International Journal of Mass Spectrometry 304 (2011) 184–194

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Mass-spectrometric analysis of proteins cross-linked to 4-thio-uracil- and 5-bromo-uracil-substituted RNA

Katharina Kramer[a], Petra Hummel[b], He-Hsuan Hsiao[a], Xiao Luo[c], Markus Wahl[c], Henning Urlaub[a,*]

[a] Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[b] IT & Electronics Service, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[c] AG Strukturbiochemie, Fachbereich Biologie-Chemie-Pharmazie, Institut für Chemie und Biochemie, Freie Universität Berlin, Takustrasse 6, 14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 12 July 2010
Received in revised form 6 October 2010
Accepted 9 October 2010
Available online 15 October 2010

Keywords:
Cross-linking
Protein
RNA
4-Thio-uracil
5-Bromo-uracil
Mass spectrometry

ABSTRACT

UV-induced cross-linking of protein–RNA complexes is widely used to study protein–RNA interactions. Mass spectrometry (MS) has been used successfully to characterise the cross-links obtained. The relatively low yield of the cross-linking reaction can be increased by using photoreactive base-analogues such as 4-thio-uracil and halopyrimidines such as 5-bromo-uracil. However, little is known about the reaction mechanism of 4-thio-uracil with amino acid side chains upon UV light exposure. Therefore the exact mass of the cross-linking product could not be predicted. In this study we performed a mass-spectrometric analysis of proteins that were cross-linked to RNA that contained 4-thio-uridine or 5-bromo-uridine. We observed the formation of three different species from 4-thio-uridine upon UV cross-linking, two indicating reaction pathways differing from the purely additive cross-linking of natural nucleotides. One cross-link originating from 5-bromo-uridine was identified. In parallel, we developed a software for the automated analysis of MS data derived from cross-linking experiments that makes use of commercially available search engines. Thus our study paves the way for future cross-linking experiments with 4-thio-uracil in more complex systems.

1. Introduction

UV-induced cross-linking is a widely used method for the investigation of interactions between proteins and DNA or RNA. UV-based cross-linking can utilise the natural reactivity of the nucleic acid bases after irradiation at a wavelength of 254 nm. After excitation by UV light, subsequent radical-based reactions lead to the formation of a covalent bond between the nucleic-acid base and an amino-acid residue, a so-called zero-length cross-link. Experimental studies have shown that cysteine, lysine, methionine, phenylalanine, tryptophan and tyrosine residues are the most reactive towards cross-linking; however, all of the common amino acids except proline can form cross-links [1].

Owing to the short half-life of the excited states of the nucleic-acid base, cross-links can only be formed to amino acids that are in contact or in close spatial proximity to the excited nucleotide [2]. The advantage of this type of cross-linking is that the three-dimensional structure of the protein–RNA complex remains largely undisturbed. However, base-pairing of the RNA and secondary structures of the protein can hinder cross-link formation, because the regions of the RNA or protein involved do not exhibit sufficient flexibility for the formation of the covalent bond.

While the actual cross-linked protein can be identified for example by gel electrophoresis combined with suitable detection methods, e.g., [3], mass spectrometry (MS) is the method of choice for the unbiased identification of the cross-linking sites and thus of the cross-linked protein. Accordingly, UV cross-linking combined with MS has been proven to be a powerful tool in the identification and characterisation of protein–RNA contact sites, e.g., in prokaryotic ribosomes [4], spliceosomal ribonucleoprotein complexes [5–8] or in *Escherichia coli*'s NusB–S10 complex [9].

However, there remain bottlenecks in the performance of comparable studies *in vivo* or on more complex protein–RNA particles that can only be isolated in limited amounts. Extending cross-linking/MS studies to these systems is hampered by (i) the low cross-linking yield under UV conditions, which only leads to approximately 1–5% of the protein population becoming cross-linked to RNA, and – as a consequence of this low cross-linking yield – (ii) relatively poor MSMS spectra, the quality and sensitivity of which are inadequate for the unambiguous identification of the cross-linked peptide moiety, and (iii) lack of suitable software tools that overcome the poorly resolved peptide fragmentation and allow the identification of the corresponding protein in a database search.

* Corresponding author. Tel.: +49 551 201 1060; fax: +49 551 201 1197.
E-mail address: henning.urlaub@mpi-bpc.mpg.de (H. Urlaub).

1387-3806/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2010.10.009

Thus, several attempts have been performed to improve the sensitivity of MS-based analyses for the identification of cross-linking sites in proteins. These are based mainly on enrichment strategies that make use of the acidic nature of the cross-linked phosphate backbone [8,10–12], or on scanning experiments in a triple quadrupole mass spectrometer [7,12].

Despite of the establishment of efficient enrichment strategies, the overall yield of UV cross-linking using the common nucleic acid bases still remains unsatisfactory. One strategy used to address this problem includes the incorporation of photoreactive groups into the nucleoprotein complex to enhance the yield of the cross-linking reaction. The nucleic acid bases can be replaced by chromophores such as 4-thio-uracil and halopyrimidines, e.g., 5-bromo- or 5-iodo-uracil. These exhibit absorption maxima at longer wavelengths (4-thio-uridine 330 nm, 5-bromo- and 5-iodo-uridine around 280 nm [13]). Previous studies have used 5-bromo-2′-deoxyuridine substitution to investigate protein–DNA interactions [14,15], while the cross-links can be characterised with electrospray ionisation (ESI) MS [16]. The combination of 5-iodo-2′-deoxyuridine substitution and cross-link analysis by matrix-assisted laser desorption/ionisation (MALDI) MS has also been applied successfully [11,17].

In addition to halo-substituted RNA/DNA, incorporation of 4-thio-uridine (referred to hereinafter as 4SU) into RNA seems to enhance the cross-linking yield even more strongly. For example, substitution of uridine with 4SU in RNA has allowed the identification of numerous proteins that are in direct contact with the U5 snRNA and their interaction sites within the U5 ribonucleoprotein particle of *Saccharomyces cerevisiae* [18]. Furthermore, the known ability of 4SU to be incorporated *in vivo* into the RNA of *E. coli* [19,20] or mammalian cells [21] could be used to develop a novel MS-based protocol for the identification of protein–RNA contact sites after UV irradiation of isolated ribonucleoprotein particles or even of whole cells. *In vivo* incorporation of 4SU has been used very recently in a cross-linking study based on immunoprecipitation [22], where a 100- to 1000-fold greater recovery of cross-linked RNA was observed, when cells treated with 4SU and irradiated at 365 nm were compared with untreated cells irradiated at 254 nm.

In all unsubstituted protein–RNA complexes investigated so far by MS (see above), UV irradiation at 254 nm leads mainly to heteroconjugates that are additive, i.e., the total mass of the conjugate is equal to the mass of the protein/peptide plus the mass of the cross-linked RNA oligonucleotide. However, to our knowledge no MS investigation of 4SU cross-links to proteins has been performed yet. A prerequisite for successful identification and an aspired automated database search for the cross-linked peptide moiety by MS analysis derived from modified nucleobases in RNA (and/or DNA) is a clear result from the cross-linking reaction, i.e., the structure and mass of the heteroconjugate that is generated upon UV irradiation.

To clarify the effect of substitutions within the RNA on the yield and the products of UV cross-linking experiments, we compared UV cross-linking of proteins to non-labelled RNA with similar cross-linking to RNA that contained site-specific substitutions with 5-bromo-uridine (5BrU) and 4-thio-uridine (4SU). These experiments were performed on the NusB–S10 complex from *E. coli*, which plays an important role in transcription antitermination and which exhibits enhanced affinity to BoxA-containing RNAs (see [9] and references therein). In addition, we also present a novel data-analysis strategy for the identification of the cross-linked peptide part using MASCOT [23] as search engine. Hitherto unknown cross-linking products between 4SU and amino acid side chains of the proteins were identified using automated and manual data analysis. Once the novel cross-linked 4SU derivates were identified, they could be easily adapted into our automated data analysis strategy.

2. Experimental

2.1. 4-Thio-uracil substituted RNA oligomer

The 19mer RNA oligonucleotides containing *rrn* BoxA (5′-CACUGCUCUUUAACAAUUA-3′; oligo 1a) without substitution and with uracils 9, 10 and 11 replaced by 4SU (5′-CACUGCUC(4SU)(4SU)(4SU)AACAAUUA-3′; oligo 1b), as well as the oligomer containing γ BoxA with uracil 4 replaced by 5BrU (5′-CGC(5BrU)CUUACACA-3′; oligo 2), were obtained from Dharmacon (ThermoFisher Scientific, Epsom, UK).

2.2. Protein expression and purification

For cross-linking experiments with 5BrU-substituted RNA, the NusB–S10 complex was expressed and purified exactly as described previously [9]. For experiments with 4SU-substituted RNA, the protocol was followed but without cleaving the N-terminal His_6-tag of NusB. Briefly, *E. coli* BL21(DE3) cells containing two different plasmids, encoding respectively NusB and S10, were grown in auto-inducing medium [24] in the presence of kanamycin and ampicillin to an OD_{600} of 0.5 at 37 °C and subsequently incubated at 20 °C for 20 h. Cells were harvested at 4 °C, washed with binding buffer (50 mM Tris–HCl, pH 7.5, 150 mM NaCl), flash-frozen in liquid nitrogen and stored at −80 °C.

