Methods 89 (2015) 138–148

Contents lists available at ScienceDirect

Methods

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

Analysis of protein–RNA interactions in CRISPR proteins and effector complexes by UV-induced cross-linking and mass spectrometry

Kundan Sharma [a], Ajla Hrle [b], Katharina Kramer [a,k], Timo Sachsenberg [c,d], Raymond H.J. Staals [e], Lennart Randau [f], Anita Marchfelder [g], John van der Oost [e], Oliver Kohlbacher [c,d,h,i], Elena Conti [b], Henning Urlaub [a,j,*]

[a] Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
[b] Structural Cell Biology Department, Max Planck Institute for Biochemistry, Martinsried, Germany
[c] Center for Bioinformatics, University of Tübingen, Tübingen, Germany
[d] Department of Computer Science, University of Tübingen, Tübingen, Germany
[e] Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands
[f] Prokaryotic Small RNA Biology, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany
[g] Department of Biology II, Ulm University, Ulm, Germany
[h] Quantitative Biology Center, University of Tübingen, Tübingen, Germany
[i] Faculty of Medicine, University of Tübingen, Tübingen, Germany
[j] Bioanalytics Research Group, Department of Clinical Chemistry, University Medical Center, Göttingen, Germany
[k] Plant Proteomics Group, Max Planck Institute for Plant Breeding Research, Cologne, Germany

ARTICLE INFO

Article history:
Received 16 February 2015
Received in revised form 19 May 2015
Accepted 4 June 2015
Available online 10 June 2015

Keywords:
Protein–RNA interactions
UV cross-linking
CRISPR-Cas
Cas7
Mass spectrometry

ABSTRACT

Ribonucleoprotein (RNP) complexes play important roles in the cell by mediating basic cellular processes, including gene expression and its regulation. Understanding the molecular details of these processes requires the identification and characterization of protein–RNA interactions. Over the years various approaches have been used to investigate these interactions, including computational analyses to look for RNA binding domains, gel-shift mobility assays on recombinant and mutant proteins as well as co-crystallization and NMR studies for structure elucidation. Here we report a more specialized and direct approach using UV-induced cross-linking coupled with mass spectrometry. This approach permits the identification of cross-linked peptides and RNA moieties and can also pin-point exact RNA contact sites within the protein. The power of this method is illustrated by the application to different single- and multi-subunit RNP complexes belonging to the prokaryotic adaptive immune system, CRISPR-Cas (CRISPR: clustered regularly interspaced short palindromic repeats; Cas: CRISPR associated). In particular, we identified the RNA-binding sites within three Cas7 protein homologs and mapped the cross-linking results to reveal structurally conserved Cas7 – RNA binding interfaces. These results demonstrate the strong potential of UV-induced cross-linking coupled with mass spectrometry analysis to identify RNA interaction sites on the RNA binding proteins.

1. Introduction

In a cell, RNA molecules almost invariably function in association with proteins. Since RNA molecules can have enzymatic activity, and are structurally more versatile than double-stranded DNA, the variety and numbers of proteins binding to RNA is significantly greater than those found associated with classical double-stranded DNA. Accordingly, a multitude of RNA-binding proteins (RBPs) have been described in prokaryotes and eukaryotes [1,2]. RNA binding by these proteins is versatile and is mediated by many different RNA-binding domains (RBDs), which can occur in various combinations within one RBP. In contrast, DNA-binding proteins such as transcription factors reveal only a very moderate variation in their DNA binding motifs.

