Complete MALDI-ToF MS analysis of cross-linked peptide–RNA oligonucleotides derived from nonlabeled UV-irradiated ribonucleoprotein particles

EVA KÜHN-HÖLSKEN,[1,4] CHRISTOF LENZ,[2,4] BJÖRN SANDER,[3] REINHARD LÜHRMANN,[3] and HENNING URLAUB[1]

[1]Bioanalytical Mass Spectrometry Group and [3]Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany
[2]Applera Deutschland GmbH, 64293 Darmstadt, Germany

ABSTRACT

Protein–RNA cross-linking combined with mass spectrometry is a powerful tool to elucidate hitherto noncharacterized protein–RNA contacts in ribonucleoprotein particles, as, for example, within spliceosomes. Here, we describe an improved methodology for the sequence analysis of purified peptide–RNA oligonucleotide cross-links that is based solely on MALDI-ToF mass spectrometry. The utility of this methodology is demonstrated on cross-links isolated from UV-irradiated spliceosomal particles; these were (1) [15.5K–61K–U4atac] small nuclear ribonucleoprotein (snRNP) particles prepared by reconstitution in vitro, and (2) U1 snRNP particles purified from HeLa cells. We show that the use of 2′,4′,6′-trihydroxyacetophenone (THAP) as MALDI matrix allows analysis of cross-linked peptide–RNA oligonucleotides in the reflectron mode at high resolution, enabling sufficient accuracy to assign unambiguously cross-linked RNA sequences. Most important, post-source decay (PSD) analysis under these conditions was successfully applied to obtain sequence information about the cross-linked peptide and RNA moieties within a single spectrum, including the identification of the actual cross-linking site. Thus, in U4atac snRNA we identified His270 in the spliceosomal U4/U6 snRNP-specific protein 61K (hPrp31p) cross-linked to U44; in the U1 snRNP we show that Leu175 of the U1 snRNP-specific 70K protein is cross-linked to U30 of U1 snRNA. This type of analysis is applicable to any type of RNP complex and may be expected to pave the way for the further analysis of protein–RNA complexes in much lower abundance and/or of cross-links that are obtained in low yield.

Keywords: RNA; protein; cross-linking; mass spectrometry; hPrp31

INTRODUCTION

RNA molecules play a fundamental role in cellular processes such as gene expression (transcription and pre-mRNA 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 particles, much attention is currently being devoted to the question of protein and RNA tertiary structures and the quaternary arrangement of the individual macromolecules in the RNP particles. Together with the fact that the number of isolated RNP (sub-) complexes from viruses, eubacteria, archaea, and eukaryotes is steadily increasing, one important challenge in this field is the development of a highly sensitive and generally applicable method that allows the identification of hitherto uncharacterized RNA binding proteins and the localization of their regions of interactions.

Cross-linking is the method of choice to link a protein covalently to its cognate RNA in RNP particles, and, most importantly, it is so far the only technique that is capable of proving that a certain protein is indeed in direct contact with RNA in RNP particles whose structures are incompletely characterized. Cross-linking between RNA and proteins can be achieved by incorporating base analogs (e.g.,

5-bromo-2′-deoxyuridine or 4-thio-uracil; Hicke et al. 1994; Sontheimer 1994; Meisenheimer and Koch 1997; Meisenheimer et al. 2000) into the RNA, either site specifically or randomly; by chemical modification of the RNA backbone (Konarska 1999); or by direct UV-irradiation at 254-nm wavelength (Görner 1994). The latter method owes its usefulness to the natural photoreactivity of RNA bases. After the cross-linking reaction, cross-linked proteins can be identified, for example, by their migration behavior in SDS-PAGE (for [^{32}P]-labeled RNA components in reconstituted RNP particles; Greuer et al. 1999; Reed and Chiara 1999; Expert-Bezancon et al. 2004), by their ability to react with a specific antibody in Western blot analysis (Möller and Brimacombe 1975; Dix et al. 1998), or by immunoprecipitation techniques (Urlaub et al. 2002; Hamon et al. 2004; Bonnal et al. 2005). Although these techniques are highly sensitive, none of them would be able to identify unexpected or unknown proteins in contact with the RNA.

Electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Karas and Hillenkamp 1988; Fenn et al. 1989) are by far the most sensitive techniques available to sequence and subsequently to identify novel proteins from a sample. Consequently, in several studies, UV cross-linking has been combined with mass spectrometry (MS) techniques in order to identify proteins that are covalently linked to nucleic acids. In this manner, proteins have been directly identified, for example, from SDS-PAGE after they have been (site-specifically) cross-linked to a [^{32}P]-labeled DNA (Coman et al. 2004) or RNA molecule (Rhode et al. 2003).

Although Jensen et al. (1993) successfully introduced the analysis of peptide–ssDNA cross-links by MALDI-ToF MS (but without sequencing), most of the preceding studies used ESI MS to sequence peptide–nucleotide cross-links by collision-induced dissociation (CID) (Smith et al. 1990), while post-source decay (PSD) (Kaufmann et al. 1993) of cross-linked heteroconjugates has rarely been performed. Very recently, Geyer et al. (2004) reported the PSD sequence analysis of a peptide–ssDNA cross-link where the entire DNA moiety was completely hydrolyzed by hydrofluoric acid, so that the cross-linked and subsequently sequenced peptide only carried the DNA base. It should also be noted that most of these MS studies investigated DNA–protein cross-links (Golden et al. 1999; Wong and Reich 2000; Steen et al. 2001; Coman et al. 2004; Doneau et al. 2004), but only a very few dealt with RNA–protein cross-links (Meisenheimer et al. 1996; Urlaub et al. 2002).

In recent years we have developed a purification strategy that has enabled us to isolate peptide–oligoribonucleotide cross-links directly from native RNP particles, such as small nuclear ribonucleoprotein particles (snRNPs) (Will and Lührmann 1997, 2001; Burge et al. 1999; Henras et al. 2004), after these have been UV-irradiated at 254 nm (Urlaub et al. 2002) and subsequently treated with specific endoproteinases and nucleases. However, this approach is subject to

limitations with regard to the sequence determination of the purified cross-links—in particular, the sequence of the cross-linked peptide moiety. Whereas the cross-linked RNA part could be identified by a combination of MALDI-ToF MS (in the linear mode and thus at a low resolution) and primer-extension analysis on the cross-linked RNA, the cross-linkedpeptide sequence and the actual cross-linked amino acid could only be unambiguously identified by N-terminal sequencing (Urlaub et al. 2002). Consequently, large amounts of sample were necessary to obtain a quantity of cross-linked material (5–10 pmol) sufficient for this type of sequence determination.

MALDI-ToF MS seems ideally suited to establish a more sensitive analytical approach, and it also offers the opportunity to reinvestigate samples on the target. Previous MALDI analyses suffered from one major weakness, though: The strongly divergent physico-chemical properties of isolated peptide–RNA heteroconjugates made it difficult to choose a suitable matrix for sample preparation. Furthermore, MALDI PSD analysis of (e.g., post-translationally) modified peptides does not allow the establishment of routine protocols, and, in comparison with ESI MS/MS, it requires much larger amounts of sample. Nonetheless we chose to improve the MALDI-ToF MS methodology owing to its above-mentioned advantages, and further to the fact that MALDI-ToF MS dovetails perfectly with our established purification protocol.

We therefore set out to find conditions for the MALDI-ToF MS-based analysis of cross-linked complexes that gave sufficient accuracy to allow the characterization of both the peptide and the RNA moiety in peptide–RNA heteroconjugates. As starting material, we chose RNP subcomplexes of the human splicing machinery with cross-linking products that have in parts been defined as: (1) [15.5K–61K–U4atac] RNP particles prepared by reconstitution in vitro (Nottrott et al. 2002) and (2) native U1 snRNP particles (Kastner and Lührmann 1999), in order to gain confidence in the new method.

Here, we describe our most recent advances in the complete sequence determination (peptide and oligoribonucleotide) of the cross-links derived from both these particles, using MALDI-ToF MS only. The use of 2′,4′,6′-trihydroxyacetophenone monohydrate (THAP) and 2,5-dihydroxy benzoic acid (DHB) as matrices enables us to analyze the samples in an accurate manner in the reflectron mode, and this in turn allows us (1) subsequently to perform mass comparisons with putative cross-linked peptide and RNA stretches and, finally, (2) to perform PSD sequence analysis in order to gain sequence and structural information about the cross-linked components within a single spectrum. Additionally, this method has enabled us to gain insight into the UV-induced photoreaction between an amino acid side chain and an RNA base. We consider this analysis to represent a significant breakthrough for the analysis of other RNPs, since it reduces the amount of

sample required on the MALDI target by a factor of 50–100 down to 100 fmol of cross-linked sample, so that low-abundance RNPs and cross-links obtained in lower yield can be investigated.

RESULTS

2,5-DHB and 2′,4′,6′-THAP matrices enhance peptide–RNA oligonucleotide detection in MALDI-ToF-MS

Linear mode MALDI-ToF mass spectrometry has previously been used to characterize peptide–RNA oligonucleotide cross-links at low mass spectrometric resolution (Urlaub et al. 2000, 2001, 2002). To provide more detailed characterization, we sought conditions that enable (1) an accurate analysis of the cross-links' intact molecular mass in the reflectron mode and (2) PSD analysis to obtain structural information about the cross-linked moieties, both at high sensitivity, that is, in the low femtomole range. We therefore first tested various matrices for their ability to produce highly accurate spectra from very small quantities of cross-linked samples (25 μg of non-cross-linked starting material). For this purpose we used [15.5K–61K–U4atac] snRNP particles that had been prepared by reconstitution in vitro (Nottrott et al. 2002) and native U1 snRNP particles. After UV irradiation of these particles, the protein moiety was digested with the endoproteinases trypsin and/or chymotrypsin. Intact RNA and RNA carrying cross-linked peptides were separated from the excess of non-cross-linked peptides by size-exclusion chromatography, and the RNA-containing fractions were hydrolyzed with RNase A and/or RNase T1. The final purification step involved chromatography on a microbore C18 RP-HPLC column, in which cross-linked peptide–RNA oligonucleotides were separated from non-cross-linked RNA oligonucleotides; the latter were mainly present in the flow-through, while cross-linked peptide–oligonucleotides eluted at a higher percentage of organic solvent and were identified during chromatography by their UV absorbance at 220 nm and 260 nm. Fractions showing an absorbance at both of these wavelengths are considered to contain peptides as well as RNA, and these were isolated and analyzed by MALDI-ToFMS. The overall purification strategy is summarized in Figure 1.

A sufficiently accurate analysis of the molecular mass of the cross-linked heteroconjugates in those fractions should enable us to determine the composition of the cross-linked peptide and the RNA fragment by comparing the measured mass of the cross-linked moiety with the theoretical masses of (1) all possible chymotryptic/tryptic fragments of the protein and (2) all possible RNase A/T1 fragments of the RNA of the initial particle.