Cells were resuspended in binding buffer containing protease inhibitors, sonicated and centrifuged at 15,000 rpm and 4 °C for 30 min. The cell lysate was incubated with glutathione-Sepharose (GE Healthcare, Munich, Germany), pre-equilibrated with binding buffer, at 4 °C for 3 h to trap the protein complex on the N-terminal GST-tag of S10. The complex was eluted with binding buffer containing 15 mM reduced glutathione. The GST-tag was cleaved with PreScission protease by incubation at 4 °C overnight. The protein complex was subsequently trapped (through the N-terminal His_6-tag of NusB) on a Ni-NTA agarose column (Qiagen, Hilden, Germany) that had been pre-equilibrated with binding buffer containing 20 mM imidazole. The column beads were washed with binding buffer containing 50 mM imidazole and the complex was eluted with binding buffer containing 500 mM imidazole. After concentration by ultrafiltration the complex was further purified by gel filtration on a Superdex 75 10/300 GL column (GE Healthcare, Munich, Germany) equilibrated with binding buffer containing 2 mM DTT. The protein solution was flash-frozen in liquid nitrogen and stored at −80 °C. Protein concentrations were determined by the Bradford assay.

2.3. UV-induced cross-linking with radiolabelled RNA oligonucleotides

The unsubstituted and 4SU-substituted oligonucleotides 1a and 1b were 5′ labelled using γ-[^{32}P]-ATP (6000 Ci/mmol) and T4 polynucleotide kinase following standard procedures. A more than 100-fold excess of 250 pmol NusB–S10 complex was mixed with [^{32}P]-labelled oligonucleotide in a reaction volume of 7.4 μL. After incubation for 30 min, the protein–RNA complex was UV-irradiated at the appropriate wavelength for 10 min (254 nm) or 1, 2, 5 and 10 min (365 nm) in a cross-linking apparatus built in-house equipped with four 8 W lamps (dimension 1.5 cm × 28.5 cm, distance 4 cm; 254 nm: F8T5BL, 365 nm: G8T5, both Sankyo Denki, Japan). The cross-linking products were analysed by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE). Bands containing radiolabelled RNA were visualised with a Typhoon 8600 Phosphorimager (GE Healthcare, Munich, Germany) and quantified by using the Quantity One software (Bio-Rad Laboratories, Munich, Germany).

2.4. UV-induced cross-linking and enrichment of cross-linked peptides for MS analysis

UV cross-linking and enrichment were performed essentially as described elsewhere [9]. 2 nmol oligo 1b and 2 nmol NusB–S10 complex in 100 μL binding buffer or 16 nmol oligo 2 and 8 nmol NusB–S10 complex in 500 μL buffer (10 mM Tris–HCl pH 7.5, 50 mM NaCl, 2 mM DTT) were incubated on ice for 30 min for complex formation. Samples were transferred to black polypropylene microplates (Greiner Bio-One, Frichenhausen, Germany), in the case of 5BrU in 100-μL aliquots. UV cross-linking was performed at 365 nm (4SU) or at 312 nm (5BrU; 8W lamps No. 298440, Herolab, Germany) for 10 min. After ethanol precipitation, the sample was digested with 1 μg RNase A and 1 U RNase T1 (both from Ambion, Applied Biosystems, Darmstadt, Germany) in the case of 4SU, or with 1.5 μg RNase A and 1.5 U RNase T1 for 5BrU. Samples were incubated with RNases for 2 h at 52 °C in the presence of 1 M urea and 50 mM Tris–HCl pH 7.9. Proteolysis with trypsin (protein:enzyme 20:1, w/w) was performed overnight at 37 °C.

The sample was desalted on a C18 (Dr. Maisch GmbH, Ammerbuch, Germany) column prepared in-house. Cross-linked peptides were enriched on a TiO_2 (GL Sciences, Tokyo, Japan) column prepared in-house. The enrichment procedure followed the protocol described in [25]: after evacuation to dryness, the sample was redissolved in 200 mg/mL 2,5-dihydroxy benzoic acid (DHB, Sigma–Aldrich, Munich, Germany) in 80% acetonitrile (ACN) containing 5% trifluoroacetic acid (TFA); it was then loaded on the TiO_2 column, washed with 200 mg/mL DHB in 80% ACN, 5% TFA, and eluted with 0.3 M ammonia. For MS analysis, the sample was dried, redissolved in 50% ACN, 0.1% formic acid (FA) and diluted with 0.1% FA.

2.5. Nano-LC separation and MS analysis

Samples were injected onto a nano-liquid chromatography system (Agilent 1100 series, Agilent Technologies, Böblingen, Germany) including a C18 trapping column (length ~2 cm, inner diameter 150 μm) in-line with a C18 analytical column (length ~15 cm, inner diameter 75 μm) packed in-house (C18 AQ 120 Å 5 μm, Dr. Maisch GmbH, Ammerbuch, Germany). Analytes were loaded on the trapping column at a flow rate of 10 μL/min in buffer A (0.1% FA, v/v) and subsequently eluted and separated on the analytical column with a gradient of 7.5–37.5% buffer B (95% acetonitrile, 0.1% FA, v/v) with an elution time of 37 min and a flow rate of 300 nL/min.

On-line ESI-MS/MS was performed with a Q-ToF (quadrupole time of flight) Ultima mass spectrometer (Waters, Manchester, UK). The instrument was operated in data-dependent acquisition mode. Survey scans were recorded for 1 s over the mass range m/z 350–1600. For precursors detected with intensities above 15 counts and a charge state of two to four, a maximum of three consecutive MS/MS scans was triggered.

2.6. Automated MS data analysis

Raw data was processed and a peak list file (.pkl) was generated by MassLynx 4.0 SP4. Data processing steps were smoothing (Savitzky Golay, smooth window (channels) ±3, number of smoothes 2) and centroiding (min peak width at half height (channels) 4, centroid top %: 80). All MSMS spectra were searched with MASCOT [23] against the entire NCBInr database to identify unmodified NusB and S10 peptides and peptides from trypsin, RNases and contaminants such as keratin. For spectra containing possible cross-links, all possible nucleotide combinations were subtracted from the precursor mass using a perl-based script, and the resulting novel peak list file was searched with MASCOT against a database comprising NusB, S10, and 12 additional proteins (see Section 3 for extended description).

Manual identification and spectra evaluation made use of several calculation programs available online, namely ProteinProspector (http://prospector.ucsf.edu) to calculate theoretical fragmentation series of peptides, and the Peptide Mass Calculator (http://rna.rega.kuleuven.ac.be/masspec/pepcalc.htm) and Mongo Oligo Mass Calculator (http://biochem.ncsu.edu/RNAmods/masspec/mongo.htm) to calculate monoisotopic masses of peptides and oligonucleotides, respectively.

3. Results

3.1. Cross-linking with 4SU

This study describes the use of photoreactive nucleobase analogues for UV-induced protein–RNA cross-linking and the MS analysis of the cross-linked heteroconjugates thus obtained. We applied this approach to the *E. coli* NusB–S10 complex bound to BoxA-containing RNA oligonucleotides [9]. We aimed to get insight how site-specifically labelled nucleobases react with amino-acid residues within a protein, and what the actual cross-linking products are.

Conditions for UV irradiation have been described earlier [12], but with the wavelength set to a longer value for substituted RNA oligonucleotides (see Section 2). We compared a non-substituted RNA oligonucleotide that contains the *rrn* BoxA (5′-CACUGCUCUUUAACAAUUA-3′; oligo 1a) with the same oligonucleotide that contained 4SU at positions 9, 10 and 11 (5′-CACUGCUC(4SU)(4SU)(4SU)AACAAUUA-3′; oligo 1b) and a γ BoxA core RNA oligonucleotide that contained 5BrU at position 4 (5′-CGC(5BrU)CUUACACA-3′; oligo 2).

As the UV cross-linking of 4SU-containing protein–RNA complexes has been shown to give a significantly enhanced cross-linking yield [18,22], we first investigated the cross-linking behaviour of the NusB–S10 complex towards unsubstituted or 4SU substituted RNA. We used ^{32}P-labelled RNA oligonucleotides and monitored the cross-linking yields of the native and 4SU-substituted *rrn* BoxA-containing RNA oligonucleotides to both proteins by SDS-PAGE.

Fig. 1 shows the protein cross-linking products after 0 min and 10 min UV irradiation of oligo 1a (lanes 1 and 2, respectively) and the cross-linking products after 0, 1, 2, 5 and 10 min in the presence of 4SU substituted oligo 1b (lanes 3, 4, 5, 6 and 7, respectively). Surprisingly, cross-linking of 4SU substituted oligo 1b only significantly enhances the cross-linking yield of protein S10, whereas the cross-linking yield of the NusB protein seems to be unaffected (compare Fig. 1, lanes 2 and 7). A quantitative analysis by phosphorimaging revealed that 4SU increases the yield of cross-linking to protein S10 by approximately 50%; however, the cross-linking yield of NusB to the 4SU-containing oligo 1b is decreased by approximately 10%. It is of note that the SDS-PAGE analysis also reveals that even under daylight conditions cross-linking between 4SU and proteins occurs, so that protection from ambient light is required when one is working with 4SU-substituted RNA.

Both of these effects – i.e., the similar cross-linking yield of NusB to non-modified and 4SU modified RNA oligonucleotide and the increase in the cross-linking yield – can be explained at least partially by the three-dimensional structure of the protein complex, as will be discussed below.

