when they have high m/z values.

Note that a successful analysis is only achieved when the samples are measured in the reflectron mode of the MALDI-TOF instrument, as only then can monoisotopic masses be considered. For measuring in the reflectron mode, the use of DHB or THAP as matrices is recommended since α-cyano-4-hydroxycinnamic acid (HCCA) gives poor or no results. Using

HCCA as a matrix for peptide–RNA oligonucleotides allows MS analysis in the instrument's linear mode only, and the mass accuracy in the linear mode is too low to distinguish unambiguously between the bases C and U, which differ by only 1 mass unit.

3.5.2. Preparation of Samples and Acquisition of MS Spectra

1. Dried fractions from the RP-HPLC separation are dissolved in not more than 10 µL 50% ACN (v/v), 0.1% TFA, and sonicated in a sonication bath for 3 min.
2. Follow these procedures next:
   a. 0.5 µL of the sample is mixed on a stainless steel sample plate with the same volume of matrix solution 1. The preparation is air dried and subjected to MS in a Voyager DE-STR (Applied Biosystems) under standard conditions in the reflectron mode. The acceleration voltage is 20 kV, the grid voltage 68%, and the delay time 250 ns. A total of 300 laser shots (N_2-pulsed laser, 20 Hz, 337 nm) are summed.
   b. Alternatively, 0.5 µL of the sample is mixed on a stainless steel sample plate with the same volume of matrix solution 2. The preparation is air dried and subjected to MS in a Voyager DE-STR under standard conditions in the reflectron mode. The acceleration voltage is 20 kV, the grid voltage 68%, and the delay time 250 ns. A total of 300 laser shots (N_2-pulsed laser, 20 Hz, 337 nm) are summed.
   c. We also have good experience with Bruker MALDI-TOF mass spectrometers (Bruker Daltonics). Bruker instruments are equipped with different sample plates, which in turn require a modified sample preparation. When DHB is used as the matrix, 0.5 µL of the sample is mixed on a Bruker Anchor 600 sample plate with the same volume of a 10 mg/mL solution of DHB in 50% ACN or water, respectively. The preparation is air dried and measured in a Reflex IV under standard conditions in the reflectron mode. The acceleration voltage is 20 kV for the IS1 and 16.9 kV for the IS2; the delay time is 400 ns. A total of 300 laser shots (N_2-pulsed laser, 5 Hz, 337 nm) are summed.

3.5.3. Spectrum Evaluation

It is often straightforward to identify the crosslinked peptide and oligonucleotide parts from MALDI MS spectra of purified peaks. This is exemplified in the following case:

Figure 5A shows a MALDI-TOF spectrum from a purified crosslink derived from UV-irradiated [U4atac-15.5 K-61 K] RNPs after digestion with chymotrypsin and RNase T1. Notably, DHB and THAP preparations yielded a resolution higher than 15,000 and mass accuracies of 30 ppm or better by using close external calibration.

To determine the peptide and RNA compositions, we compared all chymotryptic fragments of the protein moiety with all RNase T1 fragments of the RNA moiety. We found that the measured monoisotopic mass ($[M+H]^+ = 2780.818$) matches precisely only a single chymotryptic fragment of the 61 K protein, encompassing protein positions 263–273 (SSTSVLPHTGY, $[M+H]^+_{cal} = 1148.558$) and to

