A two-tracked approach to analyze RNA–protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles

Henning Urlaub,* Klaus Hartmuth, and Reinhard Lührmann*

Max-Planck-Institute for Biophysical Chemistry, Department of Cellular Biochemistry, Am Fassberg 11, D-37077 Göttingen, Germany

Abstract

Much attention is currently being devoted to questions of protein and RNA tertiary structures and to the quaternary arrangement of the individual macromolecules in ribonucleoprotein (RNP) particles. In this article we describe two complementary strategies that allow the identification of RNA–protein contact sites in assembled, nonlabeled RNP particles after UV crosslinking. The first combines immunoprecipitation of UV-irradiated RNP particles under mildly denaturing conditions followed by primer-extension analysis of the crosslinked (and thus coprecipitated) RNA. The second involves the purification of crosslinked peptide–oligonucleotide from RNP particles and the subsequent analysis of the crosslinked peptide and RNA by Edman degradation and matrix-assisted laser desorption/ionization (MALDI)–mass spectrometry (MS), respectively. Although the first approach provides a rapid method for the exact identification of RNA–protein contact sites in purified nonlabeled RNP particles, the latter adds valuable information about potential RNA binding domains within proteins and, thus, about the arrangement of these proteins within the quaternary structures of complex RNP assemblies. Recently, we applied both these strategies successfully to native purified spliceosomal RNP. These methods may be generally applicable to the analysis of RNP complexes, especially as they avoid labeling and reconstitution, both of which risk introducing artifacts. © 2002 Elsevier Science (USA). All rights reserved.

1. Introduction

In the cell, RNA is rarely present in the absence of proteins, and is often found as part of a stable RNP complex. RNA–protein interactions within ribonucleoproteins (RNPs) are crucial for the function of these particles in gene expression, gene regulation, and cell function. An understanding of the molecular mechanisms of these processes requires the identification and characterization of the RNA–protein interactions involved. Owing to the diversity of RNA structures, RNA–protein interactions are more complex than DNA–protein interactions, involving more primary and structural interaction motifs in the RNA and a greater diversity of complementary structures in the protein [1–4]. In addition, many proteins in complex ribonucleoprotein assemblies (e.g., ribosomes or spliceosomes) do not contain characterized RNA binding motifs. There-

fore, the introduction of general techniques useful for detection of novel RNA–protein interaction sites in RNPs represents an important step forward, since it is a precondition for obtaining general insight into the molecular organization and function of such particles. For example, before the atomic structures of ribosomes were known [5–8], data from various biochemical studies [9–11] had already led to models of the ribosome. These data included RNA–protein interactions inferred from biochemical results. Moreover, in the case of prokaryotic ribosomes and their subunits, such data proved useful in the initial interpretation of data from cryo electron microscopy [12,13] and crystallization [14,15] that ultimately led to the construction of three-dimensional molecular models for these particles.

The combination of data derived from crosslinking studies, from binding studies with purified components, and from cocrystallization studies with images of the spliceosomal HeLa U1 small nuclear (sn)RNP obtained by electron cryomicroscopy at 10 Å resolution has allowed the deduction of the overall structure of the U1 snRNP [16]. The structure model obtained reveals the

---

* Corresponding author. Fax: +49-(0)551-201-1197.
  E-mail addresses: henning.urlaub@mpi-bpc.mpg.de (H. Urlaub), reinhard.luehrmann@mpi-bpc.mpg.de (R. Lührmann).

ring-shaped Sm core, which consists of the seven Sm proteins (G, E, F, D1, D2, D3, B) assembled on the Sm site of U1 snRNA, and it also shows the location of the U1-snRNP-specific proteins A and 70K, bound respectively to stems II and I of the U1 snRNA. The availability of structures of spliceosomal components at atomic resolution is increasing steadily [17–20], yet structural information on the spliceosome is still limited. In the absence of further cocrystals and of binding studies with purified components, crosslinking is a very powerful tool in elucidating unknown or uncharacterized RNA–protein interactions. A variety of crosslinking techniques are available, and these have been used successfully for the identification of sites on the RNA to which proteins crosslink within spliceosomal complexes and other RNP particles. Crosslinking between RNA and proteins has been obtained by UV irradiation at 254 nm, owing to the natural photoreactivity of the RNA bases [21], or at higher wavelengths when photoreactive base analogs such as 5-bromodeoxyuridine and 4-thio-uracil were incorporated at specific sites [22, and references therein]. Alternatively, chemical modification of the RNA through the ribose backbone [23] has also proven useful in RNA–protein crosslinking by UV light. All of these established approaches suffer from one weakness: for the crosslinked sites to be mapped, the RNA must be labeled with ^{32}P, and hence the ribonucleotide particles have to be reconstituted. Although reconstitution of splicesomal particles has been used successfully to increase the crosslinking yield by the introduction of site-specific labels [24–28], it is subject to difficulties associated with the efficiency of reconstitution and, more seriously, to the risk of artifacts: complex ribonucleoprotein particles are difficult to assemble in vitro, and incomplete or incorrect assembly can result in heterogeneous populations or false-positive results.

We have therefore focused on developing an approach that enables us to map RNA–protein crosslinking sites in native, nonlabeled snRNP particles. This approach was tested on purified U1 snRNP particles and has since been extended to other snRNPs [29,30]. Our approach can be regarded as a potentially general one, suitable for the detection of single and/or multiple RNA–protein contact sites in native RNP particles. The overall strategy is outlined schematically in Fig. 1, for the example of U1 snRNPs. It comprises (i) the identification of the crosslinking sites on the RNA and the identification of the corresponding crosslinked protein by immunoprecipitation combined with primer-extension analysis, and (ii) the identification of the crosslinking sites or regions within a protein and identification of the crosslinked RNA by mass spectrometry. The method depends on the numbers of antibodies available for the different proteins of the particle and how efficiently each antibody precipitates its corresponding protein, especially under mildly denaturing

conditions (see below). In nonlabeled U1 snRNP we have so far identified three different crosslinking sites on the RNA (Fig. 1). It is clear that the latter analysis requires much more material, but, if successful, it enables one to define putative RNA binding domains within the crosslinked proteins, as outlined for the U1 70K protein and for the Sm G protein in the heptameric Sm ring (Fig. 1). In this article, we describe each approach in detail, emphasizing the most critical steps. Most examples cited refer to purified U1 snRNP particles; other systems are referred to where appropriate.

2. UV crosslinking of native snRNP particles

Total snRNPs are typically purified from HeLa nuclear extract by immunoaffinity chromatography by using an H2O anti-m₃G affinity column [31,32]. Fractions enriched in 25S [U4/U6.U5] tri-snRNPs, 20S U5 snRNPs, and 17S U2 and 12S U1 snRNPs are obtained by subsequent glycerol-gradient centrifugation of the total snRNPs. Glycerol-gradient-purified 17S U2 and 12S U1 snRNPs are applied to Mono Q ion-exchange chromatography columns, and the U1 or U2 snRNP fractions eluting from the Mono Q column are crosslinked without further treatment. It is important that in all further steps potassium chloride in buffers should be avoided (see below), and that the concentration of buffer ingredients that could decrease the crosslinking yield (e.g., glycerol) should be kept as low as possible.