3.2. General workflow for MS analysis and database search for putatively cross-linked species

Fig. 2 outlines schematically the sample preparation for the MS analysis of UV-cross-linked peptide–RNA oligonucleotide conju-

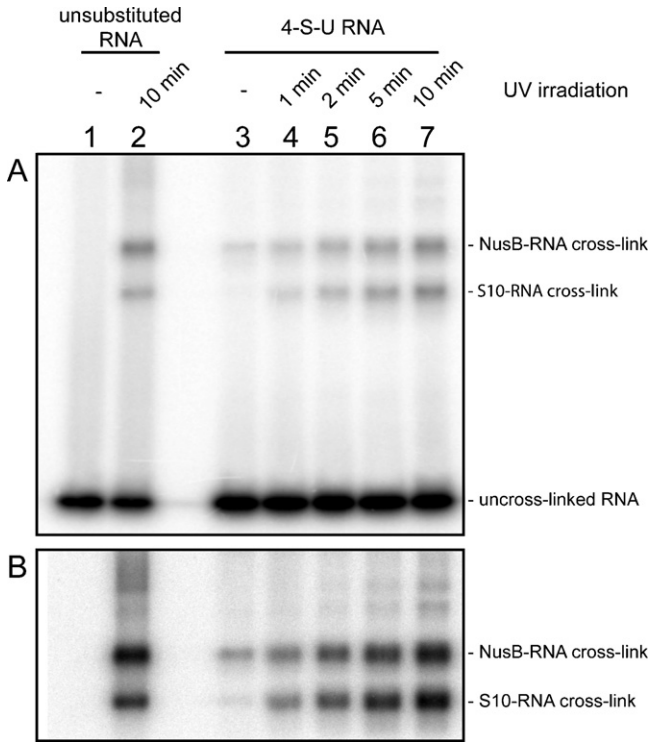

Fig. 1. Autoradiograph of cross-linking experiments with γ-[^{32}P]-ATP labelled RNA oligonucleotides. The cross-linking yield of unsubstituted (oligo 1a, lane 2) and 4SU substituted (oligo 1b, lanes 4–7) RNA oligonucleotides in cross-linking to NusB–S10-complex is compared. Lanes 1 and 3 show the controls without UV irradiation of oligos 1a and 1b, respectively. The complex with oligo 1a was cross-linked for 10 min at 254 nm (lane 2) while UV irradiation for complexes with oligo 1b was performed at 365 nm for 1, 2, 5 and 10 min (lanes 4–7). (A) Autoradiograph after 15 min exposure on Phosphorimager immediately after electrophoresis and (B) after 1 h.

gates. Briefly, UV cross-linked [NusB–S10–BoxA-containing RNA oligonucleotide] complexes are hydrolysed with ribonucleases and endoproteinase and subsequently desalted. Peptide–RNA oligonucleotides are enriched by the use of TiO_2 and analysed by LC–MSMS.

However, enriched putative peptide–RNA oligonucleotides are difficult to detect in the MS. There are several reasons for this: (i) contaminating noncross-linked RNA oligonucleotides still co-elute from the LC into the MS and are also ionised in the positive mode. These can suppress the signal intensity of putative cross-links. The same holds true for contaminating residual peptides. (ii) The abundance of the peptide–RNA conjugates is low, because of a low cross-linking yield. (iii) In connection with this, the quality of MSMS spectra of peptide–RNA oligonucleotide conjugates is still relatively poor, in particular for ion series derived from the cross-linked peptide moiety. The latter effect becomes even more problematical when fragmentation is performed under collision induced dissociation (CID) conditions in an ion-trap instrument (H.U. unpublished observations). Conversely, Q-ToF instruments deliver MSMS spectra of higher quality; however, their overall sensitivity is reduced as compared with (linear) ion-trap instruments. FT-analysers have the advantage that the measured precursor mass is highly accurate. However, when fragmentation is performed under CID conditions and analysed in the quadrupole linear ion trap (LTQ) as in ThermoFisherScientific Fourier transform ion cyclotron resonance (FT-ICR) or Orbitrap instruments (Orbitrap XL type), the MSMS spectra are again not of sufficient quality. Alternative fragmentation modes, such as higher energy collision induced dissociation (HCD) in an Orbitrap instrument [26] with subsequent fragment analysis in the FT-analyser, or the application of electron capture/transfer dissociation (ECD/ETD) seem to improve the spectrum quality [27]. However, the sensitivity of HCD fragmentation in the available Orbitrap XL instruments was not sufficient for the analysis of cross-linking experiments in our hands. (iv) There is still a lack of convenient computational approaches that ease the detection and database search of cross-linked species, in particular of the cross-linked peptide moiety.

In our previous studies we applied software tools that simply calculate all possible combinations of nucleotides and peptides (derived from a pre-defined database of proteins) on the basis of the measured precursor mass. Unfortunately, these tools first generated many putative positive hits (even under stringent search conditions, i.e., low mass deviation, data not shown) and these thus still required laborious manual confirmations. A number of algorithms have been developed for the automated analysis of chemical cross-linking, but according to [28] only Links/MS2Links [29] allow analysis of cross-links between proteins and RNA/DNA. However, for these tools the sequences of the potentially cross-linked species have to be defined, and the number of input sequences is unfortunately limited to 20. This prevents analysis of larger protein–RNA

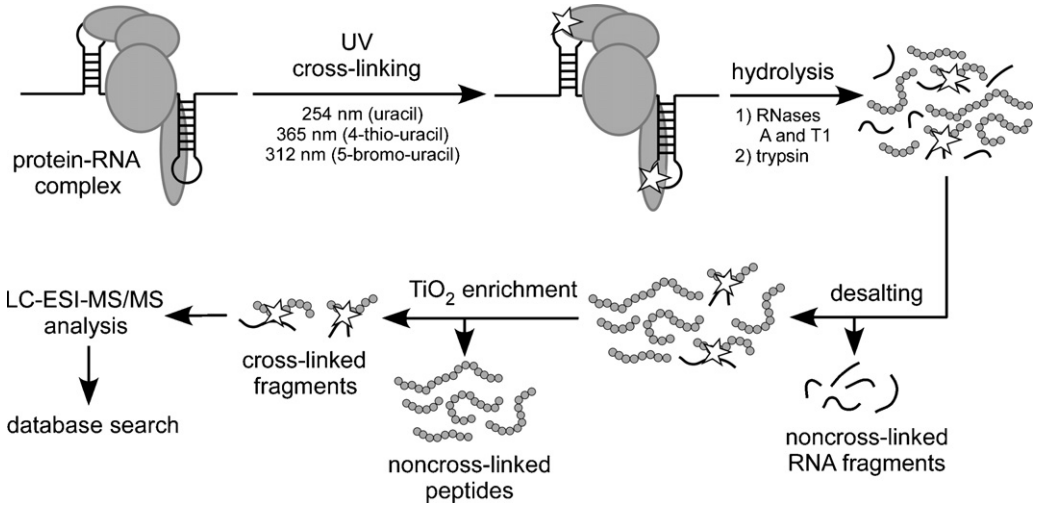

Fig. 2. Experimental scheme for UV cross-linking of a protein complex to native and 4-thio-uracil or 5-bromo-uracil substituted RNA with subsequent enrichment and LC–ESI-MSMS analysis of cross-linked fragments. Protein–RNA complexes are UV irradiated at an appropriate wavelength, hydrolysed by RNases and trypsin and desalted. Peptide–RNA oligonucleotide conjugates are then enriched via TiO_2 columns and subsequently analysed by LC–ESI-MS/MS. Finally, MS/MS data is processed and subjected to database search.