Figure 2 shows a comparison of results obtained with various matrices for MALDI-ToF MS. We initially tested α-cyano-4-hydroxycinnamic acid (CHCA), a "hot" matrix,
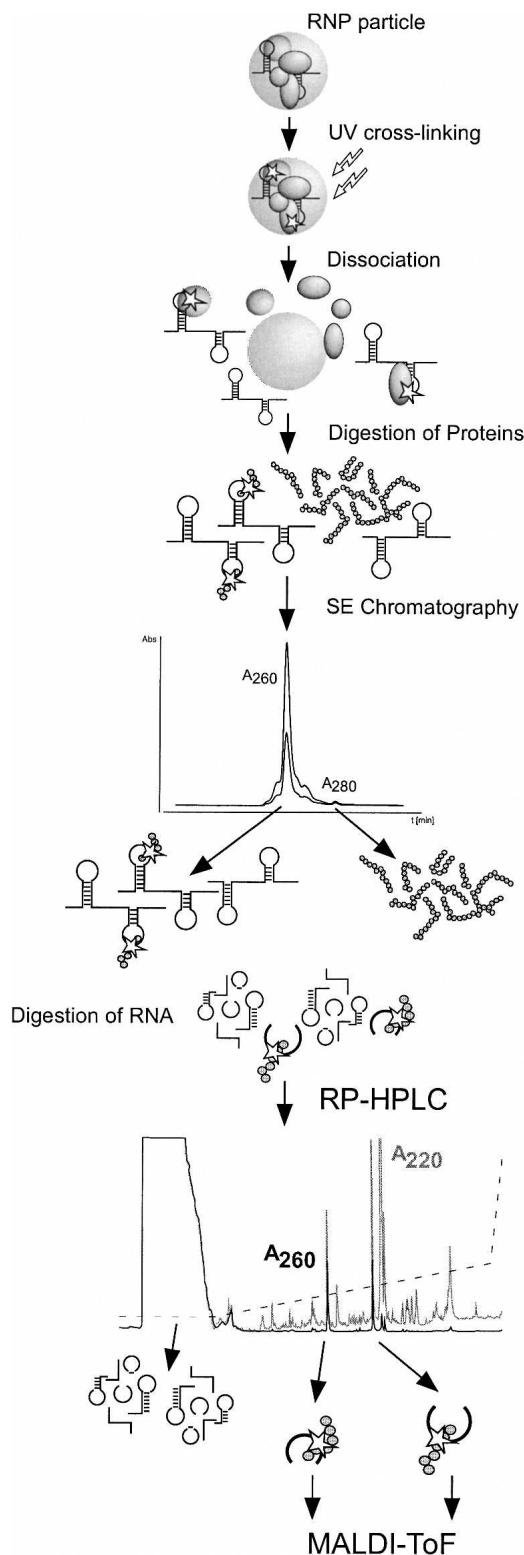
and observed a significant decomposition of cross-linked peptide–RNA oligonucleotides produced by treatment with RNase T1. Figure 2, A and B, shows the MALDI spectra recorded in reflectron and linear modes, respectively, from an RP-HPLC fraction containing a putative cross-linked peptide–RNA oligonucleotide derived from UV-irradiated [15.5K–61K–U4atac] RNP particles digested with chymotrypsin and RNase T1. Consistent with our earlier studies, Figure 2 shows that cross-linked particles with a significant RNA component (i.e., those obtained by digestion with RNase T1) could only be detected in the linear mode when CHCA was used as matrix. The linear mode, however, gave poor mass accuracy and no information about isotopic patterns. This made it difficult to determine unambiguously the cross-linked components—in particular the nucleotide composition of the cross-linked RNA oligonucleotide, since nucleotides C and U differ in mass by only one unit.

In contrast to the results obtained with CHCA, we were able to obtain high quality spectra in the reflectron mode when using DHB or THAP as matrices. Strikingly, both matrices dramatically enhance the intensity of the quasi-molecular ion $[M + H]^+$ of the isolated cross-linked species observed at a monoisotopic *m/z* of 2780.818 (Fig. 2C,D). The THAP preparation showed the highest sensitivity and the lowest amount of spontaneous fragmentation of the analyte (Fig. 2D). Both DHB and THAP preparations yielded resolutions higher than 15,000 FWHH (full width at half height). The mass accuracy with close external calibration was evaluated with standard peptides and found to be 50 ppm or better with both matrices. The enhanced sensitivity together with the high resolution and accuracy enabled us to determine the compositions of the cross-linked peptide and RNA moieties. Importantly, we also analyzed cross-linked samples in the negative mode with THAP as matrix. Previous work by Zhu et al. (1996) demonstrated that oligonucleotides analyzed in the negative linear mode with THAP as matrix gave excellent results in terms of signal intensity. We also observed an increase in signal intensity of the cross-linked precursor in the negative reflectron mode compared with the positive reflectron mode (data not shown). However, we did not use these conditions for further analysis of cross-links, since PSD analysis in the negative mode (data not shown) only showed poor fragmentation of the cross-linked peptide moiety.

Accurate mass determination of the cross-linked moiety by high-resolution MALDI-ToF spectra using THAP and DHB as matrix

We used THAP and DHB preparations for the accurate mass determination of peptide–RNA cross-links. Figure 3A shows the isotopically resolved quasimolecular ion region of a cross-linked peptide–oligonucleotide obtained

after treatment of the initially UV-irradiated particles with chymotrypsin and RNase T1 at m/z 2780.818. The sample was measured in the reflectron mode using THAP as matrix and close external calibration. To determine the peptide

FIGURE 1. Schematic representation of the purification of cross-linked peptide–RNA oligonucleotides derived from ribonucleoprotein particles, either native or prepared by reconstitution in vitro. The RNP particles are UV–irradiated and then made to dissociate. Subsequently, the protein moiety is digested with endoproteinases trypsin or chymotrypsin. RNA that carries cross-linked peptides and intact RNA are separated from non-cross-linked peptides by size-exclusion (SE) chromatography. RNA-containing fractions are hydrolyzed with ribonucleases A and/or T1, and the resulting mixture of non-cross-linked oligonucleotides and cross-linked peptide–RNA oligonucleotides is subjected to a final purification step by RP-HPLC. Fractions showing an absorbance at 220 nm and 260 nm are presumed to contain cross-linked peptide–RNA heteroconjugates and are thus analyzed in the MALDI-ToF mass spectrometer.

and RNA composition, we compared all chymotryptic fragments of the protein moiety with all RNase T1 fragments of the U4atac RNA moiety. We found that the measured monoisotopic mass matches accurately with only one chymotryptic fragment of the 61K protein, encompassing protein positions 263–273 (SSTSVLPHTGY M_{cal} = 1147.551), and with only one RNase T1 fragment of U4atac RNA, comprising nucleotide positions C42–G46 (5′-CAUAG-3′, M_{cal} = 1632.227). Assuming additive behavior of the peptide and RNA molecular masses from the cross-linking reaction, the mass accuracy was found to be 28 ppm. The next best hit from this calculation (61K protein positions 215–232, M_{cal} = 1782.019, and U4atac RNA nucleotide positions C53–G55, M_{cal} = 997.150) has a significantly higher mass deviation (132 ppm) and was thus not considered to be the actual cross-linked complex. From this analysis we conclude that protein 61K must be cross-linked to U4atac RNA between residues 263 and 273 on the protein and nucleotides C42 and G46 on the RNA.

MALDI-ToF MS analysis of another RP-HPLC fraction obtained by digestion of UV-irradiated [15.5K–61K–U4atac] RNP particles with chymotrypsin and RNases A and T1 revealed a peak at m/z 1801.646 (Fig. 3B), corresponding within 22 ppm to the same chymotryptic peptide of protein 61K comprising the same positions 263–273 (SSTSVLPHTGY, M_{cal} = 1147.551) cross-linked to an AU (M_{cal} = 653.087) dinucleotide. Six such AU dinucleotides are present in the sequence of the U4atac snRNA (A5–U6, A35–U36, A43–U44, A85–U86, A117–U118, and A129–U130). Notably, A43–U44 is located within the cross-linked T1–oligonucleotide CAUAG, encompassing nucleotide (nt) positions 42–46. We therefore conclude that 61K protein is cross-linked to this particular dinucleotide on the U4atac snRNA.

In addition, we isolated a putative cross-link after digestion of UV-irradiated trimeric RNP complexes with trypsin and RNases A and T1 from the RP-HPLC. MALDI-ToF analysis revealed two major peaks at m/z 3535.489 and m/z 3260.444. However, a database search gave no positive result for any tryptic fragment cross-linked to an RNaseA/
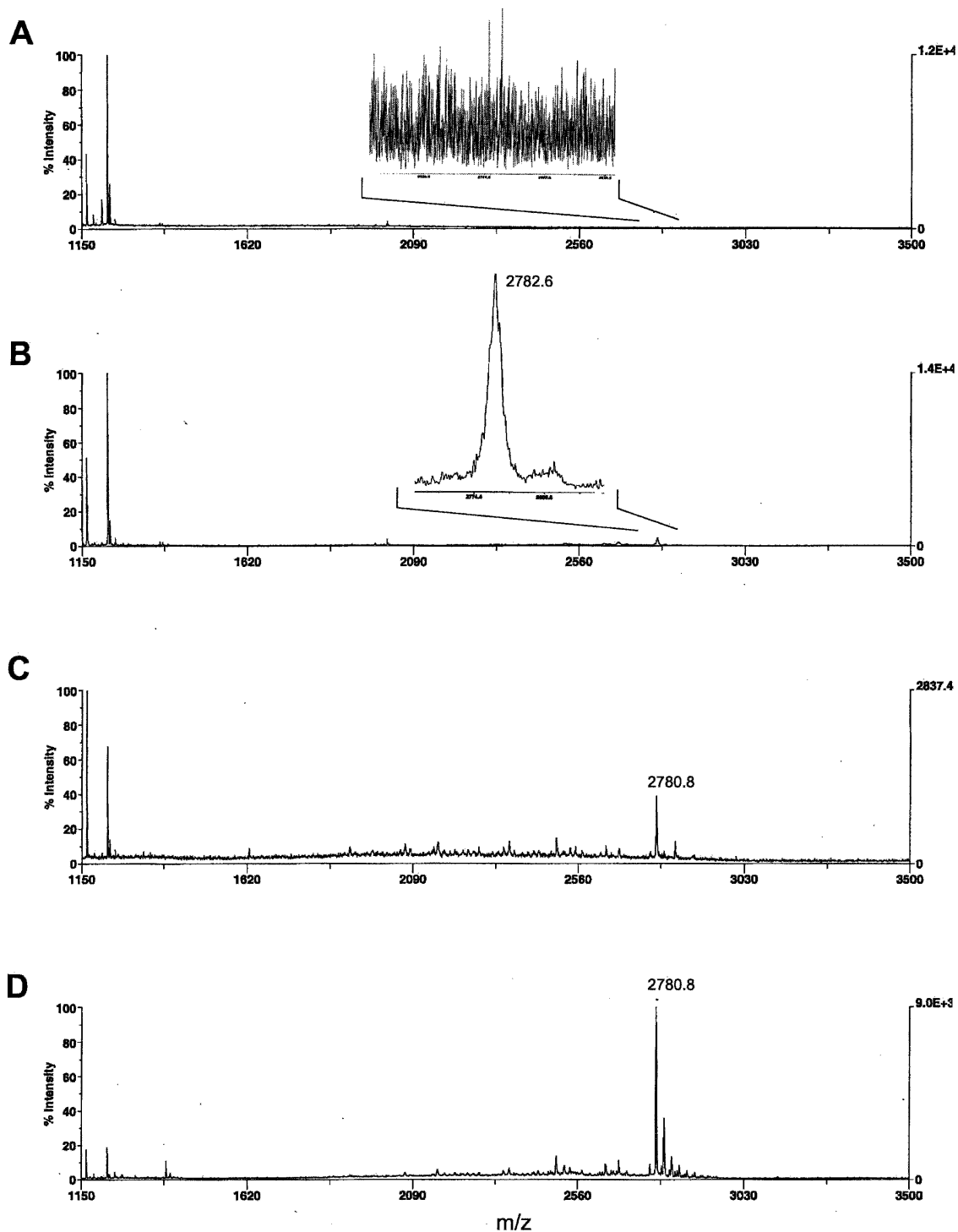
FIGURE 2. Comparison of different matrices for the analysis of cross-linked peptide–RNA heteroconjugates in MALDI-ToF. Equal volumes of cross-links in RP-HPLC fractions were mixed with different matrices and analyzed by MALDI-ToF mass spectrometry under various conditions: (*A*) with CHCA as matrix and analysis in the reflectron mode; (*B*) with CHCA as matrix and analysis in the linear mode; (*C*) with DHB as matrix and analysis in the reflectron mode; and (*D*) with THAP as matrix and analysis in the reflectron mode. Cross-links are derived from UV-irradiated [15.5K–61K–U4atac] RNPs after treatment with chymotrypsin and RNase T1, and thus are expected to carry a larger RNA moiety. All the spectra were recorded on a Voyager DE-STR MALDI mass spectrometer (Applied Biosystems) under standard conditions. In *A* and *B,* the *inset* shows magnified the *m/z* values that are expected to show a mass peak from the putatively cross-linked species.