Typically, purified native snRNP particles are adjusted to a concentration of not more than 0.1 mg/ml with a buffer containing 20 mM Tris–HCl (pH 7.0), 370 mM NaCl, 1.5 mM MgCl₂, 0.5 mM dithiothreitol (DTT), and 0.2 mM phenylmethylsulfonyl fluoride (PMSF), according to the salt concentration at which U1 snRNP particles elute from the Mono Q column. For identification of crosslinking sites on the snRNA by the primer-extension method, the samples are divided into droplets of 25 μl on precooled 10-well multitest slides (ICN Biomedicals, Inc., Aurora, OH). Alternatively, droplets can be placed on a Parafilm-wrapped aluminum block on ice or in 0.5-ml Eppendorf tubes on ice. In the last case, we use a maximum sample volume of 10 μl in each tube. Siliconized glass plates or plastic plates cannot be used, as purified snRNPs stick to these surfaces. The glass slides are then placed on an aluminum block on ice (to minimize heating during irradiation) and are UV-irradiated at a distance of 2 cm from the UV source. Eppendorf tubes are mounted directly under the lamps. For the 254-nm UV source, four 8-W germicidal lamps (G8T5, Herolab, Wiesloch, Germany) are used in parallel, mounted in a specially constructed holder. We have observed that UV irradiation for longer than 2–3 min does not increase the crosslinking yield significantly, and irradiation for longer than 3 min leads
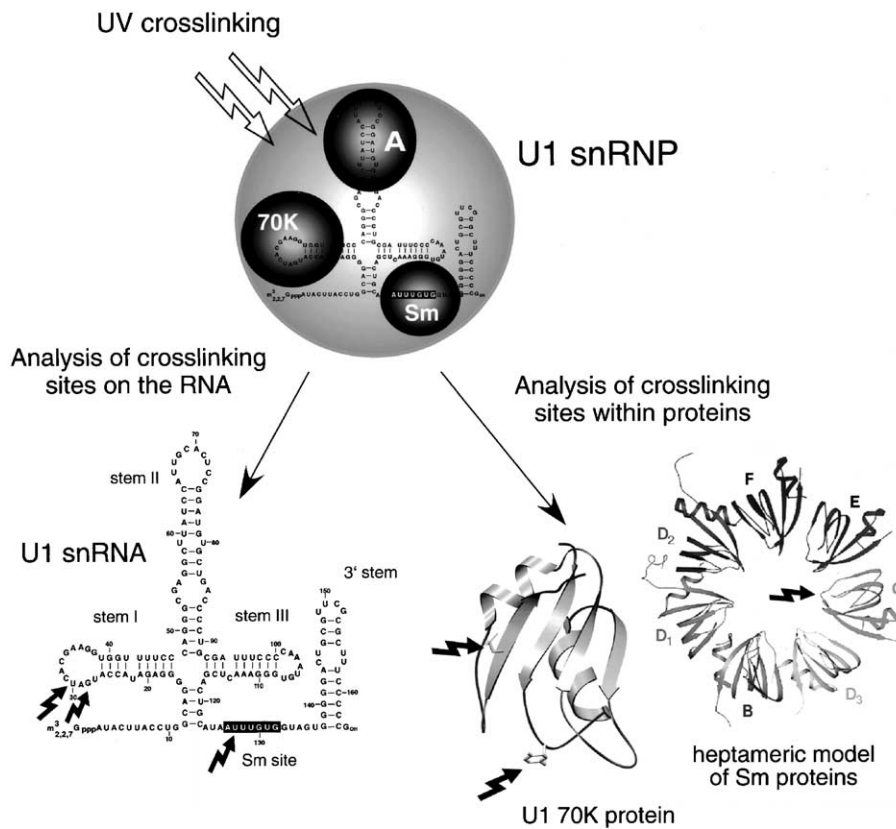
Fig. 1. Complementary strategy for identification of protein–RNA crosslinking sites in native RNP particles as outlined for UV-irradiated U1 snRNPs. UV-irradiated U1 snRNP were analyzed with the aim to map crosslinking sites on the U1 snRNA and the proteins. Arrows in the U1 snRNA secondary structure indicate the crosslinking sites on the RNA as identified by our immunoprecipitation/primer-extension method [29,30]; these were in G28 and U30 in stem I and U127 in the Sm binding site of U1 snRNA. Arrows at the 3D protein models of the U1 70K protein [29] and of the heptameric Sm protein ring [45] show the corresponding crosslinking sites within the protein as identified by Edman degradation combined with matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) [29,30].

to a substantial loss of particles. Alternatively, a UV Stratalinker 2400 (Stratagene, La Jolla, CA) can be used, but, because of the differences in lamp geometry, the crosslinking conditions have to be adjusted. For large-scale crosslinking experiments we use glass dishes with a planar surface. The depth of the sample solution is around 1 mm; the diameter of the glass dish should be chosen to match the sample volume correspondingly.

3. Analysis of the crosslinked products

3.1. Primer-extension analysis and immunoprecipitation

This approach includes three sets of experiments, outlined below for the example of UV-irradiated U1 snRNP particles.

(i) *Primer-extension analysis of the RNA derived from UV-irradiated RNP particles*: Potential RNA–protein crosslinking sites are detected as discrete reverse transcriptase stops after digestion of the crosslinked snRNP particle with proteinase K, since a few amino acid residues remain covalently attached to the RNA at the sites

of crosslinking. The actual crosslinking site is taken to be one nucleotide upstream of the stop site.

(ii) *Primer-extension analysis of UV-irradiated naked RNA, isolated from native particles*: This control experiment allows one to distinguish between, on the one hand, reverse transcriptase stops that are due to UV-induced strand breaks in the RNA, intra-RNA crosslinks, or stuttering of the reverse transcriptase (Fig. 2B, triangles) and, on the other hand, reverse transcriptase stops that are caused by a crosslinked protein.

(iii) *Immunoprecipitation of crosslinked proteins under mildly denaturing conditions and subsequent primer-extension analysis of the coprecipitated crosslinked RNA*: This experiment (Figs. 2A–C) allows the identification of a single RNA–protein crosslink. Immunoprecipitation with antibodies against the specific proteins of U1 snRNP (e.g., anti-SmG antibody), under conditions where the particle is completely dissociated, leads to the isolation of a single crosslinked protein together with a crosslinked RNA moiety. As in the other experiments, the crosslinked site is detected by primer extension after hydrolysis of the protein moiety with proteinase K.
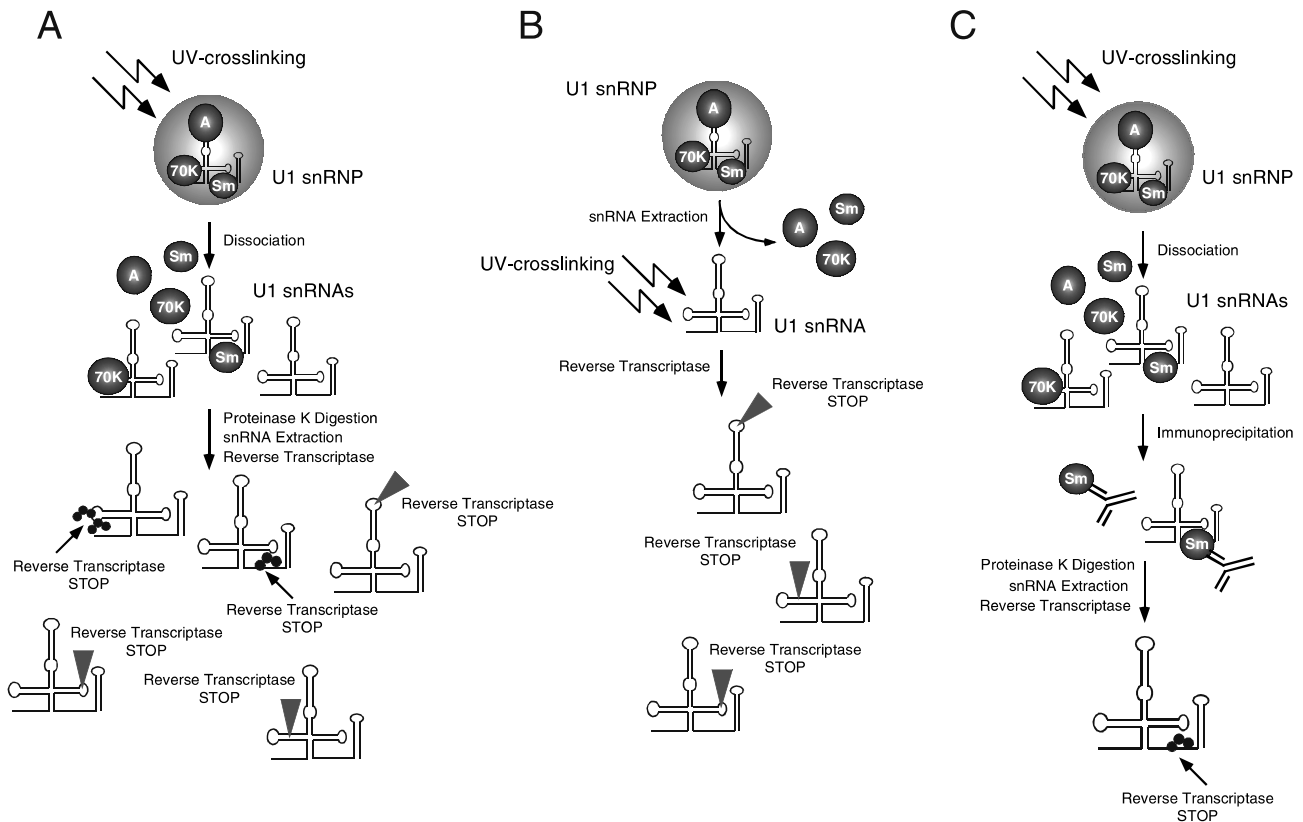
Fig. 2. Schematic representation of the procedure for identification of RNA–protein crosslinking sites in native U1 snRNPs by primer-extension analysis. (A) Primer-extension analysis of U1 snRNA derived from UV-irradiated U1 snRNP particles. (B) Primer-extension analysis of naked UV-irradiated U1 snRNA. (C) Immunoprecipitation of U1 snRNP-specific proteins, with, e.g., an antibody against Sm proteins combined with primer-extension analysis of the crosslinked and therefore coprecipitated U1 snRNA moiety. Black circles at the RNA indicate reverse transcriptase stop sites due to a protein–RNA crosslinking event. Triangles mark stops sites that are caused by UV-induced strand breaks of the RNA, intra-RNA crosslinks, or stuttering of the reverse transcriptase due to a particular RNA conformation.