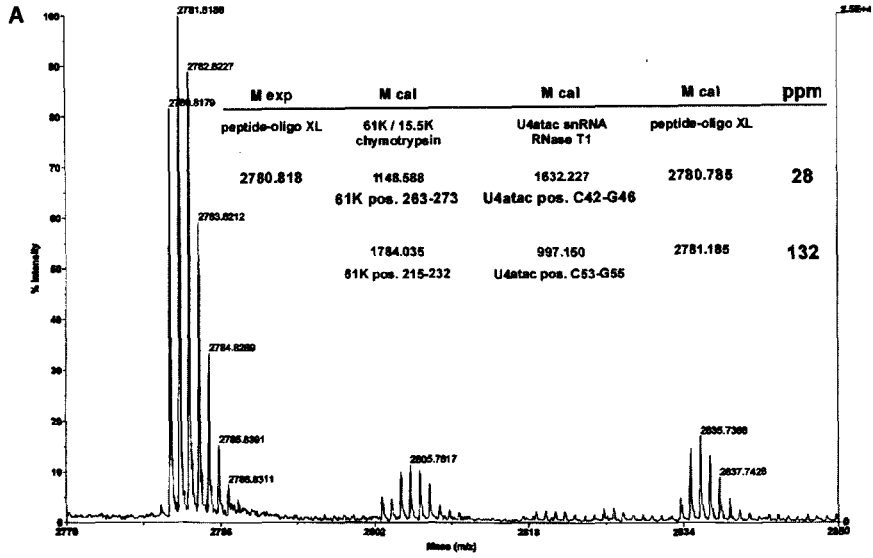


Fig. 5. MALDI-ToF (matrix-assisted laser desorption/ionization time-of-flight) mass spectra from purified peptide–RNA oligonucleotide crosslinks. (**A**) MALDI mass spectrometric (MS) spectrum of a purified crosslink derived from ultraviolet (UV)-irradiated [U4atac-15.5 K-61 K] RNP (ribonucleoprotein) treated with endoproteinase chymotrypsin and ribonuclease (RNase) T1. The spectrum was recorded on a Voyager DE-STR MALDI mass spectrometer (Applied Biosystems, Foster City, CA) under standard conditions. The measured monoisotopic mass (M_{exp}) of 2780.818 matches precisely a peptide of protein 61 K encompassing positions 263–273 (SSTSVLPHTGY, $[M+H]^+_{cal}$ = 1148.558) and to an RNase T1 fragment of U4atac RNA between nucleotide positions C42 and G46 (5′-CAUAG-3′, M_{cal} = 1632.227), M_{cal} = 1148.558 + 1632.227 = 2780.785. Notably, the mass deviation is less than 20 ppm ($M_{exp} - M_{cal}$ = 2780.818 – 2780.785 = 0.033 amu). The next-best hit from our database search was a chymotryptic fragment of protein 61 K from positions 215–232 crosslinked to a U4atac snRNA T1 fragment from positions C53–G55. Since the mass deviation between experimental and calculated mass is more than 130 ppm, this was not considered to be the actual crosslinked peptide–RNA oligonucleotide. See text for further details.

an RNase T1 fragment of U4atac RNA between nucleotide positions C42 and G46 (5′-CAUAG-3′, M_{cal} = 1632.227), 1148.558 + 1632.227 = 2780.785. The mass accuracy was better than 20 ppm (i.e., ± 0.033 amu [absolute mass units]). The next best "hit" from this calculation (61 K positions 215–232, $[M+H]^+_{cal}$ = 1784.035 and U4atac nucleotide positions C53–G55, M_{cal} = 997.150) had a significantly higher mass deviation (130 ppm) and was thus not considered to represent the crosslinked complex. From this highly accurate