A

M exp	[M+H]⁺ cal	M cal	M cal	ppm
peptide-oligo XL	61K / 15.5K chymotrypsin	U4atac RNA RNase T1	peptide-oligo XL	
2780.818	1148.588 61K pos. 263-273	1632.227 U4atac pos. C42-G46	2780.785	28
	1784.035 61K pos. 215-232	997.150 U4atac pos. C53-G55	2781.185	132

B

M exp	[M+H]⁺ cal	M cal	M cal	ppm
peptide-oligo XL	61K / 15.5K chymotrypsin	U4atac snRNA RNase T1/A	peptide-oligo XL	
1801.685	1148.588 61K pos. 263-273	653.067 U4atac pos. A43-U44	1801.645	22

C

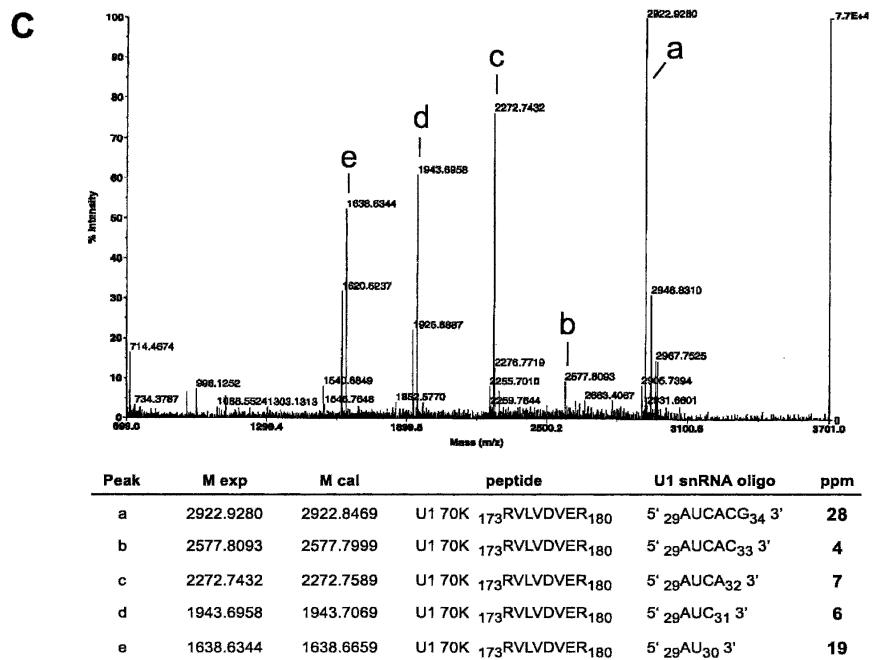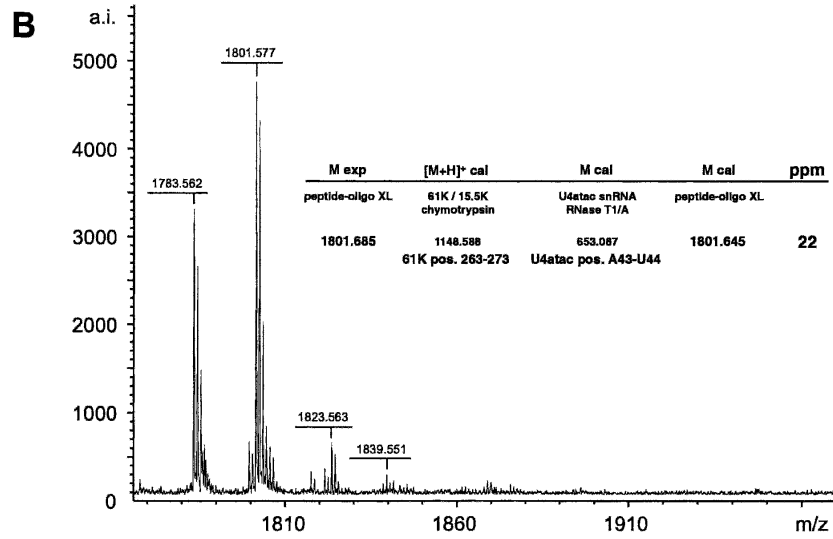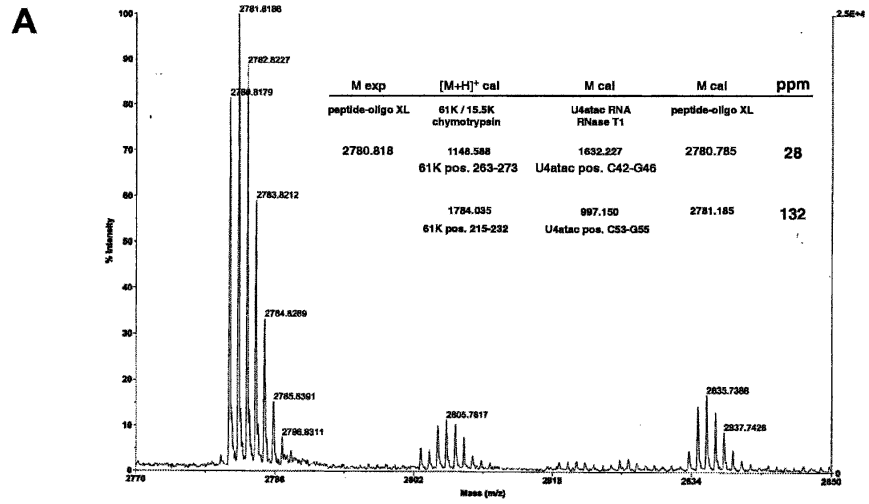| Peak	M exp	M cal	peptide	U1 snRNA oligo	ppm
a	2922.9280	2922.8469	U1 70K $_{173}$RVLVDVER$_{180}$	5' $_{29}$AUCACG$_{34}$ 3'	28
b	2577.8093	2577.7999	U1 70K $_{173}$RVLVDVER$_{180}$	5' $_{29}$AUCAC$_{33}$ 3'	4
c	2272.7432	2272.7589	U1 70K $_{173}$RVLVDVER$_{180}$	5' $_{29}$AUCA$_{32}$ 3'	7
d	1943.6958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

FIGURE 3. (Legend on next page)

T1 fragment of the U4atac snRNA, either from protein 61K or from protein 15.5K. This result illustrates the fundamental need for structural information about the cross-linked species, as yielded by PSD analysis (see below).

Figure 3C gives a third example of the determination of the cross-linked peptide and RNA moieties from a single MS spectrum under the improved conditions that we describe. This spectrum shows the MALDI-ToF MS analysis of an RP-HPLC fraction containing a cross-link derived from native UV-irradiated U1 snRNPs treated with trypsin and RNase T1. We observed several predominant peaks (a–e, see Fig. 3C). 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; Will and Lührmann 2001) with all possible U1 snRNA RNase T1 fragments in a database search, we found that these peaks correspond to a U1 70K tryptic fragment from positions 173–180 (RVLVDVER, M_{cal} = 984.572) cross-linked to a U1 snRNA T1 fragment encompassing positions 29–34 (5′-AUCACG-3′, M_{cal} = 1937.268) and to 3′ hydrolysis products of the latter (see the table in Fig. 3C). Notably, the mass accuracy over the entire spectrum is >30 ppm with close external calibration.

Note that our database search with the highest precursor of *m/z* 2922.928 gave several other hits with a similar mass accuracy: (1) residues 238–246 of protein U1-70K (ERER-RERSR, M_{cal} = 1272.675) putatively cross-linked to nt 77–81 of U1 snRNA (5′-AUGUG-3′, M_{cal} = 1649.206) with a mass accuracy of 12 ppm, and (2) positions 85–92 of protein Sm D2 (KKSKPVNK, M_{cal} = 927.587) putatively cross-linked to nt 76–81 of U1 snRNA (5′-GAUGUG-3′, M_{cal} = 1994.256) with a mass accuracy of 26 ppm. We consider these combinations to be highly unlikely, as our analysis unambiguously reveals the stepwise loss of the nucleotides G, C, A, and C. The alternative RNA oligonucleotides (5′-AUGUG-3′ and 5′-GAUGUG-3′, respectively) should have shown two losses of U, which were not observed in this experiment. The latter reason again illustrates the substantial improvement that is achieved by using matrices that allow the recording of spectra of peptide–RNA heteroconjugates with a high accuracy in the reflectron mode. Note that without any previous knowledge of sequence information and under nonimproved

MALDI conditions (i.e., analysis in the linear mode), the observed hydrolysis products of the RNA part would also have matched the alternative RNA sequences, since measuring in the linear mode does not allow one to distinguish between C and U, which differ in mass by only one unit.
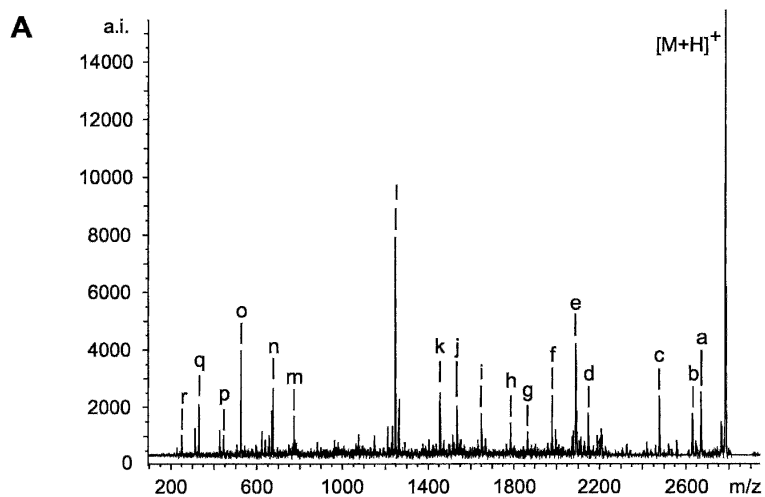
PSD analysis of cross-linked peptide–RNA heteroconjugates

Having established conditions for the sensitive and accurate analysis of peptide–RNA cross-links in the reflectron mode using THAP (or DHB) as matrix, we next performed PSD analysis on the putative peptide–RNA heteroconjugates to obtain further structural information about the cross-linked species.

Figure 4A demonstrates that our THAP (DHB) preparation also showed excellent properties in the PSD analysis of the selected peptide–RNA oligonucleotide precursors. Figure 4A shows the PSD analysis of the precursor at *m/z* 2780.818; this was calculated as being the spliceosomal 61K protein (positions 263–273) cross-linked to a pentanucleotide 5′-CAUAG-3′ (nt 42–46) of the U4atac snRNA. Strikingly, the fragment pattern obtained yields significant sequence information about the cross-linked RNA moiety. Figure 4B shows the relationship between the measured fragment ions and the RNA sequence, according to the nomenclature introduced by McLuckey et al. (1992).