By comparing the reverse transcriptase patterns found in the three experiments, it is usually possible to identify unambiguously both the crosslinked protein and the exact RNA nucleotide to which it is attached.

The experimental details for the successive steps are as follows:

3.1.1. Preparation of crosslinked RNA

3.1.1.1. Crosslinked U1 snRNA derived from UV-irradiated U1 snRNPs. A 50-µl sample of UV-irradiated native RNP particles is adjusted to 100 µl with sample buffer (sample buffer refers to the buffer in which native particles were finally purified, i.e., Mono Q ion-exchange buffer with salt concentration according to the elution of U1 snRNPs, see above), and the sample is incubated with proteinase K (Roche Diagnostics GmbH, Mannheim, Germany) at a final concentration of 1 mg/ml for 30 min at 37 °C in the presence of 1% SDS. Note that SDS immediately precipitates in the presence of potassium chloride, so sodium chloride instead of potassium chloride is recommended for buffers (see above). The RNA is extracted with phenol/chloroform, precipitated with 3 vol ethanol in the presence of

20 µg glycogen (Roche Diagnostics GmbH, Mannheim, Germany), dissolved in 100 µl 0.3M sodium acetate, again precipitated with 3 vol ethanol, dried under vacuum, and finally dissolved in 6.5 µl CE buffer (20 mM cacodylate–KOH, pH 7.0, 0.2 mM EDTA) and stored at −20 °C. See Fig. 2A.

3.1.1.2. UV-crosslinked naked U1 snRNA isolated from nonirradiated U1 snRNPs. A 50-µl sample of noncrosslinked native U1 snRNP particles is adjusted with Mono Q buffer to 100 µl and incubated with proteinase K as above. The nonirradiated naked U1 snRNA is recovered as above (without glycogen) and finally dissolved in 50 µl buffer. Two 25 µl sample droplets are UV-irradiated as above, precipitated with ethanol in the presence of 20 µg glycogen, redissolved in 0.3M sodium acetate, precipitated with ethanol, and finally stored in 6.5 µl CE buffer at −20 °C. See Fig. 2B.

3.1.1.3. Crosslinked U1 snRNA after immunoprecipitation with antibodies specific for U1 snRNP proteins. The most critical experiment during this analysis is the immunoprecipitation of crosslinked proteins under mildly

denaturing conditions. One characteristic of snRNP particles is their strong protein–protein interactions, for example, within the Sm core [33] or within an RNA-free heteromeric U5 snRNP protein complex consisting of the 220 K, 200 K, 116 K, and 40 K proteins [34]. To precipitate specifically only a single protein that is crosslinked to the RNA, strong protein–protein interactions within the particles have to be disrupted completely by the use of SDS. Importantly, such denaturing conditions must still allow the antibodies to retain their reactivity during immunoprecipitation. For most of our experiments, we use a final SDS concentration of 1% (w/v). However, when investigating the Sm protein–snRNA crosslinking sites in UV-irradiated U1 snRNPs, we increase the SDS concentration to 2% (w/v) (see below and Fig. 3, compare 3B and 3C). Fifty microliters of UV-

crosslinked sample is adjusted with SDS to a final SDS concentration of 1 or 2% (w/v), respectively. The samples are incubated for 10 min on a shaker and then heated for a minimum of 10 min at 70 °C to allow complete dissociation of the RNP complex. After cooling to room temperature, Triton X-100 is added to a final concentration of 5% (v/v), and the sample volume is then adjusted to 350 µl with phosphate-buffered saline (20 mM Na_2HPO_4, pH 8.0, 130 mM NaCl). The samples are incubated with different U1 snRNP-protein-specific antibodies bound to protein A–Sepharose (Amersham-PharmaciaBiotech, Uppsala, Sweden) for 1 h at 4 °C with end-over-end rotation. The samples are washed four times with phosphate-buffered saline (PBS) within the same Eppendorf tube. For the final washing step, the slurry is transferred into a new Eppendorf tube, washed
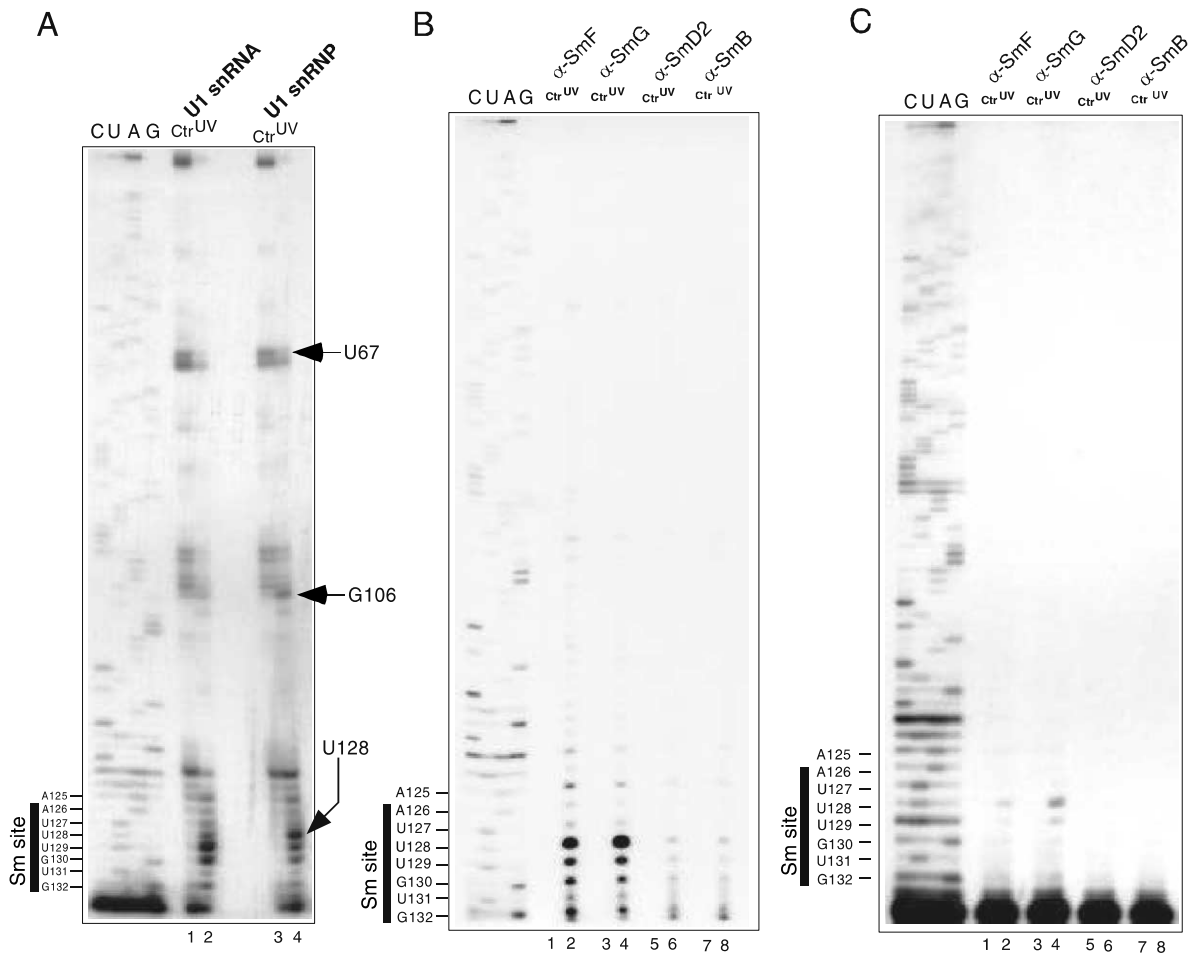