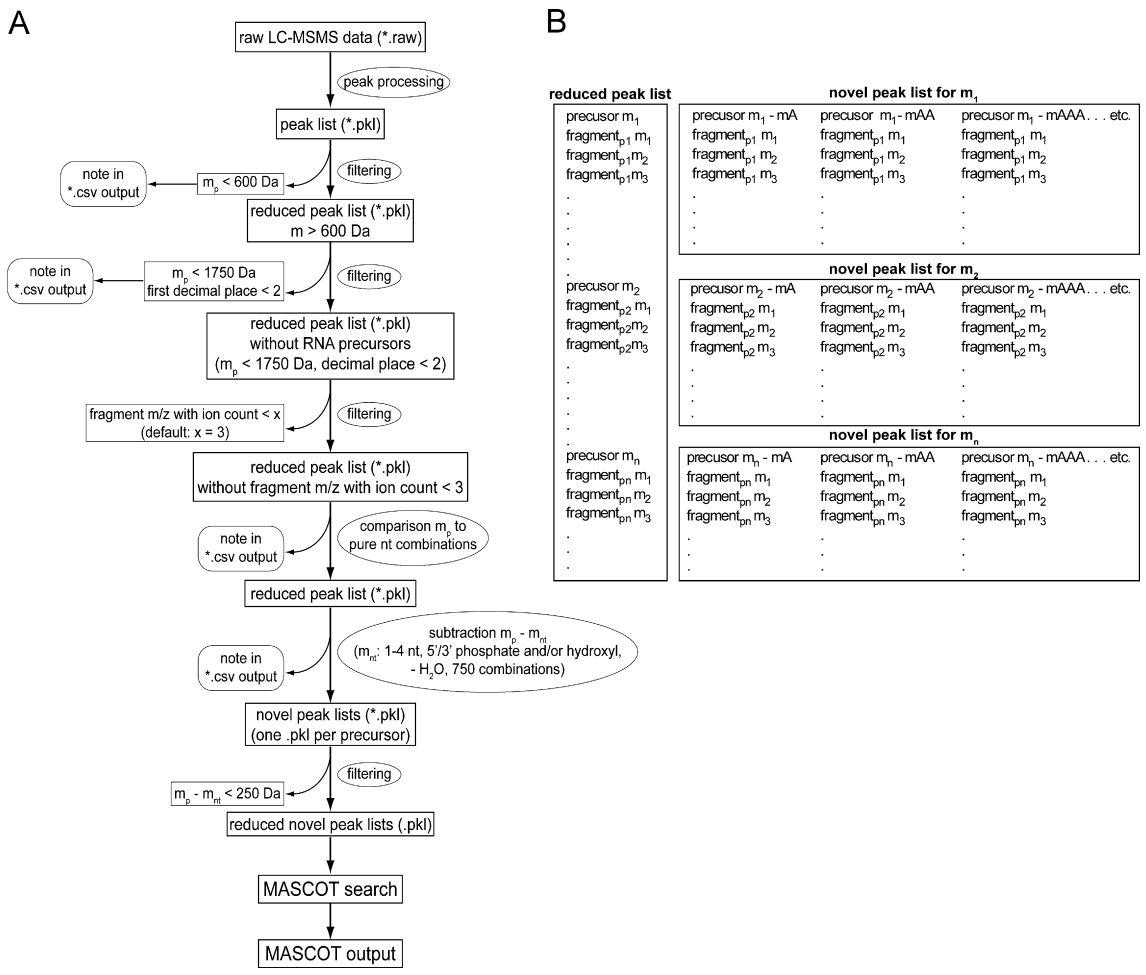

Fig. 3. Scheme of automated data analysis workflow. (A) LC–MSMS data is converted into the text-based .pkl file format. Small precursor masses (m < 600 Da), precursors corresponding to pure RNA fragments (m < 1750 Da, first decimal place <.2) are not considered for the further data analysis workflow. Noise signals below a given absolute intensity are deleted from the MSMS data (default ion count: 3). After evaluating whether the precursor mass could correspond to a pure oligonucleotide, all possible nucleotide combinations (1–4 nt, 5'/3' phosphate or hydroxyl, loss of water) are subtracted from the experimental precursor mass. If the resulting modified precursor masses are smaller than 250 Da, they are omitted from the output pkl file (last filter). For each MSMS spectrum, a separate .pkl file is created, containing copies of the MSMS data with the original and all modified precursor masses (see also panel b). Additionally, a .csv file notes those data sets that were filtered for small precursor mass or decimal place and nucleotide combinations corresponding to certain precursor masses. The original and the newly created .pkl files are then submitted to a MASCOT search. (B) Scheme of reduced peak list (left) and novel peak lists created by the program (right). For each MSMS spectrum/each precursor mass (precursor m_x) a novel peak list is created by subtracting the mass of all nucleotide combinations, e.g., A, AA, AAA. . . from the experimental precursor mass. The data of the MSMS fragments remains unchanged and is copied under each modified precursor mass ($fragment_{px}$ m_1; $fragment_{px}$ m_2; $fragment_{px}$ m_3 . . .).

complexes or unbiased searches against entire proteomes when the interaction partners have not been yet identified.

We therefore set out to create modified data files that can be used to identify the cross-linked peptide moiety by using MASCOT as search engine (Fig. 3). We developed a perl-based script on the basis of the text-based pkl file format of processed MSMS data recorded on a Waters Q-ToF instrument. The script was set up to read the original pkl and, after several processing steps, creates one pkl file for every MSMS spectrum potentially containing a cross-link. We first applied various filtering steps to exclude precursors not corresponding to cross-links and thus reduce the overall number of MSMS spectra, including the following: (i) we removed precursor masses (with their corresponding fragment masses) below 600 Da, as we anticipate a minimum of three amino acids and one nucleotide for the cross-link to be detectable at all under our conditions. (ii) We removed precursor masses that correspond to sole RNA oligonucleotides, i.e., those are smaller than 1750 Da and have a fractional mass that is 0 or 1 in the first decimal place. Note that the mass deficiency of oligonucleotides differs from that of pure peptides due to the larger content of oxygen and phosphorus atoms with a negative mass deficiency and can be used to distinguish between oligonucleotides, peptides and their heteroconjugates [30]. These constraints for exclusion of RNA oligonucleotides are consistent with data reported in this study. (iii) We reduced the data within the remaining peak list by deleting fragment masses with absolute intensities below a certain threshold (e.g., three counts). Note that this step is optional and should be applied with care.

Once the peak list has been trimmed, all possible calculated nucleotide compositions between one and a maximum of four nucleotides (including all possibilities for the 5′ and 3′ ends of the RNA fragment, i.e., both phosphates, phosphate and hydroxyl, or both hydroxyls, and loss of water) are subtracted from the remaining precursor masses in the modified peak list. Note that with the use of TiO_2 for the final enrichment step no cross-links have so far been identified that harboured a RNA oligonucleotide with more than four nucleotides. A final filtering step was included to remove generated precursor masses below 250 Da (corresponding to less than three amino acids). The output comprises of several peak lists, each originating from one single precursor mass, that still contain the centroided mass information of the fragment ions that are essentially unchanged (except for noise signals, see point (iii)

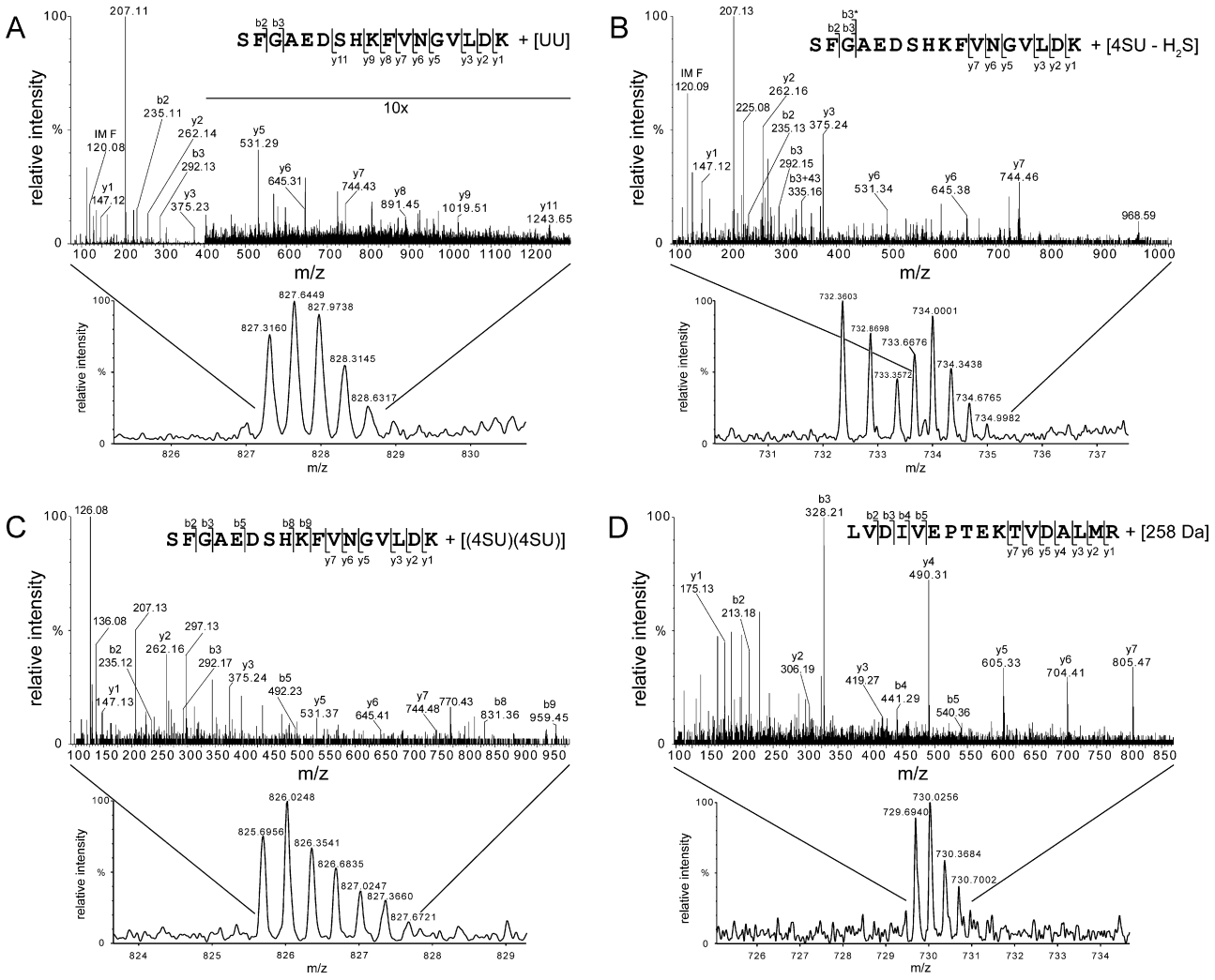

Fig. 4. MS and MSMS spectra of cross-linked species. Corresponding monoisotopic masses are shown at the bottom and the corresponding fragment spectrum with the peptide sequence at the top. (A) Spectra of NusB-peptide SFGAEDSHKFVNGVLDK (positions 113–129) cross-linked to a UU dinucleotide. Peptide sequence and identified CID fragments are indicated in the upper right. For better visualization, intensities were increased tenfold for *m*/*z* greater than 400. IM F: immonium ion of phenylalanine. (B) MS and MSMS spectra of NusB-peptide SFGAEDSHKFVNGVLDK (positions 113–129) cross-linked to 4-thio-uracil via the loss of H_2S. Peptide sequence and identified fragment ions are indicated on the upper right. The b-ion shifted 43.01 Da due to the peptide's N-terminal carbamylation is indicated with an asterisk. IM F: immonium ion of phenylalanine. (C) MS and MSMS spectrum of NusB-peptide SFGAEDSHKFVNGVLDK (positions 113–129) cross-linked to a (4SU)(4SU) dinucleotide without any terminal phosphate. Peptide sequence and identified fragments are indicated in the upper right. (D) Spectra of S10-peptide LVDIVEPTEKTVDALMR (positions 73–89) cross-linked to a species with the mass of 258.0108 Da. Peptide sequence and identified b- and y-ions are indicated in the upper right corner.

above and Fig. 3B). The unaltered precursor mass is kept in the new file, thus allowing the identification of noncross-linked peptides in the data analysis workflow. These peak lists are subsequently submitted to a database search using MASCOT as search engine. Upon submission of the files into database search, the ideal result would be one hit for the modified precursor mass corresponding to the cross-linked peptide moiety, while the difference between experimental and altered precursor masses would represent the cross-linked RNA moiety. All positive hits were confirmed manually.