Proteins that bind to RNA can modulate or stabilize RNA structures, thereby making RNA catalytically active and also mediate interactions between RNA and other macromolecules [3]. Conversely, RNA molecules can guide catalytically active proteins to their destinations. Furthermore – like the vast majority of proteins in higher eukaryotes, which are organized in protein complexes – RBPs with their cognate RNAs also serve as assembly

* Corresponding author at: Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany.

http://dx.doi.org/10.1016/j.ymeth.2015.06.005
1046-2023/

platforms for proteins, while also being able to prevent proteins from interacting with the RNA. Thus RBPs are often, if not always, organized in ribonucleoprotein complexes (RNPs) [1]. These play essential roles in the major cellular steps of gene expression and its regulation. Hence, there is major interest in the molecular characterization of RNA-binding proteins with clear emphasis on identifying putative RNA-binding sites, as these regions are often essential for a functional RNP.

The "gold standard" for characterizing molecular interactions of RBDs with their cognate RNA molecules by structure determination is co-crystallization [4,5]; others include NMR of the complex [6], or high-resolution EM of entire RNPs, as performed for the ribosome [7]. Although the number of co-structures of RBPs has been steadily increasing with more than 200 co-structures of protein–RNA complexes available in the PDB, most RBPs are still crystallized without RNA. Consequently, the molecular charac terization of the RBD requires mutation studies combined with definition of the surface charge of the protein to allow localization of the RBD. Similarly, perturbations in the chemical shift of amino acid residues in NMR that are caused by interaction with RNA can allow the localization of the RBDs [8].

In recent years, chemical protein–protein cross-linking and UV-induced protein–nucleic acid cross-linking, in combination with mass spectrometry, have emerged as complementary methods for obtaining information about the spatial arrangement of proteins in complexes and in RNPs [9,10]. In the case of UV-induced protein–RNA cross-linking, MS has been applied to identify the cross-linked proteins by standard quantitative MS-based proteomic approaches [11–13]. Subsequent database-searching has led to the identification of conserved structural motifs in these proteins [2], such as RNA-recognition motifs (RRMs) [14], K homology (KH) domains [15], zinc-finger domains [16], tudor domains [17], double-stranded RNA binding domains (dsRBDs) [18], G-patch domains [19], Sm motifs [20] etc. However, such proteomic approaches yield little or no information about (i) whether the protein cross-links to the RNA through its canonical RBD or through other domains within the protein; (ii) which RBD is involved in interaction with RNA when the proteins contains several potential RBDs; (iii) how proteins that do not harbor any known RBD (as identified by sequence) interact with RNA.

The latter situation occurs very often when prokaryotic RNA-binding proteins are investigated. These do not show primary RNA-binding sequence motifs that resemble those of eukaryotic proteins. Nonetheless, three-dimensional structures of bacterial RBPs are similar to structures of eukaryotic RBDs, for example, the bacterial HfQ protein with the characteristic Sm fold [21,22] and the prokaryotic Cas7 protein family with their RRM motifs [23,24].

We have now developed a straightforward approach that utilizes UV-induced cross-linking and mass spectrometry, not only to identify proteins that cross-link to RNA but also to identify unambiguously the cross-linked amino-acid and the cross-linked nucleotide(s) [25]. The approach is easily applicable to single (e.g., recombinant) proteins that interact with RNA but whose structure cannot be determined in complex with RNA. In contrast to other approaches, it can be also applied to assembled RNPs of any complexity, obtained either by reconstitution or by purification from extracts. Importantly, it can even be applied at the level of entire UV-cross-linked cells.

Here we describe the method for applying this approach to single recombinant proteins bound to RNA in detail. The proteins described here belong to the recently discovered prokaryotic adaptive immune defense system CRISPR-Cas [26]. In this system Cas proteins are guided by a CRISPR RNA (crRNA) to target and degrade complementary foreign nucleic acids in a manner that is functionally reminiscent of the eukaryotic RNA interference mechanism [27]. Type I, II and III CRISPR-Cas systems are classified based on their signature Cas genes (*cas3*, *cas9* and *cas10* respectively) that are further classified into different subtypes based on the presence of other Cas genes [28]. Type I and subtypes III-A and III-B form multiprotein RNPs together with different Cas proteins in addition to Cas3 or Cas10. Type II contains mainly one Cas protein, Cas9, and generates an RNP with two different RNA molecules (crRNA and tracrRNA). Some Cas proteins comprise nuclease domains, distinct helicase domains and also RRM domains that are typical for RNA-binding proteins [29]. The Cas7 family proteins, which form the backbone of the surveillance and effector complexes in Type I and Type III systems, consist of RRMs and belong to the RAMP (repeat associated mysterious proteins) superfamily [28]. Interestingly, most Cas proteins lack conserved amino-acid residues that account for RNA interaction. The diverse peripheral domains of the Cas protein family thus mediate RNA binding.