A stepwise fragmentation of the oligonucleotide moiety from both its 5′ and its 3′ ends can be observed up to the cross-linked base in the PSD spectrum; the *m/z* values of the ions consistently represent the mass of the remaining RNA moiety added to the mass of the intact cross-linked peptide sequence. The *m/z* values of the fragment ions a and b represent losses of cytosine and guanine, respectively, indicating that a cytidine is at the 5′ end and a guanosine is at the 3′ end (RNase T1 cleavage site). Together with the observed losses of nucleotides C and A and of a dinucleotide AG, these findings both support the assignment of the RNA fragment to the sequence 5′-CAUAG-3′ as previously calculated.

FIGURE 3. Highly resolved mass peaks of purified peptide–oligoribonucleotides allow the calculation of the mass of the putatively cross-linked peptide and RNA parts. (*A*) Highly resolved mass peak of a purified cross-link derived from UV-irradiated [15.5K–61K–U4atac] RNPs treated with chymotrypsin and RNase T1. The spectrum was recorded in the reflectron mode with THAP as matrix. From the measured monoisotopic mass we calculated the listed peptide sequence of the protein 61K cross-linked to an RNase T1 fragment of U4atac snRNA. The mass deviation is less then 30 ppm. The next best hit from such a calculation is also listed; notably, the mass deviation is already 132 ppm. (*B*) Highly resolved mass peak of a purified cross-link derived from UV-irradiated [15.5K–61K–U4atac] RNPs treated with chymotrypsin and RNases A/T1. From the measured monoisotopic mass we calculated the listed cross-linked sequences of protein 61K and U4atac snRNA. As in spectrum *A*, the mass deviation is <30 ppm. Note that the peak at *m/z* 1783.667 is caused by the loss of water upon formation of 2′,3′-cyclic phosphates at the 3′ end of the cross-linked RNA. (*C*) Highly resolved mass spectrum of a purified cross-link derived from UV-irradiated U1 snRNPs treated with endoproteinase trypsin and RNase T1. The measured monoisotopic masses match precisely a tryptic fragment of the U1 protein 70K from positions 173–180 cross-linked to a U1 snRNA T1 fragment from positions 29–34 and to its 3′-hydrolysis products, respectively. Notably, the mass accuracy over the entire spectrum is again >30 ppm. All spectra were recorded in the reflectron mode. *A* and *C* were recorded on a Voyager DE-STR MALDI mass spectrometer (Applied Biosystems) under standard conditions using THAP as matrix. Spectrum *B* was recorded on a Reflex IV MALDI mass spectrometer (Bruker Daltonics) with DHB as matrix.

A

a.i.

14000

12000

10000

8000

6000

4000

2000

0

$[M+H]^+$

200   600   1000   1400   1800   2200   2600   m/z

(peak labels: r, q, p, o, n, m, k, j, i, h, g, f, d, e, c, b, a)

B

Peak	Fragment	Sequence
a	w_4 + ribose*	CAUAG – C'*
b	c_4 + ribose + P_i*	CAUAG – G'*
c	y_{r4}*	AUAG*
d	y_{r3}*	UAG*
e	c_3*	CAU*
f	$c_3 - B_1$ (C) *	CAU – C'*
g	$c_3 - a_1$* / $w_4 - y_{r2}$* $c_4 - a_2$* / $w_3 - y_{r1}$*	AU + P_i* UA + P_i*
h	$a_3 - a_1$* / $c_3 - c_1$* / $w_4 - w_2$*/ $y_{r4} - y_{r2}$* $a_4 - a_4$* / $c_4 - c_4$* / $w_3 - w_1$* / $y_{r3} - y_{r1}$*	AU* UA*
i	$a_3 - a_1 - B_2$ (A)* / $c_3 - c_1 - B_2$ (A) */ $y_{r4} - y_{r2} - B_2$ (A)* / $w_4 - w_2 - B_2$ (A)* $a_4 - a_2 - B_4$ (A)*/ $c_4 - c_2 - B_4$ (A)* / $y_{r3} - y_{r1} - B_4$ (A)* / $w_3 - w_1 - B_4$ (A)*	AU – A'* UA – A'*
j	$c_3 - a_2$* / $w_3 - y_{r2}$*	U + P_i*
k	$c_3 - c_2$*/ $w_3 - w_2$*	U*
l	$a_3 - c_2 - $ ribose* / $y_{r3} - w_2 - $ ribose*	U'*
m	w_2	AG + P_i
n	y_{r2}	AG
o	$y_{r2} - B_5$ (G)	AG – G'
p	$y_{r2} - z_1$ + ribose	A + ribose
q	$y_{r2} - y_1$	A
r	$y_{r2} - w_1$	A – P_i

C

B_1 (C)

B_2 (A)

peptide ($_{263}$SSTSVLPHTGY$_{273}$)

B_3 (U)

B_4 (A)

B_5 (G)

(labels: a_1, w_4, y_{r4}, c_1, a_2, w_3, y_{r3}, c_2, a_3, w_2, y_{r2}, c_3, a_4, w_1, y_{r1}, c_4; nucleotide labels -C', -U', -G')
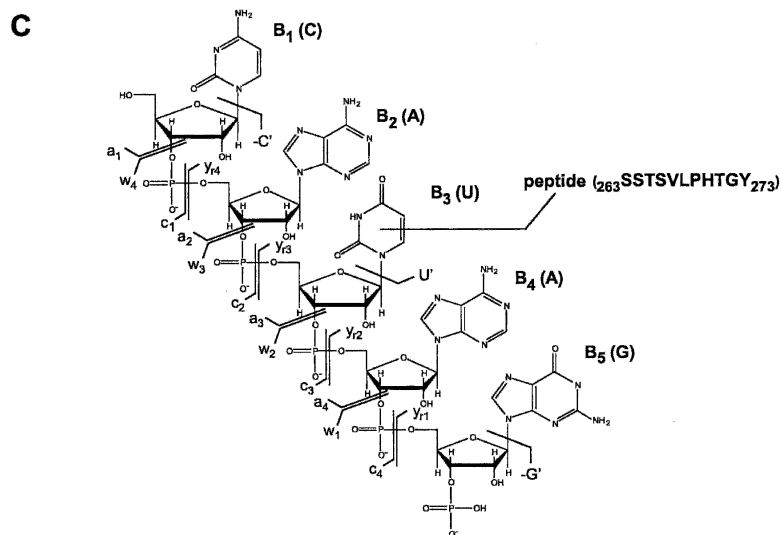
FIGURE 4. (Legend on next page)

Starting from fragment ion e, which represents the peptide carrying a CAU trinucleotide, one can follow the stepwise loss of the components of the nucleotides, that is, phosphates ($\Delta m = 80/98$), riboses ($\Delta m = 114$), and bases ($\Delta m = 111$ [C′], and 135 [A′]) up to the most prominent fragment l, which represents the peptide moiety cross-linked to a uracil. In the low-mass region of the PSD spectrum, complete fragmentation of the separate AG dinucleotide is observed (fragment ions m–r).

These PSD results demonstrate clearly that the cross-linked RNA part is indeed the pentanucleotide T1 fragment 5′-CAUAG-3′, nt 42–46 of the U4atac snRNA, which we calculated from the reflectron mode analysis. Moreover, we did not observe any loss of uridine or uracil, suggesting that U44 within the T1 fragment is the actual cross-linked nucleotide. Figure 4C summarizes schematically the observed fragmentation.

The PSD analysis of this precursor gave no sequence or structural information about the cross-linked peptide moiety. This result is not unexpected, as the relatively large oligoribonucleotide moiety easily fragments under PSD conditions and interferes with the detection of peptide fragments. A similar phenomenon was observed in ESI-CID analysis of a peptide-ssDNA cross-link (Golden et al. 1999; Steen et al. 2001). Comparable behavior is frequently observed during the analysis of glycosylated peptides (Blom et al. 2004; Zaia 2004; Zamfir and Peter-Katalinic 2004).

By performing PSD analysis of the smaller precursor at m/z 1801.646, which was provisionally assigned to the same peptide sequence being cross-linked to the shorter AU dinucleotide, we obtained sequence information about the cross-linked peptide. Figure 5A shows the corresponding PSD spectrum, and the results are summarized in Figure 5B. The spectrum shows a b-type fragment ion series of the cross-linked peptide from b_3 to b_7, revealing the peptide sequence TSVLP, and RNA fragments of w- and y-type, namely w_1, y_m, and $y_m - w_{-1}$, carrying the entire cross-linked peptide, as well as several combinations of these RNA fragments with peptide fragments of the y-series. Furthermore, the loss of adenine ($\Delta m = 135$) and adenosine ($\Delta m = 329$) and its sugar phosphate identifies uridine within the dinucleotide as the actual cross-linked nucleotide. Importantly, we observed a y-type series $y_3 - y_6$ with the resulting sequence LPHT (Fig. 5B). The absolute m/z values of this series indicate that peptide fragments are covalently linked to an AU dinucleotide or a U mononucleotide. Strikingly, the fragment with the observed m/z of 444.2 corresponds to a His–Uridine heteroconjugate, suggesting that His270 in the spliceosomal 61K protein is the amino acid actually cross-linked to U44 in the U4atac snRNA. Figure 5C summarizes the results schematically. Interestingly, the fact that we observed a fragment ion that corresponds to a His–Uridine conjugate illustrates that the cross-link reaction must be a radical-induced addition without loss of any H-radical in either component, histidine or uridine. We thus assume that the reaction takes place by a mechanism similar to that proposed by Norris et al. (1996, 1997) who used 5-halogen- (bromo- or iodo-) substituted uracil bases, with the exception that in our case the radical-induced formation of the novel covalent bond between the amino acid and the uracil base leads presumably to the opening of the C–C double bond in the uracil base, as no radical-induced loss of hydrohalide acids is possible.

As mentioned above, we isolated two other putative peptide–RNA oligonucleotides from the UV-irradiated [15.5K–61K–U4atac] trimeric complex after digestion with trypsin and RNases A/T1. However, our calculations based on the measured masses (m/z 3260.444 and m/z 3535.489) did not reveal any specific tryptic fragment derived from either the 61K or 15.5K protein cross-linked to an RNase A/T1 oligonucleotide of U4atac snRNA. Figure 6A shows the PSD sequence analysis of one of these particular precursors (m/z 3260.444). The PSD spectrum shows an almost complete series of y-type fragments up to y_{18}, revealing the sequence TGYIYHSDIVQSLPPDLR. Thus, in combination with the measured precursor mass and the observed losses of adenine and adenosine in the PSD spectrum, we iden-tified this precursor as a cross-link between protein 61K (positions 266–288, SVLPHTGYIYHSDIVQ SLPPDLR, $M_{cal} = 2606.350$) and an AU dinucleotide (presumably A43 and U44, $M_{cal} = 653.087$). The missing y_{19} (H) indicates that His270 is the actual cross-linked amino acid (Fig. 6B). Again, the losses of adenine and adenosine demonstrate that the peptide is cross-linked via the uracil rather than the adenine base. PSD analysis of the other precursor at m/z 3535.489 showed exactly the same y-type fragmentation for the peptide sequence (up to y_{18}) and w-, y_r-, and a-type fragmentation for the cross-linked RNA. Thus, the larger precursor mass is the same peptide–RNA cross-link with three additional amino acids at its N terminus