Although the program was developed and tested using MS data obtained from cross-linking experiments with native *rrn* BoxA-containing RNA (oligo 1a), we adapted it for our cross-linking experiments with 4SU. The mass of corresponding nucleotide (exact mass 340.0130 Da) was added to the four natural nucleotides, and an additional version of the program was created that subtracted all possible combinations of A, C, G, U and 4SU.

Manual identification of the cross-linked peptide was attempted on MSMS spectra that fulfilled the following criteria: (i) no identification as pure peptide by standard MASCOT database search, (ii) precursor mass above 600 Da and (iii) the decimal place did not indicate pure RNA (see above, decimal place <.2 for precursor masses <1750 Da). If a peptide sequence tag could be identified, the fragmentation series generated *in silico* for corresponding peptides was matched to the experimental fragmentation series. Once the peptide moiety of the cross-link was identified, the difference between the experimental precursor mass and the theoretical peptide mass was calculated. This mass difference represented the cross-linked RNA moiety.

3.3. 4SU cross-links in [NusB–S10–BoxA-containing RNA oligonucleotide]

We applied both the workflows outlined in Figs. 2 and 3 for detection and identification of cross-links between 4SU and the proteins NusB and S10 in a [NusB–S10–BoxA-containing oligonucleotide] complex assembled *in vitro*, in which the three uridines at positions 9 to 11 were replaced by 4SU (5′-CACUGCUC(4SU)(4SU)(4SU)AACAAUUA-3′; oligo 1b).

Fig. 4 shows the MS and MSMS fragment spectra of four cross-linked species derived from the NusB–S10 complex (SFGAED-

SHKFVNGVLDK, positions 113–129 derived from NusB; and LVDIVEPTEKTVDALMR, positions 73–89 from S10) after UV cross-linking of non-substituted oligo 1a (Fig. 4A) and 4SU-substituted oligo 1b (Fig. 4B–D). Using our strategy as outlined above, we performed a MASCOT search and identified the peptide SFGAED-SHKFVNGVLDK as a putative cross-linked peptide moiety derived from cross-linking with native oligo 1a (Fig. 4A) including the outlined b- and y-type ions that identified the peptide in the database search. Importantly, the modified database search revealed a mass difference between the peptide precursor mass found and the unaltered experimental precursor mass; this difference was 630.0241 Da (precursor 2478.9246 Da, theoretical peptide mass 1848.9005 Da). The theoretical mass of 630.0612 Da corresponds to a UU-dinucleotide. These results are consistent with our previous data [9] and thus confirm that our approach is suitable for the detection of cross-linked peptide moieties.

We next applied our search strategy to the detection of putative cross-linked peptides isolated from UV-irradiated complexes that contained RNA oligonucleotides with 4SU-substitutions (oligo 1b). Fig. 4B shows the MS and MSMS spectra from the same peptide isolated from the 4SU-substituted complex rather than the native complex. The MASCOT search assigned the listed b- and y-type ions that unambiguously identified the peptide (and the protein) in the database search with a carbamylated K121 (data not shown). Note that cross-linked complexes were hydrolysed with ribonucleases and endoproteinase in the presence of urea (see Section 2), thus leading to carbamylation (43.01 Da) at the elevated temperature of the incubation (52 °C). However, manual inspection revealed that the b3-ion was present both unmodified and also shifted 43.01 Da (Fig. 4B). We reason that this particular peptide could be present as a mixture of two isobaric forms with carbamylation at K121 and at the N-terminal amino acid. The calculated mass difference between the precursor mass (2197.9794 Da) and the carbamylated peptide (1891.9063 Da) is 306.0731 Da. The experimental mass difference of 306.07 Da must result from either a cross-link to an unsubstituted U and the subsequent loss of a water molecule, or a cross-link to 4SU with the loss of H_2S. Since the cross-linked nucleotides have the same elemental composition, we cannot distinguish on the basis of MS whether cross-linking has occurred via a 4SU or via a non-modified U within the BoxA-containing RNA oligonucleotide. However, as UV irradiation at 365 nm selectively activates 4SU, we infer that cross-link formation occurred via 4SU involving loss of H_2S. The actual cross-linked amino acid could not be identified by MSMS.

Interestingly, in the higher m/z region a fragment ion of $m/z = 968.56$ is still detectable (Fig. 4B). The mass difference between the y7 ion and this particular fragment ion is 244 Da, which might suggest that 4SU is cross-linked to the phenylalanine (see below).

Fig. 4C shows the MSMS spectrum of the same carbamylated peptide potentially cross-linked to the 4SU-substituted oligo 1b. A database search revealed that the mass difference between the carbamylated peptide (1891.9063 Da) and the unaltered precursor mass (2475.0387 Da) was 583.1324 Da. This corresponds to a 4SU dinucleotide lacking its 3′-phosphate group. However, the fragment-ion series does not allow the actual cross-linking site within the peptide to be determined, and even the site of carbamylation (although predicted by MASCOT search to be K121) cannot be confirmed.

Fig. 4D shows the MS and MSMS spectra of a peptide derived from protein S10 that was identified manually by the corresponding b- and y-type ions. Importantly, the mass difference between the unaltered precursor mass (2186.0996 Da) and that of the peptide alone (1928.0288 Da) is 258.0708 Da. Such a mass difference has never been observed in previous cross-linking studies with non-substituted RNA [5–9,12,31–34]. We also found this particular mass difference in other putatively cross-linked peptides (see below) and we assign this to a 4SU-specific derivative (see below).

3.4. MSMS reveals unique peptide–nucleotide cross-linking products derived from 4SU-substituted RNA oligonucleotides

From the MSMS spectrum shown in Fig. 4D we could only calculate a potential cross-linked adduct of 258.0708 Da. Interestingly, when applying our MASCOT-based search, we observed – in general, compared with cross-linking with unsubstituted RNA [9] – more missed cleavage at lysine residues within the listed potentially cross-linked peptide moieties than in the cross-linked product from native RNA. This leads us to the suggestion/assumption that the lysine residues could be indeed the amino acid actually cross-linked to 4SU.

Fig. 5A shows the MSMS analysis of a cross-linked peptide moiety derived from NusB (IALYELSKR, positions 87–95) with an unaltered precursor mass (1349.6914 Da) differing from that of the peptide alone (1091.6338 Da) also by 258.0576 Da. Strikingly, the y-type fragment ion pattern reveals a modification of K94 of exactly this mass. However, we also observe within this particular spectrum y4, y5, y6 and y7 fragment ions in non-modified form. We conclude that K94 is the actual modified amino acid and that the appearance of non-modified y4 to y7 ions could be due to the partial loss of the modification upon CID. The additional mass of 258.0576 Da can be calculated as corresponding to a cross-link product between lysine and 4SU nucleoside without the 3′ phosphate group (see below and Fig. 7C). The observation that an amino acid modification derived from a cross-link is so stable in CID is rather unusual. In previous studies, peptide fragments only, nucleic acid bases and single nucleotides were typically observed under CID conditions, or, in the case of longer RNA oligonucleotides, only fragment ions derived from oligonucleotide moiety were observed [5,7,8,12]. Only in very rare cases, such as reported in [5,7,8], can peptide fragments bound to nucleotide fragments be observed.

Fig. 5B shows another example of an additional mass conjugated to a lysine residue within the putatively cross-linked peptide LIDQATAEIVETAKR (positions 17–31) derived from protein S10 after cross-linking that was identified manually. The mass difference between precursor mass (1962.9734 Da) and the mass of the peptide (1656.9046 Da) was calculated to be 306.0688 Da. Interestingly, all the observed y-type ions in the fragment spectrum except for y1 are shifted by 94.05 Da, strongly indicating that the lysine residue is indeed cross-linked. A possible explanation for this additional mass is the partial fragmentation of a 4SU cross-linked via loss of H_2S to a lysine residue under CID conditions, which would leave a fragment of the nucleobase still attached to lysine (see Fig. 7B and Section 4).