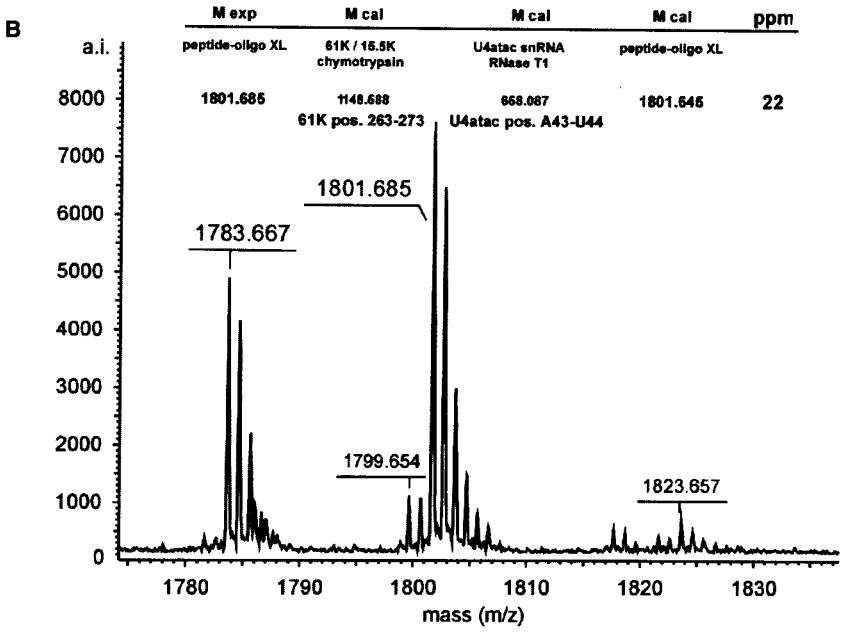
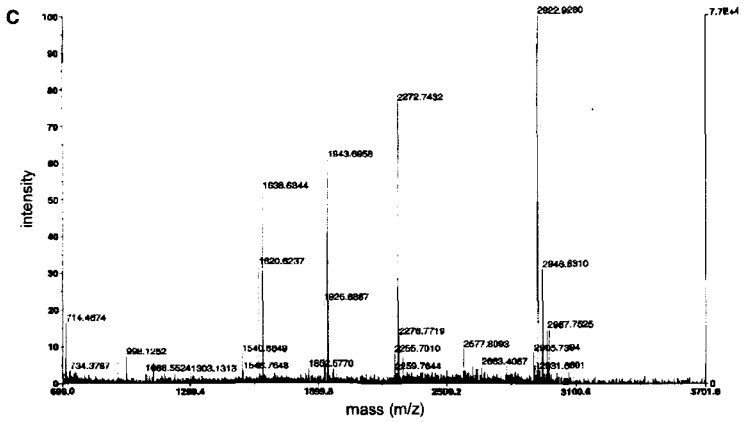


Fig. 5. (**B**) (continued) MALDI MS spectrum of a purified crosslink derived from UV-irradiated [U4atac-15.5 K-61 K] RNP treated with endoproteinase chymotrypsin and RNases A and T1. The spectrum was recorded on a Reflex IV MALDI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) under standard conditions. The measured monoisotopic mass (M_{exp}) of 1801.685 matches precisely a peptide of protein 61 K encompassing positions 263–273 crosslinked to an RNase A and T1 fragment comprising positions A43 and U44 in the 5′ stem-loop of U4atac snRNA (M_{cal} = 1801.645). As in **Fig. 4A**, the mass deviation is less than 30 ppm.

mass analysis, we conclude that the 61 K protein between residues 263 and 273 must be crosslinked to U4atac between nucleotide positions C42 and G46.

This is strongly supported by the MALDI analysis of a second crosslink derived from the same particles after digestion with chymotrypsin and RNases A and T1. Our highly accurate MALDI analysis revealed a mass peak of $[M+H]^+$ = 1801.685 (**Fig. 5B**), corresponding exactly to a peptide of protein 61 K, comprising positions 263–273 crosslinked to an AU dinucleotide. Notably, AU (A43 and U44, 5′-AU-3′, M_{cal} = 658.087) is located within the previously identified T1 fragment of U4atac.