FIGURE 4. Sequence analysis of a peptide–RNA cross-link derived from UV-irradiated [15.5K–61K–U4atac] RNPs after hydrolysis with chymotrypsin and RNase T1. (*A*) PSD spectrum of a cross-linked precursor (m/z 2780.818) with DHB as matrix. The spectrum reveals significant sequence information about the cross-linked RNA moiety, but shows no fragmentation of the peptide. The measured fragment ions are related to the RNA sequence in the table in Figure 2B according to the nomenclature introduced by McLuckey et al. (1992), revealing the RNA sequence 5′-CAUAG-3′ as previously calculated. (*B*) Fragment ions found in the PSD spectrum of the precursor at m/z 2780.818 (see *A*). Asterisks next to fragments indicate that the measured fragment masses represent RNA fragments with cross-linked peptide moiety. Y-type RNA fragment ion series are named y_{r1} to y_{rn}. Peaks g, h, and i allow several combinations of fragments and sequences, as indicated by the braces. (*C*) Schematic representation of the cross-link with the sequence of the U4atac snRNA and of the U4/U6 snRNP-specific 61K protein (hPrp31p). The fragmentation types of the RNA are indicated. Since no sequence information about the cross-linked peptide part was obtained under these conditions, the calculated peptide sequence is shown in brackets.
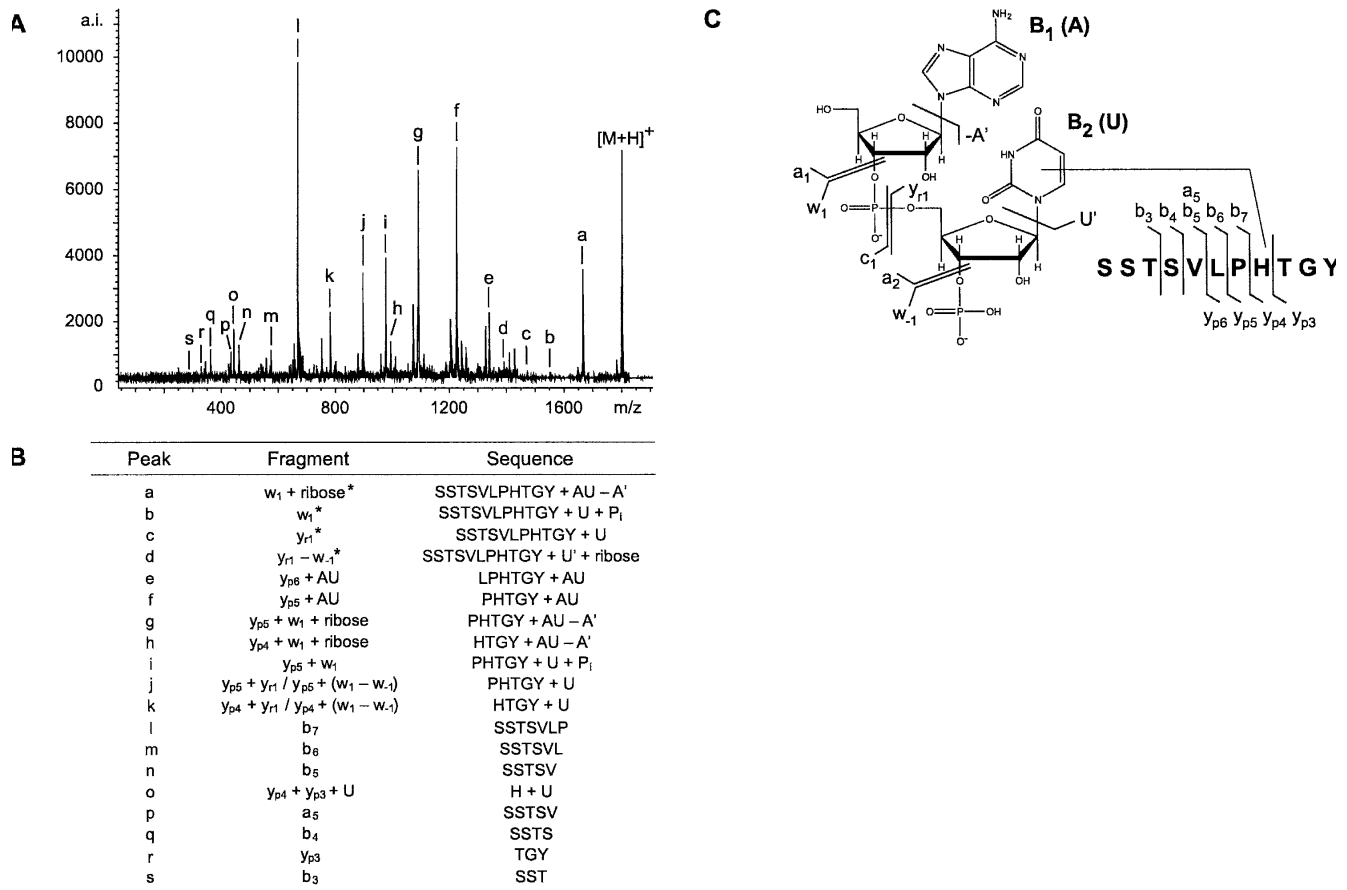
FIGURE 5. Sequence analysis of a peptide–RNA cross-link derived from UV-irradiated [15.5K–61K–U4atac] RNPs after hydrolysis with chymotrypsin and RNases A/T1. (*A*) PSD spectrum of a cross-link with a precursor of *m/z* 1801.581. The spectrum was recorded with DHB as matrix. It shows a b-type fragment ion series of the cross-linked peptide from b_3 to b_7, revealing the peptide sequence TSVLP and RNA fragments of w and y_r types, along with several combinations of these RNA fragments with peptide fragments of the y_p series; altogether this suggests that histidine 270 in the 61K protein is the amino acid actually cross-linked to nucleotide U44 in the U4atac RNA. (*B*) Fragment ions found in the PSD spectrum of the precursor at *m/z* 1801.584 (see *A*). Fragmentation nomenclature is according to McLuckey et al. (1992) except that fragmentation at the 3′-terminal phosphate is denoted as $-w_1$. Asterisks *next* to fragments indicate that the measured fragment masses represent RNA fragments with cross-linked peptide moiety. (*C*) Schematic representation of the cross-link with the sequences of the U4atac snRNA and of the U4/U6 snRNP-specific 61K protein (hPrp31p). The fragmentation types of the RNA are indicated. Since we observe Y-type fragmentation of both peptide and RNA, we refer to these as y_p and y_r, respectively.

($_{263}$SSTSVLPHTGYIYHSDIVQSLPPDLR$_{288}$, M_{cal} = 2881.461). Remarkably, the PSD analysis identified both peptides as nonspecific tryptic fragments of the 61K protein (Fig. 6B). Moreover, the fact that the AU dinucleotide (with a cross-linked U) results from an RNase A/T1 cleavage is also surprising; it may indicate that the covalently modified U is a substrate for RNases. We can exclude any nuclease activity derived from trypsin, as we exclusively isolated 61K fragments cross-linked to the RNA pentanucleotide (but no 61K fragments cross-linked to an AU dinucleotide) when UV-irradiated particles were treated with trypsin and RNase T1 only (data not shown). Such unspecific enzymatic cleavage and possible mass-spectrometric gas-phase fragmentation events must be considered in database searches when, prior to sequence analysis, putative cross-links are calculated by comparing endoproteinase fragments of the proteins with nuclease fragments of the cognate RNAs.

The results from both the MS and the PSD analyses of the cross-links derived from the [15.5K–61K–U4atac] trimeric complex reconstituted in vitro are absolutely consistent with the results from our former studies (Urlaub et al. 2000, 2002). By combining N-terminal sequencing and linear mode MALDI mass spectrometry, we showed that the U4/U6 snRNP-specific 61K protein (hPrp31p) cross-links to the 5′ stem–loop of U4atac snRNA between nucleotide positions C42 and G46. The cross-linked peptide region was identified by N-terminal sequencing, and His270 was found to be the actual cross-linked amino acid. Our studies presented here extend the earlier findings. We demonstrate that, without the need of alternative/additional sequencing methods, we are able to obtain sequence and structural information about the cross-linked peptide and RNA moieties and thus to identify unambiguously the actual cross-linking site on the peptide, His270, and on the RNA, U44.
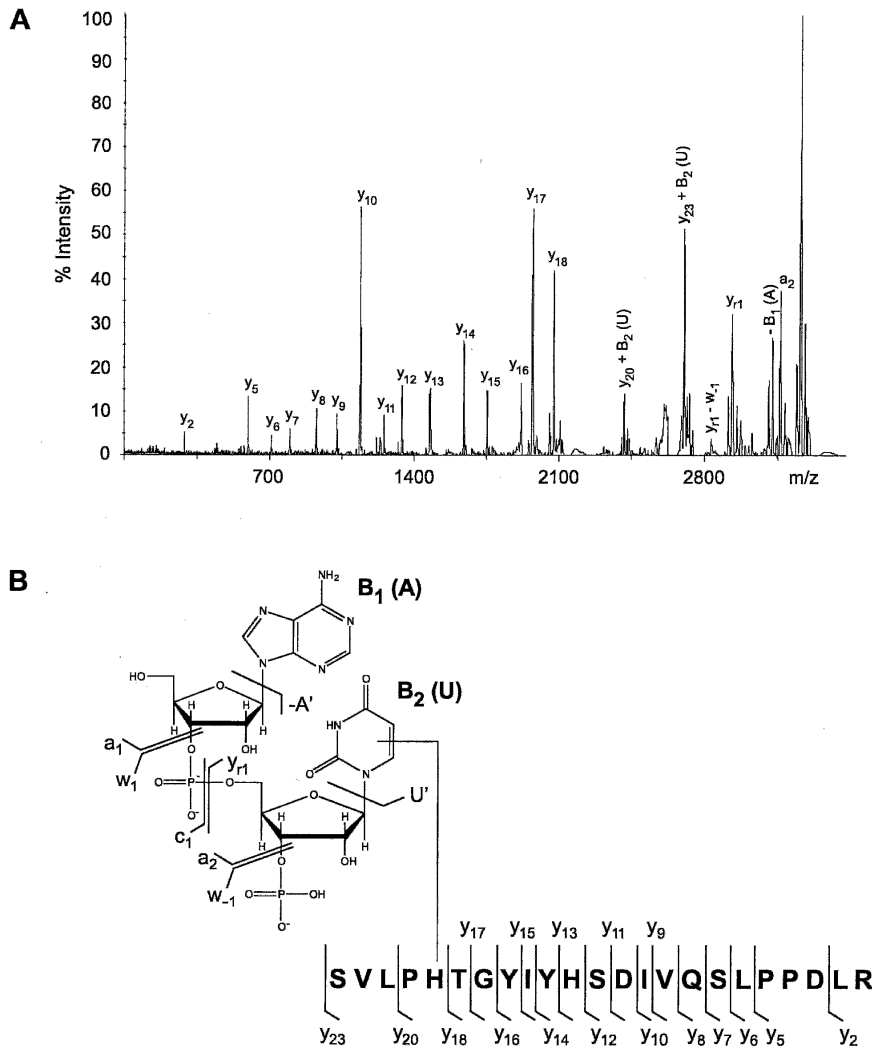
FIGURE 6. Sequence determination of a peptide–RNA cross-link derived from UV-irradiated [15.5K–61K–U4atac] RNPs after hydrolysis with trypsin and RNases A/T1. (*A*) PSD spectrum of a cross-link precursor (*m/z* 3260.444). The spectrum was recorded with THAP as matrix. The PSD spectrum shows an almost complete series of y-type fragments up to y_{18}, revealing the sequence TGYIYHSDIVQSLPPDLR. In combination with the observed losses of adenine and adenosine in the spectrum, PSD analysis identified this precursor as nontryptic peptide from protein 61K (SVLPHTGYIYHSDIVQSLPPDLR) cross-linked to an AU dinucleotide. The missing y_{19} (H) indicates that His270 is the actual cross-linked amino acid; the loss of adenine and adenosine suggests that the U rather than the A of the dinucleotide is the actual cross-linked nucleotide. The fragment ions found are directly assigned in the spectrum. (*B*) Schematic representation of both the U4atac snRNA nucleotides and the tryptic peptide of the U4/U6 snRNP-specific 61K protein (hPrp31p). Observed fragmentations within the peptide and the RNA parts are indicated. For details, see the legend of Figure 5.