Fig. 3. Immunoprecipitation of RNA–protein crosslinks under different denaturing conditions combined with primer-extension analysis of the coprecipitated RNA. (A) Primer-extension analysis of UV-irradiated naked U1 snRNA (lanes 1 and 2) compared with U1 snRNA derived from UV-irradiated U1 snRNPs (lanes 3 and 4). Lanes 1 and 3 are controls with no UV irradiation. C, U, A, and G are dideoxy sequence markers. Nucleotides within the Sm site are listed on the left. Reverse transcriptase stops that are due to putative RNA–protein crosslinks are denoted on the right. The primer used was complementary to positions 143–152 of U1 snRNA. (B) Primer-extension analysis of U1 snRNA derived from UV-irradiated U1 snRNPs after immunoprecipitation with various antibodies against Sm proteins (SmF, SmG, SmD2, SmB) in the presence of 1% SDS (see text for details). (C) primer-extension analysis of U1 snRNA derived from UV-irradiated U1 snRNPs with the same set of antibodies as in (B) but in the presence of 2% SDS. The Sm site in (B) and (C) is identified as in (A).

one more time, and finally resuspended in sample buffer containing 1% SDS (w/v). The washing procedure is critical for a successful primer-extension analysis, for two reasons. First, the additional washing step in another Eppendorf tube reduces the nonspecific background level caused by RNA that might adhere to the tube wall. Second, washing the samples with buffers containing nonidet P-40 (NP-40) leads to a dramatic increase of nonspecific background signals in the subsequent primer extension (we have no explanation for this). The RNA is extracted and recovered in the presence of glycogen as described above. Finally, the RNA is stored in 3.5 µl CE buffer at −20 °C. See Fig. 2C.

3.1.2. Primer-extension analysis

For a typical reverse transcriptase reaction, 1 µl of crosslinked RNA is incubated with 0.25 µl 10∞ HY buffer (0.5 M Tris–HCl, pH 8.4, 0.6 M NaCl, 0.1 M DTT), 0.5 µl ^{32}P-labeled cDNA oligonucleotide, and 0.75 µl H_2O for 60 s at 96 °C. After cooling to room temperature 2.15 µl H_2O, 0.25 µl 10∞ RTbuffer (0.5 M Tris–HCl, pH 8.4, 0.6 M NaCl, 0.1 M $MgCl_2$, 0.1 M DTT), 0.1 µl dNTPs (5 mM in dGTP, dATP, dTTP, and dCTP), and 0.1 µl reverse transcriptase (Seikagaku, Japan) are added to the reaction and incubated 45 min at 43 °C. The transcripts are separated on a 50% (w/v) urea and 9.6% polyacrylamide (20:1) sequencing gel in 1× Tris/borate/EDTA (pH 8.3). Gels are run at 65 W throughout. To detect crosslinking sites adjacent to the sites of the DNA primer, prerunning of the gel is recommended.

The choice of the primer is critical for the detection of crosslinking sites. For investigation of crosslinking sites within the U1 snRNP particle we use two different primers. One is complementary to nucleotide positions 63–77 in stem II of U1 snRNA, the other to nucleotide positions 143–152 in the 3′ stem–loop of U1 snRNA. The first primer proves to be useful in detection of two crosslinking sites of the U1 70K protein in stem I of U1 snRNA [29], while the latter is used to detect the crosslinking sites on the Sm site of U1 snRNA (Fig. 3) [30]. The latter primer could not be used for identification of the U1 70K crosslinking sites, since these sites are located too far downstream from the primer. In general, RNA has to be probed with different primers to obtain a complete map of crosslinking sites.

Figs. 3A–C show all three sets of experiments performed for identification of the exact crosslinking sites of Sm proteins on the U1 Sm site, thereby summarizing the most critical points of the primer–extension as mentioned above. First, by using a primer complementary to the 3′ stem of U1 snRNA (positions 143–152), reverse transcriptase stops that are different from those in the naked, UV-irradiated U1 snRNA can be detected only in the 3′ half of U1 snRNA (Fig. 3A, U128, G106, U67). Additional reverse transcriptase stops, e.g., at C31

and A29 in stem–loop I, caused by U1 70K protein crosslinking (data not shown; see, however, Ref. [29]) can be detected only by the use of a different primer (complementary to positions 63–77 of U1 snRNA). Second, immunoprecipitation performed under conditions where strong protein–protein interactions are not completely disrupted leads to "false positives," as shown for the SmF protein in Fig. 3B. Immunoprecipitation with anti-SmF, anti-SmG, anti-SmD2, or anti-SmB/B′ antibodies after dissociation of crosslinked U1 snRNPs in the presence of only 1% SDS (w/v) shows the same reverse transcriptase pattern for the SmF and SmG proteins. When crosslinked U1 snRNPs are dissociated in the presence of 2% SDS, a strong reverse transcriptase stop at U128 is observed exclusively after immunoprecipitation with the anti-SmG protein antibody (Fig. 3C), thus demonstrating that the SmG protein, but not the SmF protein, is crosslinked to U129 in the Sm site of U1 snRNA [30].

3.2. Identification of crosslinks sites at the molecular level by protein mapping methods

In addition to the primer-extension method described above, crosslinking of native or reconstituted particles can be performed on a preparative scale to identify peptide regions or domains that are in direct contact with the RNA. The experimental approach can be divided into four different steps: (1) Purification of crosslinked peptide–RNA complexes by size-exclusion chromatography, (2) purification of crosslinked peptide–oligonucleotides by reverse-phase (RP) high-performance liquid chromatography (HPLC), (3) Edman degradation of crosslinked peptide–oligonucleotide complexes, and (4) matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS).

Edman degradation identifies the crosslinked protein region in peptide–oligonucleotide complexes isolated from UV-irradiated RNP particles [35–37]. This information allows a first insight into potential new RNA binding domains of the proteins in contact with RNA, and can be extremely helpful in defining RNA binding domains in the three-dimensional structures of proteins that have not yet been cocrystallized with their RNA substrate. In addition, MALDI-MS analysis of the peptide–oligonucleotide complexes reveals the crosslinked oligonucleotide part [29,38,39]. It therefore complements the data derived from the immunoprecipitation or primer-extension method, and, if a protein has multiple crosslink sites on the RNA (as in the case of the U1 70K crosslinked to stem I of U1 snRNA), then it relates the crosslinking sites within the protein to those on the RNA. Such an analysis thus offers a powerful tool for the refinement of three-dimensional structures that have not yet reached atomic resolution.