The Cas proteins that we use to demonstrate our approach are: Type I-A Cas7 from *Thermoproteus tenax*; Type I-D Cas7 from *Thermofilum pendens*; and Type III-A Cas7 (Csm3) from *Thermus thermophilus*. These homologs belonging to the Cas7 protein family were not co-crystallized with their cognate crRNAs. The investigations shown here in detail for Csm3 from *T. thermophilus* derived from a recent study of the fully assembled CRISPR-Cas Type III-A Csm complex in which we mapped protein–RNA cross-linking sites on all the proteins within this complex [30].

2. Experimental procedures

Below we give a detailed protocol for the investigation of the molecular interaction of recombinant RNA-binding proteins with their (cognate) RNA oligonucleotides and of endogenous protein–RNA complexes isolated from prokaryotic cells using UV-induced cross-linking. The protocol allows the mapping of UV cross-linking sites between proteins and RNA at single amino acid and nucleotide resolution. The principle of this approach is that after UV-induced cross-linking of amino acid side chains within a protein to the nucleobases of an RNA the cross-linked region and the cross-linked amino acid of the protein are identified by high-resolution mass spectrometry. Mass spectrometry enables one to read the sequence of the cross-linked peptide and the composition (but not the sequence) of the cross-linked RNA. It also allows the identification of the cross-linked amino acid in cases where the spectrum is of sufficient quality (see Section 3.1). The principle behind the approach is that the RNA and the protein of interest are completely digested with endonucleases and proteases, then the cross-linked peptide–RNA oligonucleotides are separated from the non-cross-linked RNA oligonucleotides and peptides. These purified heteroconjugates are subjected to MS. The database search, performed to identify the cross-linked peptide region with its cross-linked nucleotides, is as important as the entire purification procedure, since it differs from the searches typically performed for modified peptides. However, in this article only the principle of the modified database search is described, and we refer to a more detailed description of the database search of raw MS data in a recent publication [25]. The step-by-step description of the workflow includes sample preparation, UV-induced protein–RNA cross-linking, endoproteinase and nuclease digestion of proteins and RNAs, enrichment of peptide–RNA oligonucleotide cross-links, liquid chromatography (LC)–coupled electrospray ionization (ESI) tandem mass spectrometry (MS/MS) analysis and database search of raw MS data. An overview of the workflow is provided in Fig. 1. Any protein–RNA complex can be used for the sample preparation described below. The reconstitution conditions depend on the particular nature of the protein(s) and their cognate RNAs. Isolated endogenous or reconstituted protein–RNA complexes that contain more than one protein can also be used.