Another example of the PSD analysis of a putatively cross-linked precursor is given in Figure 7A. In our MS studies shown in Figure 3C, the precursor at *m/z* 1638.666 marked as signal e was assigned to contain a tryptic U1–70K fragment encompassing positions 173–180 cross-linked to an AU dinucleotide, namely A29–U30. The PSD analysis indeed confirmed our calculation for this cross-link, as we observed several fragment ions and immonium ions of the expected peptide sequence. Despite the fact that the PSD spectrum is not of optimal quality, it provides enough

sequence information (fragment ions y_{p1}–y_{p5}, y_{p7} plus RNA) to prove that the calculated peptide sequence is indeed RVDVLVER. Consequently, the cross-linked AU dinucleotide must be A29–U30 according to Figure 3C, with the U30 cross-linked, as revealed by the loss of adenine and adenosine in the PSD spectrum. The lack of the y_{p6}-fragment ion together with the appearance of y_{p7} plus RNA—although the latter is very weak—points to L175 as the amino acid actually cross-linked. Figure 7C summarizes the results schematically. This analysis demonstrates that a sensitive and accurate MALDI-ToF analysis is a prerequisite for the identification of the cross-linked species and their sequences by PSD analysis.

SUMMARY AND CONCLUSIONS

In this study we have established that it is possible to identify the complete structure of purified peptide–RNA cross-links solely by MALDI-ToF mass spectrometry.

The analysis is based on a four-step strategy starting with the purification of peptide–RNA oligonucleotide cross-links from in vitro reconstituted and/or native ribonucleoprotein particles after UV irradiation according to our previous studies (see Fig. 1; Urlaub et al. 2000, 2001, 2002). Purified cross-links are then analyzed by MALDI-ToF in the reflectron mode under improved conditions. The accurate intact mass obtained from this analysis is used for a database search to reveal a list of putative cross-links, and finally PSD sequence analysis is applied to confirm the structure of the cross-link, including its peptide and RNA parts and the actual cross-linking site.

In this manner we examined the cross-linked peptide and RNA moieties in peptide–RNA oligonucleotides derived from UV-irradiated [15.5K–61K–U4atac] RNP complexes prepared by reconstitution in vitro and native U1 snRNP complexes and identified the actual cross-linking sites, namely His270 in the U4/U6 snRNP-specific 61K protein (hPrp31) cross-linked to U44 in the U4atac snRNA and Leu175 in the U1 snRNP specific 70K protein to U30 in the U1 snRNA. The identified cross-linking sites have been independently established in former studies by different analysis techniques (see above) and, in the
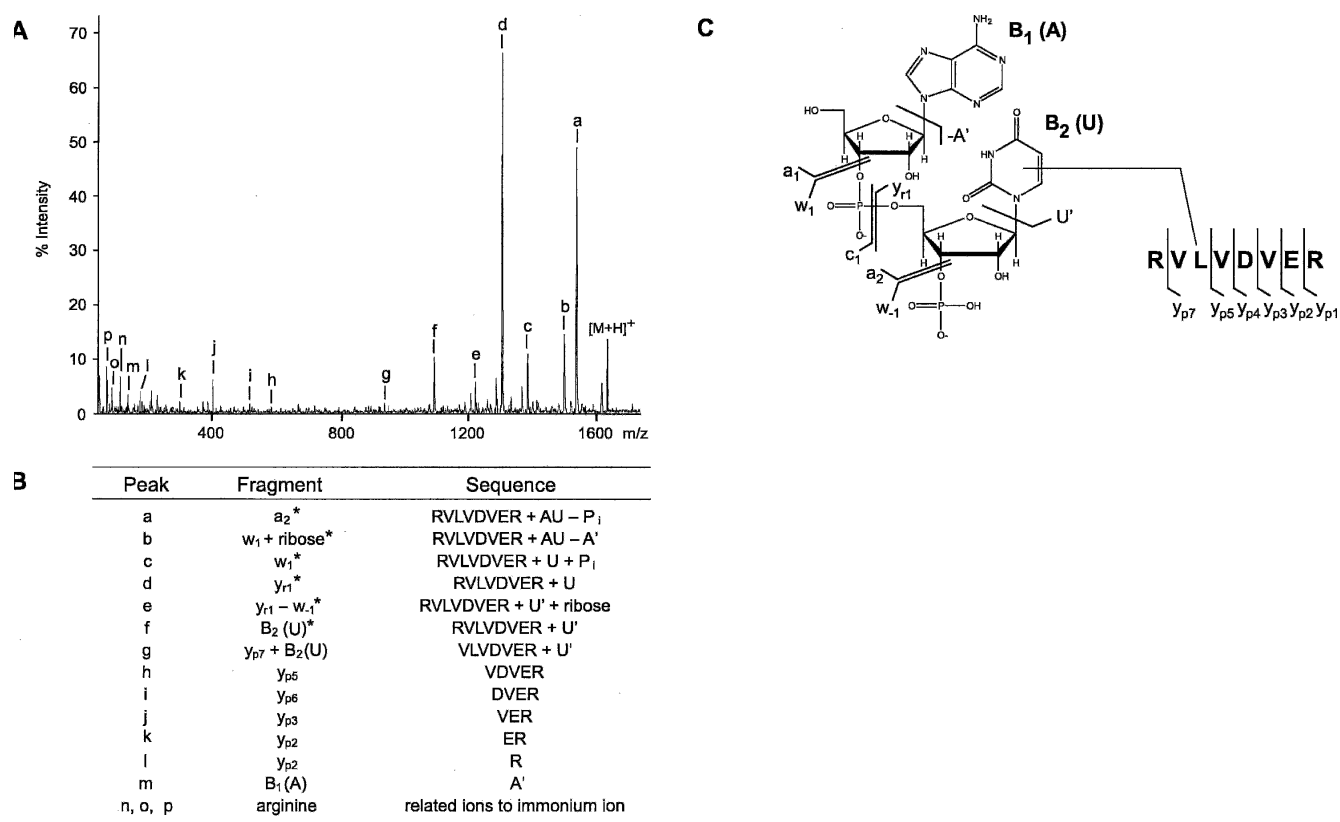
A

B

Peak	Fragment	Sequence
a	$a_2{}^*$	RVLVDVER + AU − P_i
b	w_1 + ribose*	RVLVDVER + AU − A'
c	$w_1{}^*$	RVLVDVER + U + P_i
d	$y_{r1}{}^*$	RVLVDVER + U
e	$y_{r1} - w_{-1}{}^*$	RVLVDVER + U' + ribose
f	B_2 (U)*	RVLVDVER + U'
g	$y_{p7} + B_2$(U)	VLVDVER + U'
h	y_{p5}	VDVER
i	y_{p6}	DVER
j	y_{p3}	VER
k	y_{p2}	ER
l	y_{p2}	R
m	B_1 (A)	A'
n, o, p	arginine	related ions to immonium ion

FIGURE 7. PSD spectrum of a cross-link (precursor m/z 1638.666) derived from UV-irradiated U1 snRNPs treated with endoproteinase trypsin and RNases A and T1. (*A*) The PSD spectrum shows a series of y_p-type ions from the cross-linked peptide (y_{p1}–y_{p5}), revealing the sequence VDVER. In combination with Figure 3C and the observed losses of adenine and adenosine, we identified this precursor as a cross-link of a tryptic U1–70K fragment encompassing positions 173–180 cross-linked to an AU dinucleotide, encompassing positions A29–U30 on the U1 snRNA. The measured fragment ions are related to the peptide and the RNA sequence in panel *B*. The spectrum was recorded on a Bruker Ultraflex MALDI ToF/ToF under standard PSD conditions. (*B*) Fragment ions found in the PSD spectrum of the precursor of m/z 1638.666 (see above). Asterisks *beside* fragment ions indicate RNA fragments cross-linked to peptide. For details, see the text and the legend of Figure 5. (*C*) Schematic representation of the RNA and peptide sequences of the identified cross-link. Observed fragmentation in the PSD analysis is indicated.

case of the U1 70K protein, a model of the protein with its cross-linking sites to the U1 snRNA has been proposed (Urlaub et al. 2000).

The human U4/U6 snRNP-specific 61K protein (hPrp31p) belongs to the Nop family of proteins, which also includes the box C/D snoRNP proteins Nop56 and Nop58 (Tran et al. 2004) and the archaeal homolog of these proteins, Nop5p. Recently, the crystal structure of the archaeal protein Nop5p in complex with fibrillarin was determined (Aittaleb et al. 2003; Fig. 8A). The C-terminal part of Nop5p constitutes the most strongly conserved region of the protein and is part of the so-called Nop domain (Makarova et al. 2002). Owing to accumulation of positively charged residues on the surface of this domain, the authors speculated that this region might be involved in direct interaction with the cognate archaeal box C/D RNA. Strikingly, the cross-linked His270 of the U4/U6 snRNP-specific 61K protein (hPrp31p) matches residue Lys209 of Nop5p, which is located within this conserved region of the protein (Fig. 8B). Residue Lys209 of Nop5p is the last amino acid in a disordered loop within the Nop

domain. Its approximate location can be inferred from the position of the following residue, which is ordered in the crystal structure (Fig. 8A). This observation is notably consistent with the general trend that cross-linkable regions of proteins are often located in loop regions (Urlaub et al. 1995, 2000). Such loop regions might have enough flexibility to allow the formation of a novel covalent bond between the side chain of an amino acid and the base of a nucleotide on the RNA in the vicinity.