Because the successful analysis of the crosslinked protein regions and the oligonucleotide strongly depends on the homogeneity and purity of the peptide–oligonucleotide complex, we focus in this report on the two crucial chromatography steps necessary for the isolation and purification of peptide–oligonucleotide crosslinks from native snRNP particles, namely, size-exclusion chromatography and reverse-phase high-performance liquid chromatography.

3.2.1. Purification of crosslinked peptide–RNA complexes by size-exclusion chromatography

In the first purification step, peptides that are crosslinked to RNA have to be separated from the excess of
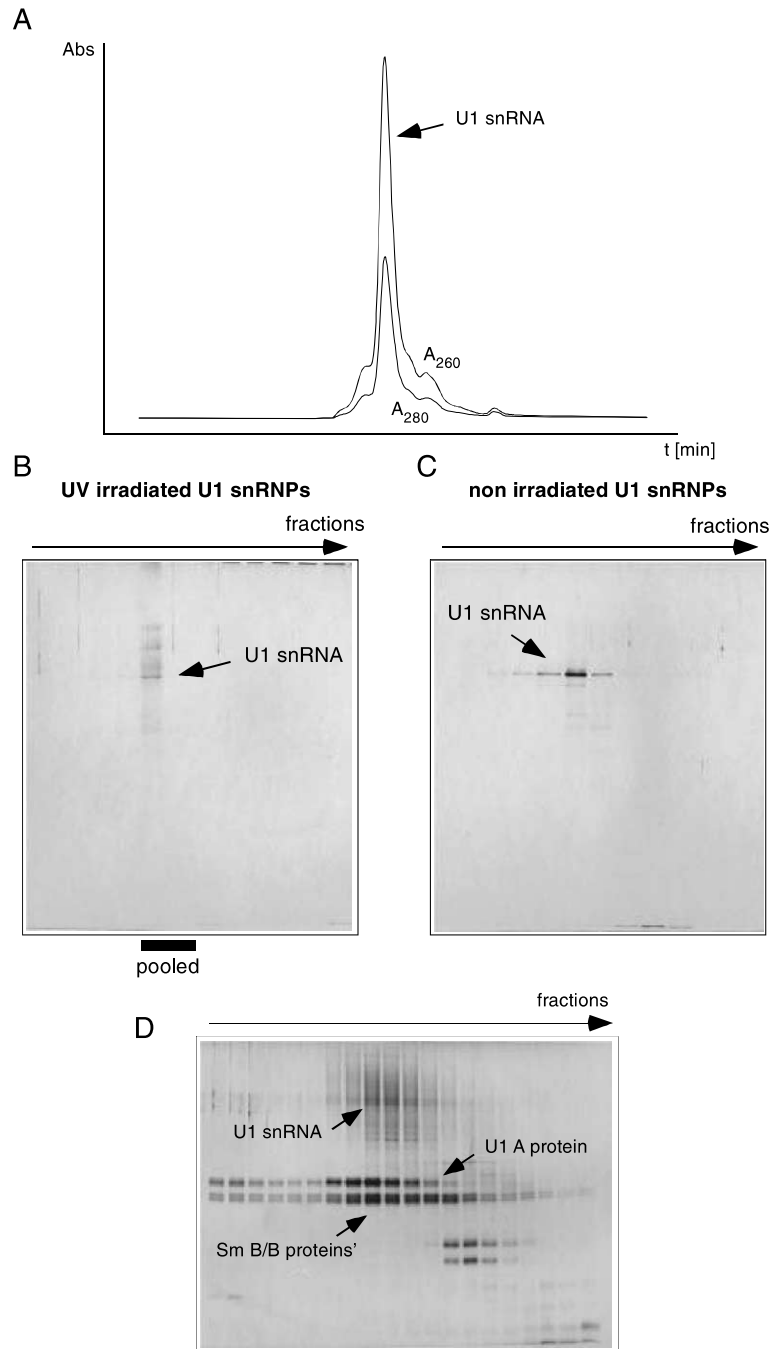

Fig. 4. Size exclusion chromatography for purification of peptide–snRNA crosslinks. (A) Superdex 75 column elution profile of U1 snRNA derived from U1 snRNPs after digestion with trypsin under denaturing conditions. (B) Silver-stained SDS–PAGE of eluted fractions from UV-irradiated U1 snRNP particles digested with trypsin. (C) Silver-stained SDS–PAGE of eluted fractions from a nonirradiated sample digested with trypsin. (D) Silver-stained SDS–PAGE from UV-irradiated U1 snRNPs digested with endoproteinase Glu-C.

noncrosslinked peptides. This is achieved by size-exclusion chromatography. Here, noncrosslinked U1 snRNA coelutes with U1 snRNA that contains crosslinked peptides, while noncrosslinked peptides are much smaller and are therefore eluted later from the column. If this step is not performed successfully, then further purification of peptide–RNA crosslinks and subsequent identification of the actual crosslinking sites is no longer possible. Copurified noncrosslinked peptides will also elute from the RP-HPLC in the final purification step (see below) and interfere with the detection of peaks containing peptide–oligonucleotide crosslinks.

Large-scale crosslinked U1 snRNP particles (see above) are precipitated with 3 vol ethanol overnight in 30-ml Corex tubes (Kendro Laboratory Products, Newtown, CT), washed with 80% ethanol, and air-dried for 10 min. The pellet is redissolved in a small volume of buffer (50 mM Tris–HCl, pH 8.0, 5 mM EDTA, 8 M urea or 1% SDS) to disrupt any protein–protein interaction, transferred to an Eppendorf tube, and heated for 10 min at 70 °C. After cooling to room temperature and diluting the sample to an appropriate concentration of urea or SDS at which endoproteinases are still active (<1 M urea or <0.1% SDS), the particles are incubated with, e.g., trypsin (sequencing grade, Roche Diagnostic GmbH) or chymotrypsin (sequencing grade, Roche Diagnostic GmbH) at a protease:substrate ratio of at least 1:20 (w/w) overnight (16 h). We also tested other endoproteinases for their ability to digest UV-irradiated particles under these conditions, to obtain different crosslinked fragments. Earlier crosslinking experiments with ribosomal subunits showed that the use of different endoproteinases led to the identification of additional crosslinking sites within the same protein [38]. Importantly, as is discussed below, the use of endoproteinases other than trypsin or chymotrypsin on snRNP particles causes serious problems during the further purification of the crosslinks.

To minimize the loading volume for the size-exclusion chromatography, the samples are precipitated with ethanol, washed, and dissolved in size-exclusion running buffer (50 mM Tris–HCl, pH 7.8, 150 mM NaCl, 5 mM EDTA) additionally containing 0.1% SDS or 1 M urea. For size-exclusion chromatography we used a Superdex 75 HR column (AmershamPharmaciaBiotech, column dimensions: 3.2×300 mm) in a SMART chromatography system (AmershamPharmaciaBiotech), at room temperature. The practical loading range is 50 µl, and the flow rate was 40 µl/min. Fig. 4A shows a typical example of U1 snRNA derived from UV-crosslinked U1 snRNPs eluting from the column after digestion of the U1 snRNP-specific proteins with trypsin in the presence of SDS. Note that the elution profile of a U1 snRNA sample derived from nonirradiated samples is identical. Aliquots of each fraction are loaded directly onto 13% SDS–PAGE and visualized by silver staining. Fig. 4B

shows the corresponding SDS–PAGE of U1 snRNA derived from crosslinked U1 snRNPs, and Fig. 4C shows a control of U1 snRNA derived from nonirradiated U1 snRNPs. On both gels only U1 snRNA is detectable, demonstrating that the protein moiety is completely digested. The fraction with U1 snRNA derived from irradiated particles (Fig. 4B) also contains higher migrating bands, revealing putative peptide–U1 snRNA crosslinks. Fig. 4D shows the silver-stained SDS–PAGE of eluted fractions from a similar experiment in which endoproteinase Glu-C was used to generate peptide–U1 snRNA crosslinks. It is obvious that under these conditions several U1 snRNP-specific proteins are not digested and they comigrate within fractions containing crosslinked and noncrosslinked U1 snRNA (Fig. 4D, e.g., the U1-specific protein A and the Sm protein B/B'). The same is observed when the endoproteinase Lys-C (Roche Diagnostic GmbH) or Arg-C (sequencing grade, Roche Diagnostic GmbH) is used. However, the less stringent conditions required for activity of these particular endoproteinases fails to induce complete dissociation of the particles.