Figure 5C gives a third example of the determination of the crosslinked peptide and RNA moieties from a single MS spectrum. The spectrum shows the MALDI MS analysis of a crosslink derived from native UV-irradiated U1



M exp	M cal	peptide	U1 snRNA oligo	ppm
2922.9280	2922.8469	U1 70K $_{173}$RVLVDVER$_{180}$	5' $_{29}$AUCACG$_{34}$ 3'	**28**
2577.8093	2577.7999	U1 70K $_{173}$RVLVDVER$_{180}$	5' $_{29}$AUCAC$_{33}$ 3'	**4**
2272.7432	2272.7589	U1 70K $_{173}$RVLVDVER$_{180}$	5' $_{29}$AUCA$_{32}$ 3'	**7**
1943.6958	1943.7069	U1 70K $_{173}$RVLVDVER$_{180}$	5' $_{29}$AUC$_{31}$ 3'	**6**
1638.6344	1638.6659	U1 70K $_{173}$RVLVDVER$_{180}$	5' $_{29}$AU$_{30}$ 3'	**19**

Fig. 5. (C) (continued) MALDI MS spectrum recorded under our highly accurate conditions from purified peptide–RNA oligonucleotide crosslinks derived from UV-irradiated U1 snRNP hydrolyzed with endoproteinase trypsin and RNase T1. The spectrum was recorded on a Voyager DE-STR MALDI mass spectrometer under standard conditions. The measured monoisotopic masses (M_{exp} = 2922.928, 2577.809, 2272.743, 1943.696, 1638.634) match precisely a tryptic fragment of the U1 protein 70K from positions 173–180 crosslinked to a U1 snRNA T1 fragment from positions 29–34 and to its 3′-hydrolysis products, respectively. Importantly, the mass deviation over the entire spectrum was less than 30 ppm.

snRNPs treated with trypsin and RNase T1. By comparing all possible tryptic fragments from the 10 U1 proteins (70K, U1A, U1C, and the Sm proteins B/B′, D1–D3, E, F, and G) with all possible U1 snRNA RNase T1 fragments, we found that the mass peaks correspond to a U1 70K tryptic fragment from positions 173–180 (RVLVDVER) crosslinked to a U1 snRNA T1 fragment encompassing positions 29–34 (5′-AUCACG-3′) and to the 3′ hydrolysis products of the latter. Notably, the mass accuracy over the entire spectrum is better than 30 ppm using close external calibration.

Our preparations (DHB and THAP) also showed excellent properties in the PSD analysis of selected peptide–oligonucleotide mass peaks, revealing structural information about the crosslinked peptide moiety. An example is given in **Fig. 5D**, which shows a PSD sequence analysis under standard conditions in a

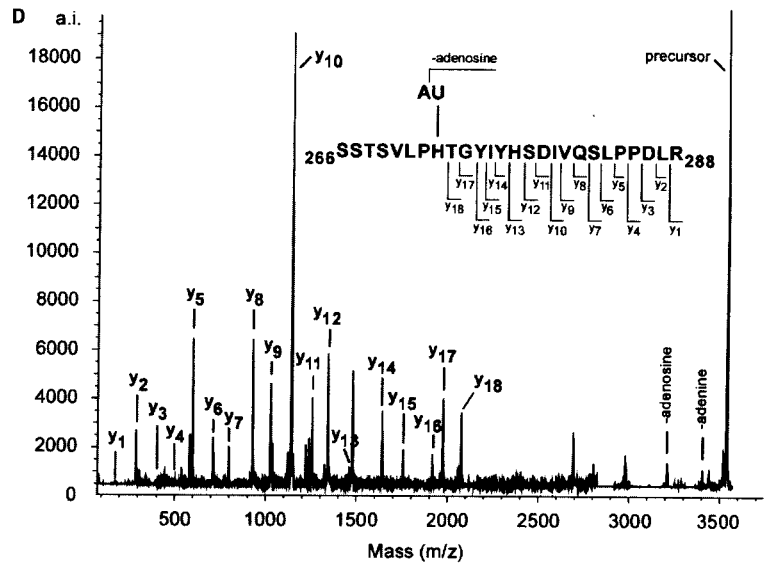


Fig. 5. (**D**) (continued) Postsource decay (PSD) spectrum of a selected crosslink (precursor mass m/z = 3535.489) derived from [U4atac-15.5 K-61 K] RNP treated with trypsin and RNases A and T1. The spectrum was recorded on a Reflex IV MALDI mass spectrometer. The sequence with the experimentally found y-type ion series is listed within the spectrum. Importantly, PSD sequence analysis revealed that His270 is the potentially crosslinked amino acid, and U44 rather than A43 within the identified T1/A fragment (5′-AU-3′) is the actual crosslinked nucleotide.