We consider the entirely MS-based analysis presented here to be of great importance not only for the identification of putative RNA binding regions in proteins (as described above), but also for future studies of hitherto noncharacterized and low abundant protein–RNA complexes directly isolated from cells for two reasons: First, due to the improvements in sensitivity by a factor of 50–100 so that 100 fmol of cross-linked peptide–RNA oligonucleotide on the MALDI target is sufficient, it is obvious that the starting amount for such an analysis can be drastically reduced accordingly and/or that samples with a lower cross-linking yield can
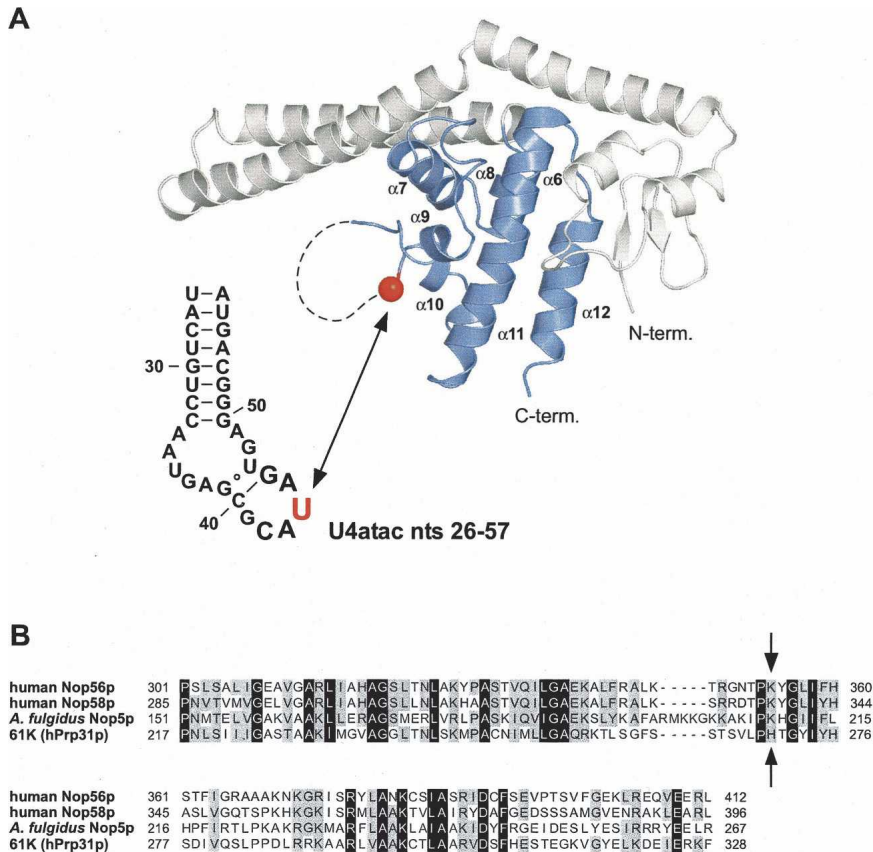
FIGURE 8. The cross-linked amino acid His270 of the U4/U6 snRNP-specific 61K protein (hPrp31p) matches residue Lys209 in the central Nop56/Nop58 homology domain of the protein Nop5p from *Archaeoglobus fulgidus,* the archaeal homolog of the human box C/D snoRNP proteins Nop56p/Nop58p and the U4/U6 snRNP-specific 61K protein (hPrp31p). (*A*) Schematic representation of the cross-linking site between U44 (shown in red) in the 5′ stem–loop of the human U4atac snRNA and the disordered loop within the Nop domain of the crystal structure of the Nop5p protein of *A. fulgidus.* Amino acid Lys209 (marked with a red ball) in the crystal structure of the archaeal Nop5p (Aittaleb et al. 2003) corresponds in the sequence alignment to His270 of the human 61K protein, which is found to be cross-linked to U4atac snRNA (see *B*). The parts of the crystal structure highlighted in dark blue represent the C-terminal part of the Nop56p/Nop58p homology domain of the Nop5p protein from α-helices α6 to α12. See the text for details. (*B*) Alignment of the C-terminal part of the Nop domain of the human U4/U6-snRNP-specific 61K protein (hPrp31p, residues 217–328; Makarova et al. 2002, acc. no. NP_056444) with the homologous sequences of the human C/D box snoRNP proteins Nop56 (amino acids 301–412, acc. no. Y12065), Nop58p (amino acids 285–396, acc. no. AF123534), and of the archaeal homolog of human Nop56p/Nop58p, Nop5p, (amino acids 151–267; Aittaleb et al. 2003; acc. no. NP_070912). Amino acids marked by an arrow represent the position of the residue highlighted in the crystal structure (*A. fulgidus* Nop5p, Lys209; see *A*) and the corresponding cross-linked amino acid His270 in the 61K protein (hPrp31p). Sequence alignments were performed by using the Clustal method. Identical residues are boxed in black; conserved residues are highlighted in gray.

be investigated. Second, establishing the MS-based sequencing obviates the need of additional experiments to elucidate sequence and structure of the cross-linked components.

The first achievement gained in this analysis is the sensitive and highly accurate analysis of cross-links in the reflectron mode by MALDI-ToF mass spectrometry with THAP or DHB as matrices. Monoisotopic resolution can be achieved up to 6000 *m/z*, enabling mass accuracies of >50 ppm even with external calibration. At this degree of accuracy the

cross-linked species can often be assigned to a certain protein in a restricted database search simply by combining monoisotopic masses of peptide fragments with monoisotopic masses of RNA fragments. Even if no unambiguous assignment can be made based on intact mass alone, the achieved accuracy might already be sufficient to narrow down the analysis significantly. It is obvious that the number of false positives will steadily increase with the number of proteins in larger RNP complexes (e.g., the spliceosomal 17S U2 or 25S [U4/U6.U5] snRNP), thus increasing the importance of reducing ambiguity of the assignment to an absolute minimum. In cases where either no hit (e.g., due to unspecific cleavage of the peptide and/or RNA) or too many hits in a database search are obtained, sequence information of either the cross-linked RNA or, more importantly, the cross-linked peptide species is an absolute requirement for the unambiguous assignment of the cross-links.

Therefore, the second improvement in our approach is the PSD analysis of cross-links, which reveals sequence information about the cross-linked peptide and RNA moieties and, ultimately, the actual site of cross-linking. Sequencing of the cross-linked peptide component has always been the bottleneck in previous analyses (Urlaub et al. 2000, 2001, 2002), mainly because too much sample was required for, for example, automated N-terminal sequencing, and the very different physico-chemical properties of the various cross-linked species made it difficult to find appropriate MALDI MS conditions for sequence analysis. Here, PSD analysis with THAP or DHB as matrices has the main advantage that when peptide–RNA heteroconjugates with cross-linked di- or trinucleotides (i.e., relatively short RNA fragments) are being studied, sequence information about both the peptide and the RNA moieties can be obtained in a single experiment under standard conditions.

In a recent study, Geyer et al. (2004) sequenced the peptide part of a peptide–ssDNA cross-link by PSD, but in this particular case the DNA moiety was completely hydrolyzed by hydrofluoric acid, so that the peptide carried only the base, and DNA sequence information was lost.

We further did not test 3-hydroxypicolinic acid (3-HPA) as MALDI matrix for our purpose. 3-HPA is suitable for MALDI analysis of nucleotides (Wu et al. 1994; Vallone et al. 2004), but shows poor ionization response for peptide analytes. Furthermore, in a similar MALDI analysis of post-translational modifications in RNase A and/or T1 fragments of prokaryotic ribosomal 5S rRNA (Kirpekar et al. 2000), the authors observed that 3-HPA was useful only when higher *m/z* ranges were considered, while only THAP showed good properties in the lower and middle *m/z* ranges.

Despite these achievements in the sequencing of the peptide–RNA oligonucleotide cross-links and the overall advantages of MALDI MS—sensitivity, mass accuracy, sample preparation, and ease of instrument handling—some critical points also have to be taken into account. Sequence information about the cross-linked peptide is only obtained from PSD analysis when the cross-linked RNA moiety is relatively small, that is, after digestion with RNases A and T1. In this case, fragmentation of the peptide backbone dominates the PSD spectrum.

The sequencing of cross-links with larger RNA fragments revealed the RNA sequence, but no information about the peptide as in similar investigations of cross-linked peptide–ssDNA conjugates by ESI MS/MS. Under standard conditions information mainly about the cross-linked di- or trinucleotide ssDNA oligomers was obtained, while the parts of cross-linked peptide sequence were obtained only after drastic changes in the collision energy (Steen et al. 2001). Unfortunately, PSD analyses do not offer any instrumental parameters to adjust the fragmentation conditions accordingly. High-energy collision-induced dissociation experiments on MALDI-ToF/ToF instruments might offer improved MS/MS capabilities for this purpose.

Since RNAs often contain repetitive di- and trinucleotide (and even longer) sequence stretches, it is difficult to map the exact cross-linked region on the RNA by only one experiment, when no cross-linkable site-specific label on the RNA has been introduced. However, when one is investigating native RNP particles—for example, with a view to discovering which proteins are in contact with the RNA at all or in the search for new RNA binding motifs in proteins—the cross-linked RNA fragment does not need to be long. In studies of this kind, our method—coupled with PSD—is the method of choice, and we plan to exploit its potential in future studies of RNAs with dynamic protein binding patterns and of low-abundance RNPs.

With the establishment of PSD analysis, the sequencing of low-abundance cross-links and cross-links derived from low-abundance protein–RNA complexes is possible. Further improvements in UV cross-linking and in purification of the cross-linked samples will be necessary in order to exploit this method to the full.

MATERIALS AND METHODS

Chemicals and reagents

Water, acetonitrile, ethanol, sodium chloride, magnesium chloride, ethylenediaminetetraacetic acid disodium salt (EDTA), urea, and sodium dodecylsulfate (SDS) were obtained from Merck. Trifluoroacetic acid (TFA), CHCA, and 2′,4′,6′-THAP were purchased from Fluka, 2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethane sulfonic acid (HEPES) from Calbiochem, 2,5-dihydroxy benzoic acid (DHB) from Sigma-Aldrich, RNases A and T1 from Ambion, chymotrypsin from Roche, and porcine trypsin from Promega. CHCA was recrystallized from dry ethanol. Peptide standard calibration mixture was obtained from Bruker.

Sample preparation

Samples containing cross-linked peptide–RNA oligonucleotide heteroconjugates were purified according to Urlaub et al. (2000, 2001, 2002). Briefly, spliceosomal ternary [15.5K–61K–U4atac] RNP complexes obtained by reconstitution in vitro (Nottrott et al. 2002) or native spliceosomal U1 snRNP particles were UV-irradiated at 254 nm. The protein moiety was digested with endoproteinases trypsin and/or chymotrypsin. RNA cross-linked to peptides together with intact RNA was separated from the excess of non-cross-linked peptides by size-exclusion chromatography on an HR75 column coupled to a SMART system (Amersham Pharmacia Biotech/GE Healthcare). Fractions containing RNA were digested with ribonucleases T1 and/or A. Peptide–oligoribonucleotide species were purified from the mixture by RP-HPLC with a C2/C18 column (100 mm × 2.1 mm; GraceVydac) coupled to a SMART system or with an RP C18 column (150 mm × 0.3 mm; MicroTech Scientific) coupled to a 140C microgradient system (Applied Biosystems) running at a flow rate of 2 µL/min. Gradients of water/0.1% TFA (solvent A) and acetonitrile/0.1% TFA (solvent B) or water/0.1% TFA (solvent A) and 80% acetonitrile/0.085% TFA, respectively, were used in the HPLC systems. Fractions that showed an absorbance at 220 and 260 nm were presumed to contain cross-linked peptide–RNA heteroconjugates and thus subjected to further analysis by MALDI MS.

Mass spectrometry

MALDI-ToF measurements were performed with a Voyager DE-STR instrument (Applied Biosystems) equipped with a pulsed nitrogen laser (20 Hz, 337 nm) and with a Reflex IV instrument (Bruker Daltonics) equipped with a pulsed nitrogen laser (5 Hz, 337 nm). For the Voyager DE-STR, the following conditions were used: acceleration voltage 20 kV, grid voltage 92% (linear)/64% (reflectron)/65% (PSD), and delay time 350 ns (linear)/250 ns (reflectron)/100 ns (PSD). A total of 300 laser shots were summed for each spectrum or PSD segment. For the Reflex IV, the acceleration voltage was 20 kV for the IS1 and 16.9 kV for the IS2, and the delay time was 400 ns. A total of 300 laser shots (5 Hz, 337 nm) were summed. PSD spectra were recorded with an acceleration voltage of 25 kV (IS1) and 20.2 kV (IS2). The delay time was 200 ns, and a total of 210 laser shots were summed for each PSD segment. The PSD spectrum in Figure 7A was recorded on an Ultraflex ToF/ToF I instrument with PAN upgrade (Bruker Dal-

tonics) equipped with a pulsed nitrogen laser (20 Hz, 337 nm). The following PSD conditions were used: acceleration voltage 8.0 kV for IS1, 7.1 kV for IS2, reflector voltage 29 kV for R1, 14.5 kV for R2, LIFT voltage 19 kV for LIFT1, 2.0 kV for LIFT2, and a total of 1500 shots were summed.

For the measurements on the Voyager DE-STR, 0.5 μL of sample was mixed on a stainless steel sample plate with the same volume of CHCA (5 mg/mL in 50% ACN/0.1% TFA), DHB (10 mg/mL in 50% ACN/0.1% TFA), or THAP (10 mg/mL in 50% ACN/0.5% TFA) matrix solution. The preparation was allowed to dry at room temperature and subjected to MS. For the MALDI-ToF measurements on the Reflex IV, 0.5 μL of the sample was mixed on a Bruker Anchor600 or Anchor400 sample plate with the same volume of DHB (10 mg/mL in 50% ACN or water, respectively) solution. The preparation was air dried as before and subjected to MS. Close external calibration with peptide standards was performed on all instruments.