3.2.2. Purification of crosslinked peptide–oligonucleotides by reverse-phase high-performance liquid chromatography

The eluted fractions containing U1 snRNA and U1 snRNA with crosslinked peptides are precipitated with ethanol overnight and dissolved in 150–200 µl of 2 mM EDTA. The U1 snRNA is digested with ribonuclease T1 and/or ribonuclease A for 2 h at 52 °C. Note that a large excess of RNase T1 (up to 20–30 µg) is used, whereas in the case of RNase A not more than 2–5 µg is used (see below). After treatment with RNase, the samples are adjusted to 50 mM Tris–HCl, pH 7.8, and subjected to a second digestion with 1 µg trypsin or chymotrypsin at 37 °C overnight. This step increases the yield of crosslinked peptide–oligonucleotides eluting from RP-HPLC [38]. We assume that after the first round of digestion before size-exclusion chromatography crosslinked peptides have several missed cleavage sites, because of protection of the crosslinked full-length snRNA. Such large peptide–oligonucleotide complexes are more hydrophobic and therefore might stick to the C_{18}-RP column. A second digestion, performed after hydrolysis of the RNA, generates smaller peptides that are rapidly eluted from the C_{18}-RP-HPLC column. Note that RNases within the sample are also digested: while RNase T1 is surprisingly stable against endoproteolytic treatment, RNase A is digested and its peptides might be eluted in HPLC and influence the detection of the peptide–oligonucleotide crosslinks. Thus, the amount of RNase A used must be kept as small as possible (see above).

The digested sample (200 µl) is injected directly onto a C_{18}-RP-HPLC column (281TP5215, Vydac, Hesperia, CA; column dimensions 2.1 mm i.d. × 150 mm, particle

size 5 μm, pore size 300 Å) mounted in a SMART chromatography system (AmershamPharmaciaBiotech) equipped with 50-ml pumps. Elution begins at 5% Solvent B (Fig. 5A). Solvent A is water containing 0.1% (v/v) trifluoroacetic acid (TFA); solvent B is acetonitrile containing 0.085% (v/v) TFA. In principle, in RP-HPLC running at 5% solvent B, RNA oligonucleotides are present in the flow-through, while peptides and crosslinked peptide–oligonucleotides are eluted at higher concentrations of solvent B. Therefore, the gradient is started after elution of the injection peak. A typical gradient applied was: (1) isocratic elution of the injection peak at 5% solvent B for 30 min, (2) 5% solvent B to 45% solvent B for 120 min, (3) 45% solvent B to 80% solvent B for 20 min. An additional isocratic step that varies in time is performed at 10% solvent B, as larger snRNA oligonucleotides always eluted at that percentage of solvent B (Fig. 5A). The gradient is continued when the baseline is reached. Fractions eluting from the column that exhibit strong absorbance at 220 and 260 nm are presumed to contain a crosslinked peptide–oligonucleotide (Fig. 5A, marked by asterisks). These fractions are dried under vacuum and stored at −20 °C until analysis by Edman degradation (Fig. 5B) and MALDI-MS (Fig. 5C). Note that prolonged storage times drastically reduce the amount of detectable crosslinks in Edman degradation and MALDI-MS. During RP-HPLC analysis of crosslinked samples, multiple crosslinked peptide–oligonucleotides are always eluted from the column (Fig. 5A, asterisks). The majority of the peaks contain peptides with the same site crosslinked to RNA. The fractions differ in peptide length (incomplete proteolytic digestion) and in oligonucleotide length (incomplete RNA hydrolysis). In addition, multiple peaks are derived from one crosslinked peptide–oligonucleotide and are due to different states of TFA denaturation. Note that strong absorbance at 280 nm is always indicative of contamination by larger peptides or small proteins (e.g., RNase T1, Fig. 5A) and fractions showing this should not be analyzed.

3.2.3. Edman degradation

Crosslinked peptide–oligonucleotides are dissolved in an appropriate volume of 50% acetonitrile containing 0.1% TFA, and two-thirds of the sample is subjected to automated N-terminal sequencing (this was performed on a Procise Sequencer, Applied Biosystems, Foster City, CA). During Edman degradation the crosslinked amino acid residue cannot be detected, owing to its covalent modification, and it is therefore identified as a gap in the Edman degradation products. This is outlined schematically in Fig. 5B for a tryptic fragment of the U1 70K protein crosslinked to U1 snRNA. N-Terminal sequencing identifies a peptide with the sequence RVXVDVER. In each cycle except the third, amino acids can be clearly identified. The sequence can be as-

signed to a tryptic fragment derived from U1 70K, $_{173}$RVLVDVER$_{180}$. Within this oligopeptide, Leu-175 is in the position of the nondetectable amino acid and is therefore considered to be the actual crosslinking site within the peptide.

3.2.4. Matrix-assisted laser desorption/ionization mass spectrometry

The remaining one-third of the sample is analyzed by mass spectrometry in a time-of-flight (TOF) MALDI mass spectrometer. As outlined schematically in Fig. 5C, the difference between the mass of the crosslinked oligonucleotide and that of the peptide (as determined by Edman sequencing of the crosslinked peptide moiety) allows determination of the nucleotide composition of the crosslinked oligonucleotide. The exact calculation procedure for the crosslinked oligonucleotide moiety of isolated complexes is described in detail elsewhere [29,38,39]. The RNA modification database, http://medlib.med.utah.edu/RNAmods/, provides several software tools useful for calculations concerning RNA in mass spectrometry.

Samples can be measured by two principal methods:

1. The sample (0.5–0.8 μl) in 50% acetonitrile, 0.1% TFA is mixed with 1.2 μl of matrix solution (saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma–Aldrich Co., St. Louis, MO)) in water:acetonitrile 3:2, containing 0.1% (v/v) TFA on the target and air-dried. Mass spectra are recorded on a VG-TOFSpec (Fisons, Manchester, UK) in the linear positive ion mode with an acceleration voltage of 22 kV by summing over 30–50 laser shots (N$_2$-pulsed laser, 337 nm, 4 ns) according to Thiede et al. [40].
2. Alternatively, 0.5 μl of the sample is mixed on a stainless-steel sample plate with the same volume of a 10 mg/ml solution of DHB (2,5-dihydroxybenzoic acid, Sigma–Aldrich Co.) matrix in 50% acetonitrile containing 0.1% TFA. The preparation is air-dried and used for MALDI-TOF analysis on a Voyager DE-STR (Applied Biosystems) without further processing. The samples are measured in the reflector mode, with acceleration voltage 20 kV, grid voltage 68%, delay time 250 ns, by summing over 300 laser shots (N$_2$-pulsed laser, 20 Hz, 337 nm).