Reflex IV MALDI-TOF instrument (Bruker Daltonics). The species analyzed is a 61 K–U4atac crosslink (selected precursor mass m/z = 3535.489) derived from the UV-irradiated [U4atac-15.5 K-61 K] RNPs after digestion with trypsin and RNases A and T1. Importantly, the measured precursor mass of m/z = 3535.489 did not yield a specific tryptic 61 K fragment crosslinked to an RNase A/T1 fragment of U4atac snRNA in our database search. The PSD spectrum shows a complete series of C-terminal fragment ions (so-called y-type ions) up to y18, revealing the sequence TGYIYHSDIVQSPPDLR (**Fig. 5D**). In combination with the measured precursor mass, we thus identified the crosslink as a fragment of protein 61 K encompassing positions 263–288 (SSTSVLPHTGYIYHSDIVQSP PDLR) crosslinked to an AU dinucleotide (A43 and U44, 5′-AU-3′). The missing y19 (H) and y20 (P) indicate that His270 is the actual crosslinked amino acid. The fragments of a peptide derived from the U4/U6-specific protein 61 K encompassing positions 263–288 (SSTSVLPHTGYIYHSDIVQSPPDLR) up to y18 (residues 270–288). The observed losses of adenine and adenosine, on the other hand, suggest that the RNA is crosslinked via a uracil rather than an adenine base.

Acknowledgments

We thank Thomas Conrad, Gabi Heyne, Peter Kemkes, and Hossein Kohansal for their excellent technical assistance in preparation of snRNPs from HeLa cells; Uwe Plessmann and Monika Raabe for their help in HPLC; and Christof Lenz from Applied Biosystems Europe for his support in MALDI MS on the Voyager DE-STR. This work is supported by a YIP grant from the EURASNET (within the sixth EU framework) to H.U. and by a BMBF grant (031U215B) to R.L.

4. Notes

1. The buffer contains 20 m*M* HEPES KOH (pH 7.5), 250 m*M* NaCl, 1.5 m*M* $MgCl_2$, and 0.5 m*M* DTT in the case of U1 snRNPs, in 20 m*M* HEPES KOH (pH 7.9), 1.5 m*M* $MgCl_2$, 250 m*M* NaCl, 0.5 m*M* DTT, and 0.2 m*M* EDTA in the case of 25 S[U4/U6.U5] tri-snRNPs, or in 20 m*M* HEPES KOH (pH 7.5), 150 m*M* NaCl, 1.5 m*M* $MgCl_2$, 0.1% Triton X-100, 0.2 m*M* EDTA in the case of reconstituted [U4atac-15.5 K-61 K] protein–RNA complexes. The ideal sample volume is 1 mL.
2. If the starting sample volume is smaller, it is adjusted with the respective buffer to yield a solution depth of 1 mm. However, the sample concentration should not be below 0.05 mg/mL. If the starting sample volume is larger, glass dishes with a greater diameter are used, if available; alternatively, the sample can be irradiated in successive batches in the same dish. The latter procedure has the advantage that the sample recovery rate is higher than when glass dishes with a larger diameter are used. In particular, protein–RNA particles that contain numerous relatively large proteins and relatively small RNAs tend to stick to the glass surface; saturation of the glass surface (with sample) minimizes this effect. An example of such a complex is the 25 S [U4/U6.U5] tri-snRNP, which contains large proteins such as the U5-specific proteins 220 K (human Prp8) and 200 K, the U4/U6-specific proteins 90 K and 60 K, and snRNAs of length 116, 145, and 106 nt, respectively *(12)*.
3. For sample volumes of 1 mL or less, the sample is distributed into three (or fewer) 1.5-mL Eppendorf tubes and precipitated as above. For larger sample volumes, we recommend using Corex glass tubes (Kendro Laboratory Products) for the precipitation. Note that the Corex tubes must be completely free of RNases; our laboratory uses a set of Corex tubes reserved exclusively for this purpose.
4. The sample volume depends on the size of the SE column. We usually dissolve the sample in 50 µL; however, large-scale preparations require volumes of up to 200 µL or more. As a rule of thumb, 100 pmol to 1 nmol of RNA in a sample volume of 50 µL leads to excellent separation on a 75 HR column (3.1 × 300 mm, Amersham Biotech/GE Healthcare) in the Smart System (Amersham Bioscience/GE Healthcare). For larger volumes, we perform several runs on small SE columns instead of using a larger column on a different system. Larger columns generate larger volumes of eluted sample, and the detector systems of preparative