3.5. Cross-linking to 5BrU-substituted RNA

We also investigated cross-linking at 312 nm of the NusB–S10 complex with a 5BrU-substituted γ BoxA core RNA oligonucleotide (5′-CGC(5BrU)CUUACACA-3′, oligo 2). Fig. 6 shows the MSMS spectrum of a peptide derived from the protein NusB (positions 9–41, sequence ARECAVQALYSWQLSQNDIAD-VEYQFLAEQDVK). The listed y-type ion series was assigned manually and a mass difference between the peptide precursor mass (3829.8257 Da) and the unaltered precursor mass (5124.8575 Da) of 1295.0318 Da was calculated. This mass difference cannot easily be explained. Previous cross-linking studies on bromo-uracil derivatives by MS showed that cross-linking led to the loss of HBr and that the C5–C6 bond of uracil remained intact. The appearance of two marker ions in the lower m/z region, i.e., the nucleobases guanine and cytidine at $m/z = 152$ and 112 Da respectively [7], and the location of the 5BrU within oligo 2 (see above),

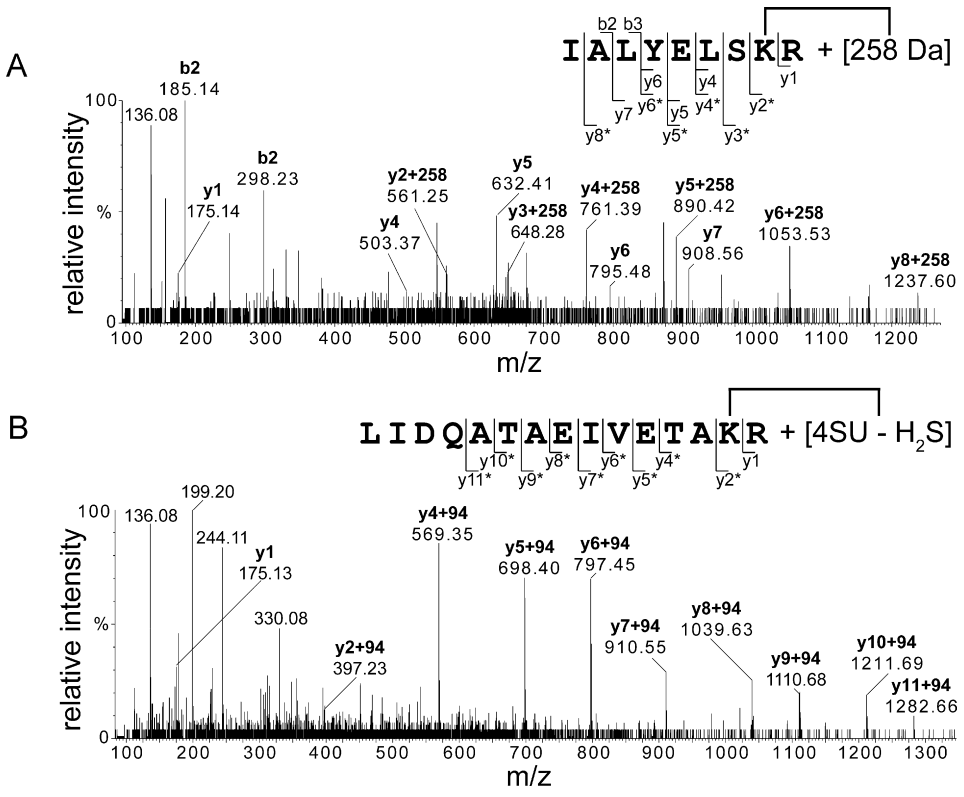

Fig. 5. MSMS spectra of two cross-linked species that display a mass shift of peptide fragments due to cross-linked fragments of a 4-thio-uridine. (A) MSMS spectrum of NusB-peptide IALYELSKR (positions 87–95) carrying an additional mass of 258.0576 Da. Peptide sequence and identified fragment ions are indicated on the upper right. Y-ions that are shifted 258.05 Da are marked with an asterisk. (B) MSMS spectrum of cross-link comprised of S10-peptide LIDQATAEIVETAKR (positions 17–31) and a 4-thio-uracil nucleotide after loss of H_2S. Peptide sequence and identified fragment ions are indicated in the upper right corner. Fragment ions exhibiting an additional mass of 94.05 Da are indicated by an asterisk.

all suggest that the cross-linked RNA sequence is 5′-GCC(5BrU). Indeed, the observed mass difference of 1295.0318 Da is obtained when one takes into account the mass of the RNA oligonucleotide, the loss of HBr and the addition of 2 H atoms and 1 O atom (18.0106 Da). Whether the latter is "simply" the addition of water, or whether the C5–C6 double bond is hydrated and an amino acid within the peptide is oxidised, cannot be answered on the basis of the MSMS spectrum of the cross-link alone.

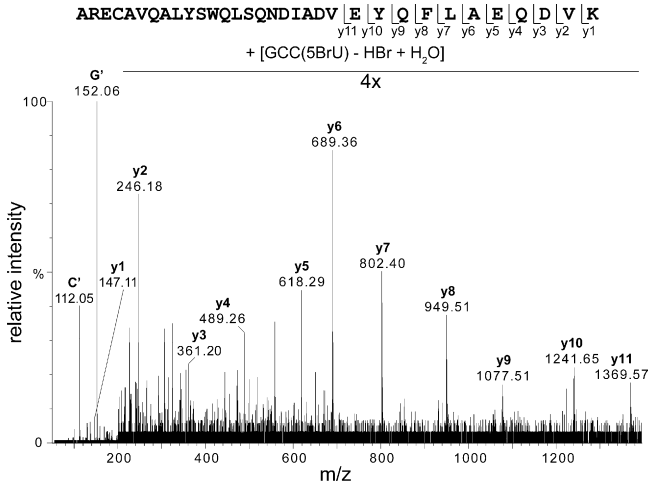

Fig. 6. MSMS spectrum of NusB-peptide ARECAVQALYSWQLSQNDIADVEYQFLAEQDVK (positions 9–41) cross-linked to a GCC(5BrU) oligonucleotide via loss of HBr and uptake of H_2O. Peptide sequence and identified y-type ions are indicated above the spectrum. For better visualization, signal intensities were increased 4-fold for *m*/*z* greater 200.

4. Discussion

In this study we compared cross-linking of proteins to non-substituted RNA with cross-linking to 4SU- or 5BrU-substituted RNA to investigate the overall mass of the cross-linking products with substituted RNAs. For this purpose we carried out cross-linking on a [NusB–S10–BoxA-containing RNA oligonucleotide] complex reconstituted *in vitro* [9] and further used and improved a database search for the identification of cross-linked heteroconjugates after UV irradiation.

4.1. Automated (computational) data analysis workflow

As the novel data analysis workflow was at first developed considering the results from cross-linking experiments with unsubstituted RNA, and, more importantly, as the mass of the cross-links containing 4SU was not known, not all cross-links described above were identified in an automated data analysis. However, our strategy was successful in the identification purely additive cross-link of NusB-peptide SFGAEDSHKFVNGVLDK (positions 113–129) to a (4SU)(4SU) dinucleotide (see Fig. 4C) and all cross-links in which the RNA moiety exhibited a mass of 306 Da (SFGAEDSHK-FVNGVLDK (NusB, positions 113–129), Fig. 4B; LIDQATAEIVETAKR (S10, positions 17–31), Fig. 5B). In the latter cases the identification was possible because 306 Da also corresponds to a cross-link to an unsubstituted uridine via the loss of H_2O.

After all cross-linking products derived from 4SU had been characterized by MS, the loss of H_2 upon cross-linking, i.e., an additional mass of 258 Da (Fig. 7C), was implemented into the program. In addition, two novel posttranslational modifications (PTM) were defined in the MASCOT search engine. The first PTM was set to