Measuring in the reflector mode has the advantage over the linear mode that monoisotopic masses of the components can be considered. MALDI-MS in the linear mode does not allow the unambiguous differentiation of C and U, since these differ in mass by only one mass unit. Thus, different compositions of the crosslinked oligonucleotide have to be considered (for details, see Refs. [29,38,39]). Note that, for measuring in the reflector mode, the use of DHB as matrix is recommended, since α-cyano-4-hydroxycinnamic acid gives poor or no results.
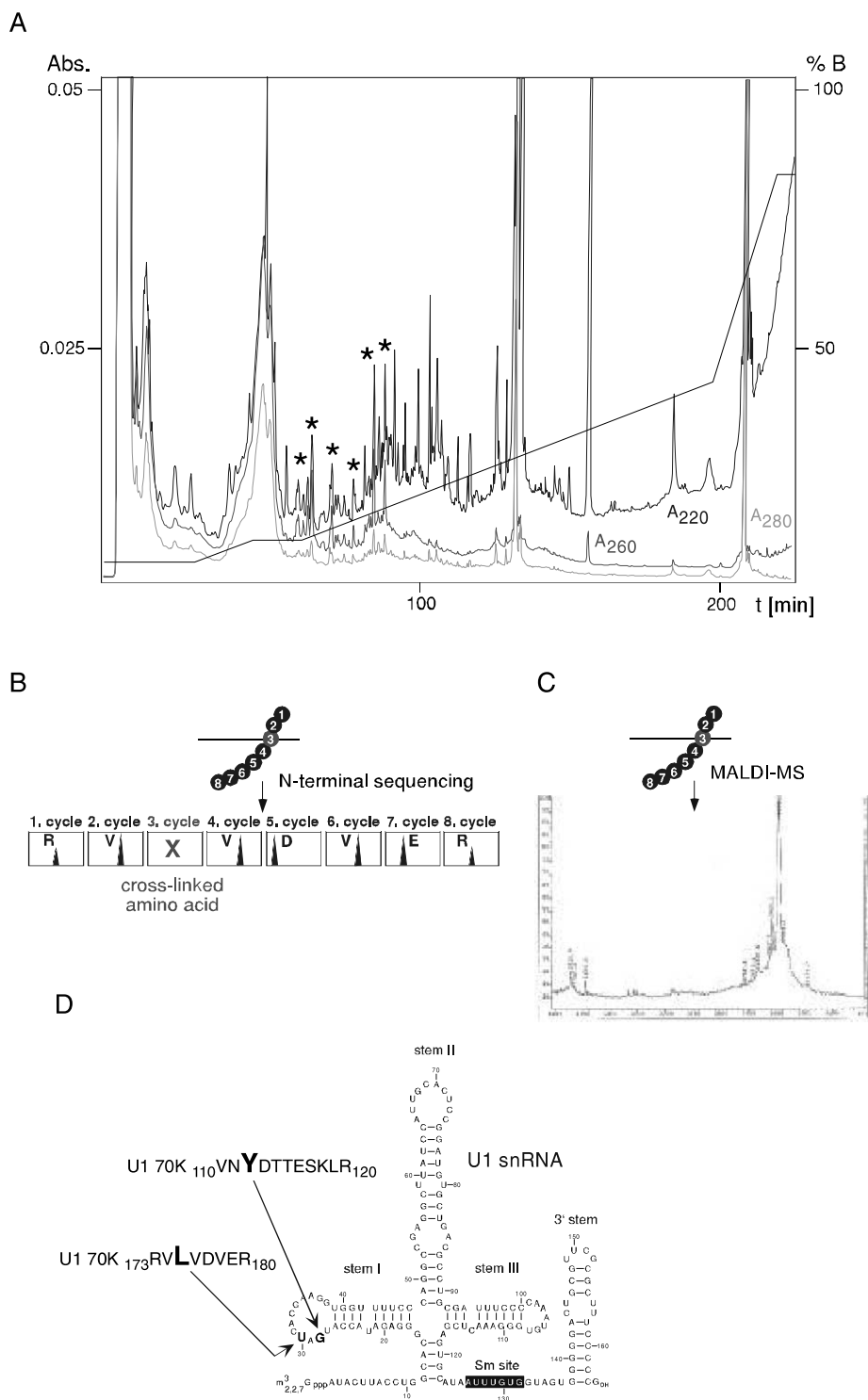
Fig. 5. Final purifcation of peptide–oligonucleotide crosslinks and subsequent analysis of the actual crosslinking site. (A) RP-HPLC of peptide–oligonucleotide crosslinks derived from UV-irradiated U1 snRNPs digested with trypsin and RNase A and T1. Asterisks mark fractions that were considered to contain crosslinks. These fractions were subjected to automated N-terminal sequence analysis (see B) and MALDI-MS (see C). (B) Schematic presentation of identification of crosslinking sites within peptides by Edman degradation. During analysis of the Edman degradation products the crosslinked amino acid cannot be detected. It is denoted with an ×. (C) Schematic presentation of analysis of the oligonucleotide moiety in crosslinked peptide–oligonucleotide complexes. The mass of the crosslinked RNA moiety can be calculated from the mass difference of the measured complex and the mass of the peptide moiety as determined by Edman degradation. From this mass difference the nucleotide composition can be determined. Software tools are available from http://medlib.med.utah.edu/RNAmods/. (D) Summary of crosslinking data obtained by Edmann degradation and MALDI-MS analysis from purified U1 70K peptide–U1 snRNA oligonucleotide complexes. The peptide containing the crosslinked Tyr112 is derived from UV-irradiated U1 snRNPs after digestion with chymotrypsin, whereas that carrying the crosslinked Leu175, after digestion with trypsin. In both cases a mixture of RNase A and T1 was employed.

As summarized in Fig. 5D, the combination of Edman degradation and MALDI-MS enabled us to relate two crosslinking sites within the U1 70K protein to the corresponding crosslinking sites in stem I of U1 snRNA.

4. Concluding remarks

Analysis of the molecular RNA–protein arrangement in mammalian snRNP particles complements the extensive characterization of their protein composition [41,42]. Our two-tracked approach contributes in two ways: (1) It allows the rapid and exact identification of RNA–protein crosslinking sites in isolated snRNP particles. Although it is not as sensitive as when labeled components are used, it has the advantage that it can reveal multiple crosslinking sites between one protein and its cognate RNA, by the combination of primer-extension analysis and immunoprecipitation. (2) It adds valuable data about potential new RNA binding domains within the snRNP-associated proteins. In this manner, we have demonstrated that regions of the Sm1 motif within the Sm B and G proteins interact directly with RNA [30]. So far, the biochemical characterization of Sm motifs 1 and 2 within the Sm proteins has not proceeded beyond establishing their involvement in protein–protein interactions among the Sm proteins [43–45]. More recently, multivariate statistical analysis of negative-stained electron microscopic images and crystallization of Sm-like proteins from the archaebacterium *Archaeoglobus fulgidus* [46,47] revealed a seven-membered ring structure of the assembled Sm proteins on an oligo(U) RNA stretch and, consistent with our data, a direct interaction of a loop structure (loop 3) within the Sm motif 1 of the Sm-like proteins with the oligo(U) RNA.

Another feature of our approach is the possibility of confirming protein–RNA interactions in particles reconstituted in vitro. Once a protein–RNA interaction in native particles is established by our immunoprecipitation or primer-extension method, crosslinking and subsequent analysis of the RNA–protein crosslinking sites in particles reconstituted in vitro should reveal similar interaction patterns, thus confirming the proper conformation and activity of reconstituted components. Working with reconstituted particles also provides an opportunity for the identification of protein sites in contact with RNA, especially when the yield of isolated native particles might be too low.

Furthermore, with the availability of three-dimensional cryo-EM structures of isolated snRNPs, determining where proteins interact with their cognate RNAs is becoming crucial for the interpretation of these structures. In the case of U1 snRNP, our approach has proven a very useful tool for the identification of hitherto uncharacterized protein–RNA interactions at the molecular level (Figs. 1 and 5D). We identified two

distinct amino acids in the RNA binding domain of U1 70K directly contacting bases of two nucleotides in stem I of U1 snRNA. On the basis of these molecular interactions, we proposed a structural model of the protein–RNA interactions that is reminiscent of the protein–RNA interactions of other proteins containing RNA binding domains like U1a and Sxl [29,48,49]. Our approach further demonstrated for the first time that the Sm-site RNA is in direct contact with the inner surface of a heptameric Sm protein ring [30,45] and thus supports the assumption that the Sm-site RNA might traverse the Sm protein ring. Such structural information should be of value in the refinement of three-dimensional structures.