chromatography systems are generally not as sensitive as those in semimicro- or microsystems.

5. The RNA hydrolysis must be complete. Therefore, we recommend using a sufficient amount of RNases. However, when RNase A and T1 are used together, the amount of RNase A should be kept as low as possible as RNase A is readily cleaved by endoproteinases in a second digestion (*see* **Subheading 3.4.2., step 4**), and its fragments interfere with the detection of the crosslinks during HPLC. In contrast, RNase T1 remains almost intact and elutes as a single peak from the RP column around 30% solvent B (*see* **Subheading 3.4.1., Fig. 4**).
6. This second endoproteinase treatment is necessary as it substantially improves the yield of crosslinked peptide–RNA oligonucleotides in the RP-HPLC. We conclude that the first digestion (**Subheading 3.3.2., step 1**) generates mainly larger peptide fragments because of steric hindrance of the endoproteinase caused by the large intact RNA.

References

1. Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G., and Lührmann, R. (2000) 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. *J. Biol. Chem.* **275**, 41458–41468.
2. Urlaub, H., Raker, V. A., Kostka S., and Lührmann, R. (2001) Sm protein–Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.* **20**, 187–196.
3. Urlaub, H., Hartmuth, K., and Lührmann, R. (2002) A two-tracked approach to analyze RNA–protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods* **26**, 170–181.
4. Nottrott, S., Urlaub, H., and Lührmann, R. (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.* **21**, 5527–5538.
5. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438.
6. Handa, N., Nureki, O., Kurimoto, K., et al. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585.
7. Gautier, T., Berges, T., Tollervey, D., and Hurt, E. (1997) Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol. Cell. Biol.* **17**, 7088–7098.
8. Will, C. L., Kastner, B., and Lührmann, R. (1994) Analysis of ribonucleoprotein interactions in *RNA processing*. In: *A Practical Approach*, Vol. 1 (Higgins, S. J., and Hames, B. D., eds.), IRL Press, Oxford, U.K., pp. 141–177.
9. Kastner, B. (1998) Purification and electron microscopy of spliceosomal snRNPs. In: *RNP Particles, Splicing and Autoimmune Diseases* (Schenkel, J., ed.), Springer-Verlag, Berlin, pp. 95–140.