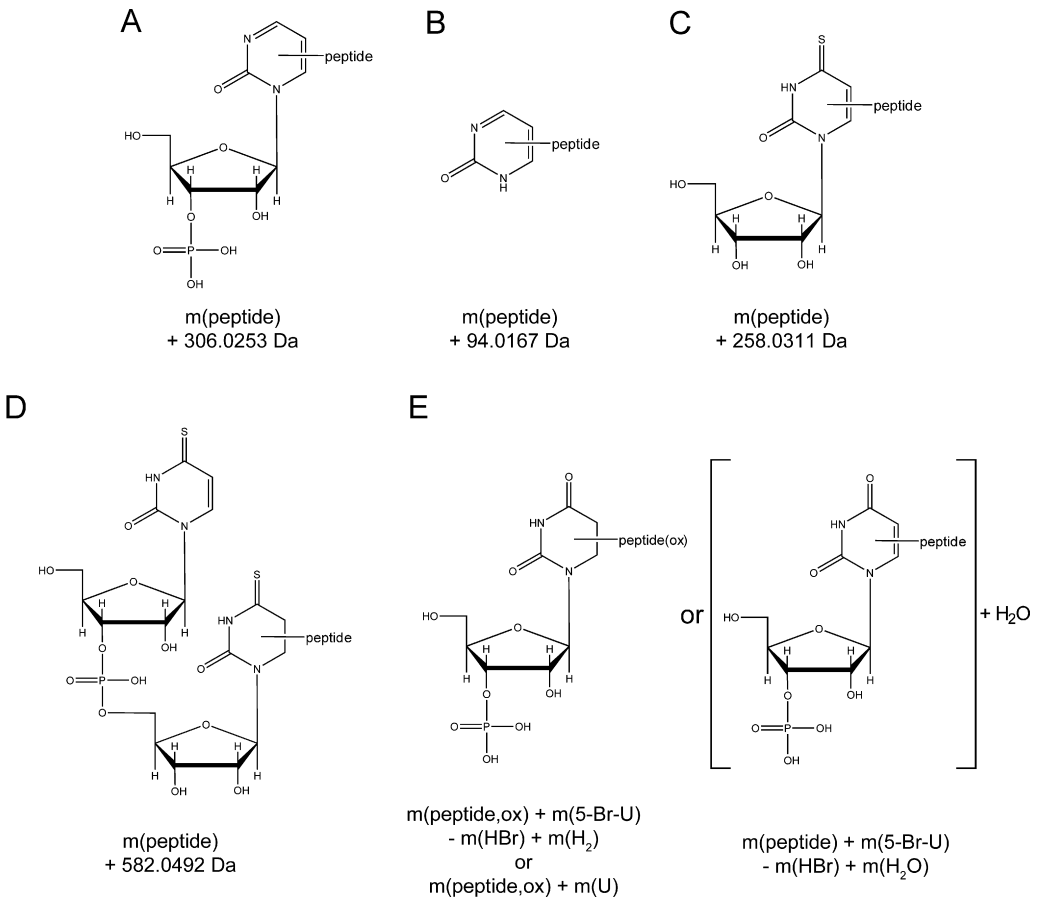

Fig. 7. Proposed structures resulting from cross-linking of proteins and substituted RNAs. The theoretical mass of the cross-linked species is given as the sum of peptide mass m(peptide) and the mass of the cross-linked nucleotide fragment below each structure. (A) Nucleotide fragment accounting for an additional mass of 306.0253 Da as observed for several cross-linked peptides, e.g., the examples shown in Figs. 4 and 5B. (B) Nucleotide fragment accounting for an additional mass of 94.0167 Da as observed for the y-ions of S10-peptide LIDQATAEIVETAKR (Fig. 5B). (C) Proposed structure accounting for an additional mass of 258.0311 Da as observed for several peptides as well as for the CID fragments of NusB-peptide IALYELSKR (Fig. 5A). (D) Proposed structure of the 4SU dinucleotide cross-linked to NusB-peptide SFGAEDSHKFVNGVLDK (Fig. 4C). (E) Proposed structure of the cross-linked region derived from a 5-bromo-uracil substituted RNA (Fig. 6).

have a mass of 306 Da and a potential neutral loss of 212 Da (corresponding to 94 Da remaining after neutral loss). This would enable the identification of e.g., the spectrum shown in Fig. 5B, with a mass difference of precursor and peptide mass of 306 Da and peptide fragments shifted 94 Da. The second PTM was defined as having an overall mass of 258 Da and a neutral loss of 258 Da, respectively, thus enabling the identification of spectra as the example shown in Fig. 5A.

Adaptation of the observed cross-linking products into our novel data analysis strategy at the appropriate stage, i.e., either in our program or in MASCOT, enabled the successful identification of all cross-links shown in this study by reanalysing the MS data. In all examples presented here, the cross-linked peptide was the only identified hit for the respective MS/MS data in the MASCOT search against the small database comprising 14 proteins. MASCOT did not report any matches to the other combinations of the same MS/MS fragments with different modified precursor masses. The same search was repeated against the NCBInr database with the taxonomy limited to *E. coli*. In all cases except for the cross-link shown in Fig. 3B, the cross-linked peptide was the highest scoring hit. All other, false positive matches exhibited noticeable lower scores and higher expectation values. In addition, all other matches could be ruled out as false positives, either simply because they corresponded to proteins not contained in the sample, or by manual evaluation of the search results.

4.2. Structural implications of 4SU–peptide cross-links

By using UV irradiation conditions suitable for 4SU cross-linking we found peptides IALYELSKR (positions 87–95), SDVPYKVAINEAIELAK (positions 96–112), SFGAEDSHKFVNGVLDK (positions 113–129) derived from protein NusB, and the peptides LIDQATAEIVETAKR (positions 17–31) and LVDIVEPTEKTVDALMR (positions 73–89) derived from protein S10, cross-linked to BoxA-containing RNA oligonucleotide.

MS and MSMS of these peptides showed several novel conjugated masses, i.e., 582 Da, 306 Da, 258 Da, and 94 Da. These masses represent cross-link adducts derived from 4SU, and the corresponding structures of the conjugates are outlined in Fig. 7. The extra mass of 306 was found in the cross-linked peptide SFGAEDSHKFVNGVLDK (positions 113–129) derived from NusB that was also identified in our previous study as cross-linked to a CU dinucleotide in non-substituted *rrn* BoxA-containing RNA [9]. However, the mass of 306 can represent both a conjugated uridine with a 3′–5′ cyclic phosphate or a derivative of 4SU conjugated to the peptide with loss of H_2S. We favour the latter possibility for two reasons (see Fig. 7A): (i) UV irradiation was conducted at 365 nm, which is in the absorbance maximum of 4SU but is far from that of uridine, and (ii) we have sequenced another conjugate the precursor mass of which also showed an extra mass of 306 Da (Fig. 5B). In MSMS of this particular precursor we observed a mass tag of 94 Da still attached to a lysine residue. The extra mass of 94 Da represents a

derivative of the nucleobase 4SU attached to a lysine via loss of H_2S and breakage of the N-glycosidic bond upon CID fragmentation (Fig. 7B). Both these findings suggest a mechanism in which – upon UV irradiation – lysine is covalently bound to 4SU with subsequently loss of H_2S. The exact structure is not revealed directly by our data, i.e., we cannot deduce directly whether the ε-N-atom of lysine replaces the sulphur atom and is thus conjugated to the C-1 atom of uracil or whether it is conjugated via the C5 or C6 atom of the uracil derivative.

Within this particular NusB peptide (SFGAEDSHKFVNGVLDK) we could not identify the actual cross-linked amino acid, neither in the peptide found to be cross-linked to non-substituted RNA (Fig. 4A) nor in that cross-linked to 4SU (Fig. 4B and C). Remarkably, the MSMS spectrum in the latter case reveals an almost complete y-type ion series up to y7 (V123) and a gap between the y7 and a fragment at $m/z = 968$. However, the mass difference between both these fragments matches exactly the mass of phenylalanine (147 Da) plus 94 Da (potentially corresponding to the above described mass tag) and the loss of NH_3 (which is frequently observed in y-type ion series). Based on this observation, we conclude that F124 might be the cross-linked amino acid.

Interestingly, the UV cross-linking reaction to 4SU does not seem necessarily always to result in the loss of H_2S. We have identified two cross-linked peptides (IALYELSKR, positions 87–95) from NusB and LVDIVEPTEKTVDALMR (positions 73–89 from S10) that reveal an additional mass of 258 Da. This mass corresponds to a 4SU nucleoside. Consequently, these peptides must have been cross-linked to 4SU without the loss of H_2S but also leaving the C5–C6 double bond in 4SU intact (Fig. 7C), i.e., two hydrogen atoms are lost upon cross-linking.

In the case of UV-induced cross-linking to halo-substituted nucleobases, previous MS studies showed that cross-linking led to a loss of e.g., HBr [16] and thus an addition of the peptide side chain at the C5 position of uracil. In our case we cannot distinguish whether we observed a similar cross-linking reaction and addition of a water molecule to heteroconjugate or whether the C5–C6 double bond is hydrated and an amino acid residue of the cross-linked peptide is oxidised in addition. However, as the MSMS spectrum is only recorded up to m/z 2000 and no b-ions are visible, it does not reveal unambiguously this particular peptide fragment at which the cross-linked nucleobase is still attached.

4.3. Comparison of our cross-linking experiments with substituted RNAs to previous experiments with unsubstituted RNA

The gel-electrophoretic analysis of the [NusB–S10–BoxA-containing RNA] complex has shown that the incorporation of 4SU drastically enhances the cross-linking yield of only S10 protein whereas the cross-linking of NusB to the RNA seems to be unaffected. This result was surprising, in particular when considering the location of cross-linked peptides within the two proteins. In previous cross-linking studies two peptides derived from NusB have been found as cross-linking sites to this particular *rrn* BoxA-containing RNA oligonucleotide, i.e., SDVPYKVAINEAIELAK (positions 96–112), and FVNGVLDK (positions 122–129). Importantly, the location of the NusB-peptides cross-linked to 4SU is congruent with that of the previous study with non-substituted RNA except that we found the peptides SFGAEDSHKFVNGVLDK (positions 113-129) and IALYELSKR (positions 87-95) also cross-linked to the poly-U stretch within the BoxA. Remarkably, the latter precedes directly the NusB peptide SDVPYKVAINEAIELAK (positions 96–112) and the other stretch encompassing positions 113–129 harbours the sequence FVNGVLDK. In fact, more peptides were found to be cross-linked to the poly-U stretch of the BoxA-containing RNA oligonucleotide used here as compared with experiments with non-substituted RNA [9]. The fact that the yield of cross-link did not increase when 4SU was used suggests that these particular peptides, i.e., SDVPYKVAINEAIELAK and FVNGVLDK, are highly reactive towards BoxA-containing RNA oligonucleotide even under less effective UV irradiation conditions.