References

[1] C.G. Burd, G. Dreyfuss, Science 265 (1994) 615–621.
[2] J.-M. Perez-Canadillas, G. Varani, Curr. Opin. Struct. Biol. 11 (2001) 53–58.
[3] P.B. Moore, Annu. Rev. Biochem. 68 (1999) 287–300.
[4] D.J. Klein, T.M. Schmeig, P.B. Moore, T.A. Steitz, EMBO J. 20 (2001) 4214–4221.
[5] N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, Science 289 (2000) 905–920.
[6] F. Schluenzen, A, Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, A. Yonath, Cell 102 (2000) 615–623.
[7] B.T. Wimberly, D.E. Brodersen, W.M. Clemons Jr., R.J. Morgan-Warren, A.P. Carter, C. Vonrheinn, T. Hartsch, V. Ramakrishnan, Nature 407 (2000) 327–339.
[8] M.M Yusupov, G.Z. Yusupova, A. Baucom, K. Lieberman, T.N. Earnest, J.H. Cate, H.F. Noller, Science 292 (2001) 883–896.
[9] J. Walleczek, R. Martin, B. Redl, M. Stöffler-Meilicke, G. Stöffler, Biochemistry 28 (1989) 4099–4105.
[10] H.F. Noller, R. Green, G. Heilek, V. Hoffarth, A. Hüttenhofer, S. Joseph, I. Lee, K. Lieberman, A. Mankin, C. Merryman, T. Powers, E.V. Puglisi, R.R. Samaha, B. Weiser, Biochem. Cell. Biol. 73 (1995) 997–1009.
[11] F. Mueller, R. Brimacombe, J. Mol. Biol. 271 (1987) 545–565.
[12] H. Stark, M.V. Rodnina, J. Rinke-Appel, R. Brimacombe, W. Wintermeyer, van Heel, Nature 389 (1997) 403–406.
[13] A. Malhotra, R. Penczek, R.K. Agrawal, I.S. Gabashvili, R.A. Grassucci, R. Jühnemann, N. Burkhardt, K.H. Nierhaus, J. Frank, J. Mol. Biol. 280 (1998) 103–116.
[14] W.M. Clemons Jr., J.L.C. May, B.T. Wimberly, J.P. McCutcheon, M.S. Capel, V. Ramakrishnan, Nature 400 (1999) 833–840.
[15] N. Ban, P. Nissen, J. Hansen, M. Capel, P.B. Moore, T.A. Steitz, Nature 400 (1999) 841–847.
[16] H. Stark, P. Dube, R. Lührmann, B. Kastner, Nature 409 (2001) 539–542.
[17] G. Varani, K. Nagai, Annu. Rev. Biophys Biomol. Struct. 27 (1998) 407–445.
[18] I. Vidovic, S. Nottrott, K. Hartmuth, R. Lührmann, R. Ficner, Mol. Cell 6 (2000) 1331–1342.
[19] C.L. Kielkopf, N.A. Rodinova, M.R. Green, S.K. Burley, Cell 106 (2001) 595–605.
[20] Z. Liu, I. Luyten, M.J. Bottomley, A.C. Messias, S. Houngninou-Molango, R. Sprangers, K. Zanier, A. Krämer, M. Sattler, Science 294 (2001) 1098–1102.
[21] R. Brimacombe, W. Stiege, A. Kyriatsoulis, P. Maly, Methods Enzymol. 164 (1988) 287–309.

[22] B.S. Cooperman, Methods Enzymol. 164 (1988) 341–361.

[23] M.M. Konarska, Methods 18 (1999) 22–28.

[24] M.M. Moore, P.A. Sharp, Science 256 (1992) 992–997.

[25] J.R. Wyatt, E.J. Sontheimer, J.A. Steitz, Genes Dev. 6 (1992) 2542–2553.

[26] I. Dix, C.S. Russell, R.T. O'Keefe, A.J. Newman, J.D. Beggs, RNA 4 (1998) 1675–1686.

[27] R. Reed, M.D. Chiara, Methods 18 (1999) 3–12.

[28] Y.T. Yu, Methods Enzymol. 318 (2000) 71–88.

[29] H. Urlaub, K. Hartmuth, S. Kostka, G. Grelle, R. Lührmann, J. Biol. Chem. 275 (2000) 41458–41468.

[30] H. Urlaub, V. Raker, S. Kostka, R. Lührmann, EMBO J. 20 (2001) 187–196.

[31] C.L. Will, B. Kastner, R. Lührmann, in: S.J. Higgins, B.D. Hames (Eds.), RNA Processing, A Practical Approach, vol. I, IRL Press, Oxford, 1994, pp. 141–177.

[32] B. Kastner, in: J. Schenkel (Ed.), RNP Particles, Splicing and Autoimmune Diseases, Springer-Verlag, Berlin/New York, 1998, pp. 95–140.

[33] V.A. Raker, K. Hartmuth, B. Kastner, R. Lührmann, Mol. Cell. Biol. 19 (1999) 6554–6565.

[34] T. Achsel, K. Ahrens, H. Brahms, S. Teigelkamp, R. Lührmann, Mol. Cell. Biol. 18 (1998) 6756–6766.

[35] M.H. Metz-Boutigue, J. Reinbolt, J.P. Ebel, C. Ehresmann, B. Ehresmann, FEBS Lett. 245 (1989) 194–200.

[36] M.C. Willis, K.A. LeCuyer, K.M. Meisenheimer, O.C. Uhlenbeck, T.H. Koch, Nucleic Acids Res. 22 (1994) 4947–4952.

[37] H. Urlaub, V. Kruft, O. Bischof, E.C. Müller, B. Wittmann-Liebold, EMBO J. 14 (1995) 4578–4588.

[38] H. Urlaub, B. Thiede, E.C. Müller, R. Brimacombe, B. Wittmann-Liebold, J. Biol. Chem. 272 (1997) 14547–14555.

[39] B. Thiede, H. Urlaub, H. Neubauer, G. Grelle, B. Wittmann-Liebold, Biochem. J. 334 (1998) 39–42.

[40] B. Thiede, B. Wittmann-Liebold, M. Bienert, E. Krause, FEBS Lett. 357 (1995) 65–69.

[41] C.L. Will, R. Lührmann, Curr. Opin. Cell Biol. 9 (1997) 320–328.

[42] C.L. Will, R. Lührmann, in: A.R. Krainer (Ed.), Eukaryotic mRNA Processing, Oxford Univ. Press, Oxford/New York/Tokyo, 1997, pp. 130–173.

[43] H. Hermann, P. Fabrizio, V.A. Raker, K. Foulaki, H. Hornig, H. Brahms, R. Lührmann, EMBO J. 14 (1995) 2076–2088.

[44] B. Séraphin, EMBO J. 14 (1995) 2089–2098.

[45] C. Kambach, S. Walke, R. Young, J.M. Avis, E. de la Fortelle, V.A. Raker, R. Lührmann, J. Li, K. Nagai, Cell 96 (1999) 375–387.

[46] T. Achsel, H. Stark, R. Lührmann, Proc. Natl. Acad. Sci. USA 98 (2001) 3685–3689.

[47] I. Törö, S. Thore, C. Mayer, J. Basquin, B. Séraphin, D. Suck, EMBO J. 20 (2001) 2293–2303.

[48] C. Oubridge, N. Ito, P.R. Evans, C.H. Teo, K. Nagai, Nature 372 (1994) 432–438.

[49] N. Handa, O. Nureki, K. Kurimoto, I. Kim, H. Sakamoto, Y. Shimura, Y. Muto, S. Yokoyama, Nature 398 (1999) 579–585.