Database search and computational analysis of peptide–RNA oligonucleotide cross-links

To calculate all theoretically possible combinations of RNA and peptide fragments that match the experimentally determined mass of a cross-linked peptide–RNA oligonucleotide, an exhaustive search algorithm over all possible specific and unspecific fragments of given protein and RNA sequences in the database was implemented within the Visual Fortran environment (Hewlett-Packard). The algorithm allows the following user-defined settings to specify and constrain the search: (1) mass deviation, (2) the endoproteinases used, (3) the endonucleases used, (4) the maximum number of missed cleavages on protein and RNA sequences, (5) sequence tags of the putative cross-linked peptide and RNA moiety, and (6) UV-cross-linkable amino acids (C, F, H, K, L, M, W, Y) based on Saito et al. (1983 a,b), Saito and Sugiyama (1990), Williams and Konigsberg (1991), and Meisenheimer et al. (1996, 2000). The two latter constraints greatly reduce the list of theoretically possible cross-links in cases where there are ambiguous results. Furthermore, the algorithm also takes into account unspecific cleavages and strand breaks at the N and C termini of the protein and the 5′ and 3′ ends of the RNA. Therefore, the output list contains, at most, 16 subgroups of putatively matching cross-links, among which specific cleavage of the protein and RNA moieties has the highest priority.

REFERENCES

Aittaleb, M., Rashid, R., Chen, Q., Palmer, J.R., Daniels, C.J., and Li, H. 2003. Structure and function of archaeal box C/D sRNP core proteins. *Nat. Struct. Biol.* **10:** 256–263.

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., and Brunak, S. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4:** 1633–1649.

Bonnal, S., Pileur, F., Orsini, C., Parker, F., Pujol, F., Prats, A.C., and Vagner, S. 2005. Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site *trans*-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA. *J. Biol. Chem.* **280:** 4144–4153.

Burge, C.B., Tuschl, T., and Sharp, P.A. 1999. Splicing of precursors to mRNA by the spliceosome. In *The RNA world* (eds. R.F. Gestel et al.) pp. 525–560. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Coman, M.M., Jin, M., Ceapa, R., Finkelstein, J. O'Donnell, M., Chait, B.T., and Hingorani, M.M. 2004. Dual functions, clamp opening and primer-template recognition define a key clamp loader subunit. *J. Mol. Biol.* **342:** 1457–1469.

Dix, I., Russell, C.S., O'Keefe, R.T., Newaman, A.J., and Beggs, J.D. 1998. Protein–RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *RNA* **4:** 1239–1250.

Doneau, C.E., Gafken, P.R., Bennett, S.E., and Barofsky, D.F. 2004. Mass spectrometry of UV-cross-linked protein-nucleic acid complexes: Identification of amino acid residues in the single-stranded dna-binding domain of human replication protein A. *Anal. Chem.* **76:** 5667–5676.

Expert-Bezancon, A., Sureau, A., Durosay, P., Salesse, R., Groeneveld, H., Lecaer, J.P., and Marie, J. 2004. hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of β-tropomyosin exon 6B. *J. Biol. Chem.* **279:** 38249–38259.

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.

Geyer, H., Geyer, R., and Pingoud, V. 2004. A novel strategy for the identification of protein–DNA contacts by photocrosslinking and mass spectrometry. *Nucleic Acids Res.* **32:** e132.

Golden, M.C., Resing, K.A., Collins, B.D., Willis, M.C., and Koch, T.H. 1999. Mass spectral characterization of a protein–nucleic acid photocrosslink. *Protein Sci.* **8:** 2806–2812.

Görner, H. 1994. Photochemistry of DNA and related biomolecules: Quantum yields and consequences of photoionization. *J. Photochem. Photobiol. B.* **26:** 117–139.

Greuer, B., Thiede, B., and Brimacombe, R. 1999. The cross-link from the upstream region of mRNA to ribosomal protein S7 is located in the C-terminal peptide: Experimental verification of a prediction from modeling studies. *RNA* **5:** 1521–1525.

Hamon, S., Le Sommer, C., Mereau, A., Allo, M.R., and Hardy, S. 2004. Polypyrimidine tract-binding protein is involved in vivo in repression of a composite internal/3′-terminal exon of the *Xenopus* α-tropomyosin pre-mRNA. *J. Biol. Chem.* **279:** 22166–22175.

Henras, A.K., Dez, C., and Henry, Y. 2004. RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr. Opin. Struct. Biol.* **14:** 335–343.

Hicke, B.J., Willis, M.C., Koch, T.H., and Cech, T.R. 1994. Telomeric protein–DNA point contacts identified by photocrosslinking using 5-bromodeoxyuridine. *Biochemistry* **33:** 3364–3373.

Jensen, O.N., Barofsky, D.F., Young, M.C., von Hippel, P.H., Swenson, S., and Seifried, S.E. 1993. Direct observation of UV-cross-linked protein-nucleid acid complexes by matrix-assisted laser desorption ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **7:** 496–501.

Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60:** 2299–2301.

Kastner, B. and Lührmann, R. 1999. Purification of U small nuclear ribonucleoprotein particles. In *RNA–protein interaction protocols, methods in molecular biology* (ed. S.R. Haynes), vol. 118, pp. 289–298. Humana Press, Totowa, NY.

Kaufmann, R., Spengler, B., and Lützenkirchen, F. 1993. Mass spectrometric sequencing of linear peptides by production analysis in a reflectron

time-of-flight mass spectrometer using matrix-assisted laser desorption ionization. *Rapid Commun. Mass Spectrom.* **7:** 902–910.

Kirpekar, F., Douthwaite, S., and Roepstorff, P. 2000. Mapping post-transcriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry. *RNA* **6:** 296–306.

Konarska, M.M. 1999. Site-specific derivatization of RNA with photo-crosslinkable groups. *Methods* **18:** 22–28.

Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P., and Lührmann, R. 2002. Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6·U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J.* **21:** 1148–1157.

McLuckey, S.A., Van Berkel, G.J., and Glish, G.L. 1992. Tandem mass-spectrometry of small multiply charged oligonucleotides. *J. Am. Soc. Mass Spectrom.* **3:** 60–70.

Meisenheimer, K.M. and Koch, T.H. 1997. Photocross-linking of nucleic acids associated to proteins. *Crit. Rev. Biochem. Mol. Biol.* **32:** 101–140.

Meisenheimer, K.M., Meisenheimer, P.L., Willis, M.C., and Koch, T.H. 1996. High yield photocrosslinking of a 5-iodocytidine (IC) substituted RNA to its associated proteins. *Nucleic Acids Res.* **24:** 981–982.

Meisenheimer, K.M., Meisenheimer, P.L., and Koch, T.H. 2000. Nucleoprotein photo-cross-linking using halopyrimidine-substituted RNAs. *Methods Enzymol.* **318:** 88–104.

Möller, K. and Brimacombe, R. 1975. Specific cross-linking of proteins S7 and L4 to ribosomal RNA by UV irradiation of *Escherichia coli* ribosomal subunits. *Mol. Gen. Genet.* **141:** 343–355.

Norris, C.L., Meisenheimer, P., and Koch, T.H. 1996. Mechanistic studies of the 5-iodouracil chromophore relevant to its use in nucleo-protein photo-cross-linking. *J. Am. Chem. Soc.* **118:** 5796–5803.

———. 1997. Mechanistic studies relevant to bromouridine enhanced nucleoprotein photocrosslinking. Possible involvement of an excited tyrosine residue of the protein. *Photochem. Photobiol.* **65:** 201–207.

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.

Reed, R. and Chiara, D. 1999. Identification of RNA–protein contacts within functional ribonucleoprotein particles by RNA site-specific labeling and UV crosslinking. *Methods* **18:** 3–12.

Rhode, B.M., Hartmuth, K., Urlaub, H., and Lührmann, R. 2003. Analysis of site-specific protein–RNA cross-links in isolated RNP complexes, combining affinity selection and mass spectrometry. *RNA* **9:** 1542–1551.

Saito, I. and Sugiyama, H. 1990. Photoreactions of nucleic acids and their constituents with amino acids and related compounds. In *Bioorganic photochemistry* (ed. H. Morrison), pp. 317–340. John Wiley and Sons, New York.

Saito, I., Sugiyama, H., and Matsuura, T. 1983a. Isolation and characterization of a thymine–lysine adduct in UV-irradiated nuclei. The role of thymine–lysine photoaddition in photo-cross-linking of proteins to DNA. *J. Am. Chem. Soc.* **105:** 6989–6991.

———. 1983b. Photochemical reactions of nucleic acids and their constituents with amino acids and related compounds. *Photochem. Photobiol.* **38:** 735–743.

Smith, R.D., Loo, J.A., Edmonds, C.G., Barinaga, C.J., and Udseth, H.R. 1990. New developments in biochemical mass spectrometry: Electrospray ionisation. *Anal. Chem.* **62:** 882–899.

Sontheimer, E.J. 1994. Site-specific RNA crosslinking with 4-thio-uridine. *Mol. Biol. Rep.* **20:** 35–44.

Steen, H., Petersen, J., Mann, M., and Jensen, O.N. 2001. Mass spectrometry analysis of a UV-cross-linked protein–DNA complex: Tryptophans 54 and 88 of *E. coli* SSB cross-link to DNA. *Protein Sci.* **10:** 1989–2001.

Tran, E., Brown, J., and Maxwell, E.S. 2004. Evolutionary origins of the RNA-guided nucleotide-modification complexes: From the primitive translation apparatus? *Trends Biochem. Sci.* **7:** 343–350.

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.

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.

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.

Urlaub, H., Hartmuth, K., and Lührmann, R. 2002. A two-tracked approach to analyze RNA–protein crosslinking sites in native, non-labeled small nuclear ribonucleoprotein particles. *Methods* **26:** 170–181.

Vallone, P.M., Fahr, K., and Kostrzewa, M. 2004. Genotyping SNPs using a UV-photocleavable oligonucleotide in MALDI-TOF MS. *Methods Mol. Biol.* **297:** 169–178.

Will, C.L. and Lührmann, R. 1997. Protein functions in pre-mRNA splicing. *Curr. Opin. Cell. Biol.* **9:** 320–328.

———. 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13:** 290–301.

Williams, K.R. and Konigsberg, W.H. 1991. Identification of amino acid residues at the interface of protein–nucleic acid complexes by photochemical cross-linking. *Methods Enzymol.* **208:** 516–539.

Wong, D.L. and Reich, N.O. 2000. Identification of tyrosine 204 as the photo-cross-linking site on the DNA–EcoRI DNA methyltransferase complex by electrospray ionisation mass spectrometry. *Biochemistry* **39:** 15410–15417.

Wu, K.J., Shaler, T.A., and Becker, C.H. 1994. Time-of-flight mass spectrometry of underivatized single-stranded DNA oligomers by matrix-assisted laser desorption. *Anal. Chem.* **66:** 1637–1645.

Zaia, J. 2004. Mass spectrometry of oligosaccharides. *Mass Spectrom. Rev.* **23:** 161–227.

Zamfir, A. and Peter-Katalinic, J. 2004. Capillary electrophoresis-mass spectrometry for glycoscreening in biomedical research. *Electrophoresis* **25:** 1949–1963.

Zhu, Y.F., Chung, C.N., Taranenko, N.I., Allman, S.L., Martin, S.A., Haff, L., and Chen, C.H. 1996. The study of 2,3,4-trihydroxyaceto-phenone and 2,4,6-trihydroxyacetophenone as matrices for DNA detection in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **10:** 383–388.