In contrast to NusB, the cross-linking yield of S10 to BoxA-containing RNA oligonucleotide is drastically enhanced when 4SU substituted RNA is used (Fig. 1). Surprisingly, we could not identify those peptides that were found in the previous study to be cross-linked to unsubstituted *rrn* BoxA-containing RNA oligonucleotide. Instead we found three different regions within S10 to be cross-linked, i.e., LIDQATAEIVETAKR (positions 17–31), LVDIVEPTEKTVDALMR (positions 73–89) and GPIPLPTR (positions 38–45). Strong cross-linking of the 4SU containing RNA towards these regions must be the reason for the enhanced cross-linking yield. However, the location of these peptides is not congruent with our previous study, in which the cross-linked peptide were mainly located in a loop region encompassing positions 46–67 in S10 which was deleted in the NusB–S10 complex used for crystallisation [9]. Currently, we do not have any satisfactory explanation for this discrepancy. Interestingly, another peptide derived from S10 (GPIPLPTR, positions 38–45) that was also found to be cross-linked to 4SU-substituted RNA (data not shown) is located in the crystal structure at the interface between both the proteins NusB and S10 [9]. Moreover, the cross-linked peptide derived from NusB encompassing positions 9–41 that was found to be cross-linked to 5BrU position 4 of the γ BoxA core RNA oligonucleotide also harbours amino acid residues that participate in the interaction surface of both the proteins. The question whether the newly identified peptides in S10 and whether indeed the interface between both the proteins is also involved in the interaction with RNA can only be answered by additional biochemical and structural studies.

Here, we demonstrate that identification of cross-linked peptides can lead to valuable conclusions about protein–RNA interaction surfaces. However, in many cases (in this and in our previous study) we could not identify the actual cross-linked amino acid which would be essential for complete interpretation of the cross-link data on the molecular level.

In summary, this feasibility study will proof useful – in combination with here further developed computational approached – for a broader MS-based analysis of UV cross-linking in more complex protein–RNA assemblies and/or of entire cells.

Acknowledgements

We thank Johanna Lehne, Uwe Pleßmann and Monika Raabe for their excellent technical assistance and Klaus Hartmuth for helpful discussions.

At the moment it is not planned to provide a download of the perl script for the general public, however it can be made available upon personal request.

References

[1] M.D. Shetlar, J. Christensen, K. Hom, Photochemical addition of amino acids and peptides to DNA, Photochem. Photobiol. 39 (1984) 125–133.
[2] E.I. Budowsky, M.S. Axentyeva, G.G. Abdurashidova, N.A. Simukova, L.B. Rubin, Induction of polynucleotide-protein cross-linkages by ultraviolet irradiation. Peculiarities of the high-intensity laser pulse irradiation, Eur. J. Biochem. 159 (1986) 95–101.
[3] W. Tate, B. Greuer, R. Brimacombe, Codon recognition in polypeptide chain termination: site directed crosslinking of termination codon to *Escherichia coli* release factor 2, Nucleic Acids Res. 18 (1990) 6537–6544.
[4] H. Urlaub, B. Thiede, E.C. Muller, R. Brimacombe, B. Wittmann-Liebold, Identification sequence analysis of contact sites between ribosomal proteins and rRNA in *Escherichia coli* 30 S subunits by a new approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing, J. Biol. Chem. 272 (1997) 14547–14555.
[5] E. Kuhn-Holsken, C. Lenz, B. Sander, R. Luhrmann, H. Urlaub, Complete MALDI-ToF MS analysis of cross-linked peptide-RNA oligonucleotides derived

from nonlabeled UV-irradiated ribonucleoprotein particles, RNA 11 (2005) 1915–1930.
[6] E. Kuhn-Holsken, O. Dybkov, B. Sander, R. Luhrmann, H. Urlaub, Improved identification of enriched peptide RNA cross-links from ribonucleoprotein particles (RNPs) by mass spectrometry, Nucleic Acids Res. 35 (2007) e95.
[7] C. Lenz, E. Kuhn-Holsken, H. Urlaub, Detection of protein–RNA crosslinks by NanoLC–ESI-MS/MS using precursor ion scanning and multiple reaction monitoring (MRM) experiments, J. Am. Soc. Mass Spectrom. 18 (2007) 869–881.
[8] F. Martin Richter, H.H. Hsiao, U. Plessmann, H. Urlaub, Enrichment of protein–RNA crosslinks from crude UV-irradiated mixtures for MS analysis by on-line chromatography using titanium dioxide columns, Biopolymers 91 (2009) 297–309.
[9] X. Luo, H.H. Hsiao, M. Bubunenko, G. Weber, D.L. Court, M.E. Gottesman, H. Urlaub, M.C. Wahl, Structural and functional analysis of the *E. coli* NusB–S10 transcription antitermination complex, Mol. Cell 32 (2008) 791–802.
[10] H. Steen, J. Petersen, M. Mann, O.N. Jensen, Mass spectrometric analysis of a UV-cross-linked protein-DNA complex: tryptophans 54 and 88 of *E. coli* SSB cross-link to DNA, Protein Sci. 10 (2001) 1989–2001.
[11] V. Pingoud, H. Geyer, R. Geyer, E. Kubareva, J.M. Bujnicki, A. Pingoud, Identification of base-specific contacts in protein-DNA complexes by photocrosslinking and mass spectrometry: a case study using the restriction endonuclease SsoII, Mol. Biosyst. 1 (2005) 135–141.
[12] E. Kuhn-Holsken, C. Lenz, A. Dickmanns, H.H. Hsiao, F.M. Richter, B. Kastner, R. Ficner, H. Urlaub, Mapping the binding site of snurportin 1 on native U1 snRNP by cross-linking and mass spectrometry, Nucleic Acids Res. (2010).
[13] K.M. Meisenheimer, T.H. Koch, Photocross-linking of nucleic acids to associated proteins, Crit. Rev. Biochem. Mol. Biol. 32 (1997) 101–140.
[14] E.E. Blatter, Y.W. Ebright, R.H. Ebright, Identification of an amino acid-base contact in the GCN4-DNA complex by bromouracil-mediated photocrosslinking, Nature 359 (1992) 650–652.
[15] B.J. Hicke, M.C. Willis, T.H. Koch, T.R. Cech, Telomeric protein–DNA point contacts identified by photo-cross-linking using 5-bromodeoxyuridine, Biochemistry 33 (1994) 3364–3373.
[16] M.C. Golden, K.A. Resing, B.D. Collins, M.C. Willis, T.H. Koch, Mass spectral characterization of a protein-nucleic acid photocrosslink, Protein Sci. 8 (1999) 2806–2812.
[17] H. Geyer, R. Geyer, V. Pingoud, A novel strategy for the identification of protein–DNA contacts by photocrosslinking and mass spectrometry, Nucleic Acids Res. 32 (2004) e132.
[18] I. Dix, C.S. Russell, R.T. O'Keefe, A.J. Newman, J.D. Beggs, Protein–RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*, RNA 4 (1998) 1239–1250.
[19] A. Favre, R. Bezerra, E. Hajnsdorf, Y. Lemaigre Dubreuil, A. Expert-Bezancon, Substitution of uridine in vivo by the intrinsic photoactivable probe 4-thiouridine in *Escherichia coli* RNA. Its use for *E. coli* ribosome structural analysis, Eur. J. Biochem. 160 (1986) 441–449.
[20] R. Bezerra, A. Favre, In vivo incorporation of the intrinsic photolabel 4-thiouridine into *Escherichia coli* RNAs, Biochem. Biophys. Res. Commun. 166 (1990) 29–37.
[21] A. Favre, G. Moreno, M.O. Blondel, J. Kliber, F. Vinzens, C. Salet, 4-Thiouridine photosensitized RNA-protein crosslinking in mammalian cells, Biochem. Biophys. Res. Commun. 141 (1986) 847–854.
[22] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano Jr., A.C. Jungkamp, M. Munschauer, A. Ulrich, G.S. Wardle, S. Dewell, M. Zavolan, T. Tuschl, Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP, Cell 141 (2010) 129–141.
[23] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis 20 (1999) 3551–3567.
[24] F.W. Studier, Protein production by auto-induction in high density shaking cultures, Protein Expr. Purif. 41 (2005) 207–234.
[25] M.R. Larsen, T.E. Thingholm, O.N. Jensen, P. Roepstorff, T.J. Jorgensen, Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns, Mol. Cell. Proteomics 4 (2005) 873–886.
[26] J.V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann, Higher-energy C-trap dissociation for peptide modification analysis, Nat. Methods 4 (2007) 709–712.
[27] K.L. Krivos, P.A. Limbach, Sequence analysis of peptide:oligonucleotide heteroconjugates by electron capture dissociation and electron transfer dissociation, J. Am. Soc. Mass Spectrom. 21 (2010) 1387–1397.
[28] D. Fabris, E.T. Yu, Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS3D, J. Mass Spectrom. 45 (2010) 841–860.
[29] E.T. Yu, A. Hawkins, I.D. Kuntz, L.A. Rahn, A. Rothfuss, K. Sale, M.M. Young, C.L. Yang, C.M. Pancerella, D. Fabris, The collaboratory for MS3D: a new cyberinfrastructure for the structural elucidation of biological macromolecules and their assemblies using mass spectrometry-based approaches, J. Proteome Res. 7 (2008) 4848–4857.
[30] S. Pourshahian, P.A. Limbach, Application of fractional mass for the identification of peptide-oligonucleotide cross-links by mass spectrometry, J. Mass Spectrom. 43 (2008) 1081–1088.
[31] H. Urlaub, K. Hartmuth, S. Kostka, G. Grelle, R. Luhrmann, 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 (2000) 41458–41468.
[32] H. Urlaub, K. Hartmuth, R. Luhrmann, A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles, Methods 26 (2002) 170–181.
[33] H Urlaub, E. Kuhn-Holsken, R. Luhrmann, Analyzing RNA–protein crosslinking sites in unlabeled ribonucleoprotein complexes by mass spectrometry, Methods Mol. Biol. 488 (2008) 221–245.
[34] H. Urlaub, V.A. Raker, S. Kostka, R. Luhrmann, Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure, EMBO J. 20 (2001) 187–196.