10. Kastner, B., and Lührmann, R. (1999) Purification of U small nuclear ribonucleoprotein particles. In: *RNA–Protein Interaction Protocols, Methods in Molecular Biology*, Vol. 118 (Haynes, S. R., ed.), Humana Press, Totowa, NJ, pp. 289–298.
11. Burge, C. B., Tuschl, T., and Sharp, P. A. (1999) Splicing of precursors to mRNA by the spliceosome. In: *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 525–560.
12. Will, C. L., and Lührmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301.
13. Blum, H., Beier, H., and Gross, H. J. (1987), Improved silver staining of plant proteins, RNA and DNA polyacrylamid gels. *Electrophoresis* **8**, 93–99.
14. Yisraeli, J. K., and Melton, D. A. (1989) Synthesis of long, capped transcripts in vitro by SP6 and T7 RNA polymerases. *Methods Enzymol.* **180**, 42–50.
15. Milligan, J. F., and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51–62.
16. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning, a Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
17. Raker, V. A., Hartmuth, K., Kastner, B., and Lührmann, R. (1999) Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell. Biol.* **19**, 6554–6565.
18. Kambach, C., Walke, S., Young, R., et al. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**, 375–387.
19. Meisenheimer, K. M., Meisenheimer, P. L., and Koch, T. H. (2000) Nucleoprotein photo-cross-linking using halopyrimidine-substituted RNAs. *Methods Enzymol.* **18**, 88–104.
20. Nottrott, S., Hartmuth, K., Fabrizio, P., et al. (1999) Functional interaction of a novel 15.5 kD [U4/U6.U5] tri-snRNP protein with the 5′ stem-loop of U4 snRNA. *EMBO J.* **18**, 6119–6133.
21. Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5 kD protein bound to a U4 snRNA fragment. *Mol. Cell* **6**, 1331–1342.
22. Urlaub, H., Kruft, V., Bischof, O., Müller E. C., and Wittmann-Liebold, B. (1995) Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *EMBO J.* 14, 4578–4588.
23. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
24. Urlaub, H., Thiede, B., Müller, E. C., and Wittmann-Liebold, B. (1997) Contact sites of peptide-oligoribonucleotide cross-links identified by a combination of peptide and nucleotide sequencing with MALDI MS. *J. Protein Chem.* **16**, 375–383.
25. Yates, J. R., 3rd. (2004) Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 297–316.

26. Mann, M., Hendrickson, R. C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**, 437–473.
27. Wang, R., and Chait, B. T. (1994) High-accuracy mass measurement as a tool for studying proteins. *Curr. Opin. Biotechnol.* **5**, 77–84.
28. McLachlin, D. T., and Chait, B. T. (2001) Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr. Opin. Chem. Biol.* **5**, 591–602.
29. Berkenkamp, S., Kirpekar, F., and Hillenkamp, F. (1998) Infrared MALDI mass spectrometry of large nucleic acids. *Science* **281**, 260–262.
30. Kirpekar, F., and Krogh, T. N. (2001) RNA fragmentation studied in a matrix-assisted laser desorption/ionisation tandem quadrupole/orthogonal time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **15**, 8–14.
31. McCloskey, J. A., Whitehill, A. B., Rozenski, J., Qiu, F., and Crain, P. F. (1999) New techniques for the rapid characterization of oligonucleotides by mass spectrometry. *Nucleosides Nucleotides* **18**, 1549–1553.
32. Fenn, J. B., Mann, M., Meng C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionisation for mass spectrometry of large biomolecules. *Science* **246**, 64–71.
33. Karas, M., and Hillenkamp, F. (1988), Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**, 2299–2301.
34. Koy, C., Mikkat, S., Raptakis, E., et al. (2003) Matrix-assisted laser desorption/ionization-quadrupole ion trap-time of flight mass spectrometry sequencing resolves structures of unidentified peptides obtained by in-gel tryptic digestion of haptoglobin derivatives from human plasma proteomes. *Proteomics* **3**, 851–858.
35. Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988) Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153.
36. Steen, H., and Jensen, O. N. (2002) Analysis of protein–nucleic acid interactions by photochemical cross-linking and mass spectrometry. *Mass Spectrom. Rev.* **21**, 163–182.
37. Kaufmann, R., Spengler, B., and Lützenkirchen, F. (1993) Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption ionization. *Rapid Commun. Mass Spectrom.* **7**, 902–910.