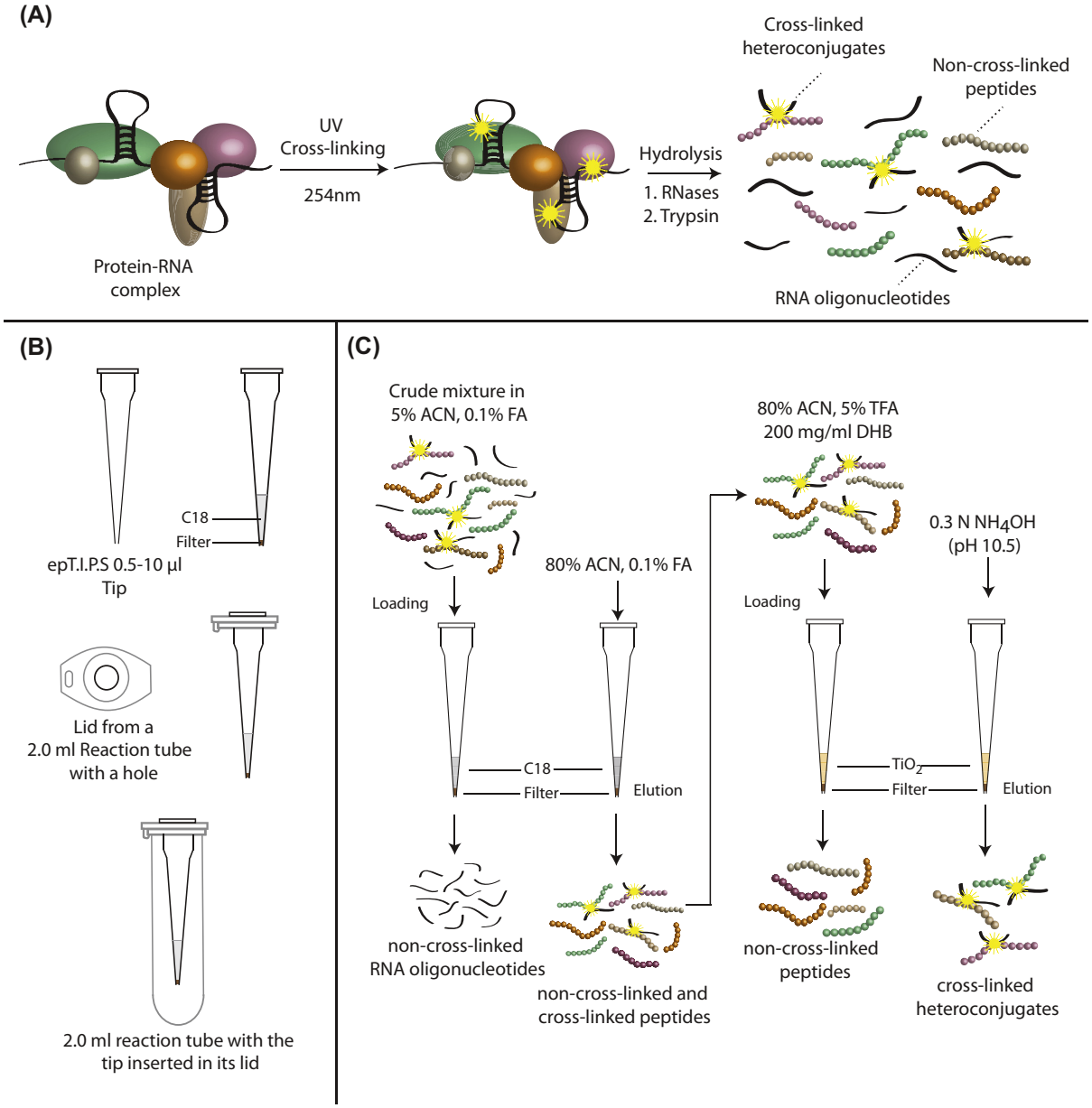

Fig. 1. Overview of the UV-induced protein–RNA cross-linking and purification and enrichment of cross-linked heteroconjugates. (A) Protein–RNA complex is UV-irradiated at 254 nm and hydrolyzed by RNases and trypsin resulting in a crude mixture of cross-linked and non-cross-linked peptides and RNA fragments. (B) Preparation of a C18 column for the C18 reversed-phase chromatography. (C) A schematic representation for C18 reversed phase chromatography and TiO_2 enrichment for removal of non-cross-linked RNA oligonucleotides and non-cross-linked peptides as described in Sections 2.5.1 and 2.5.2.

2.1. Sample preparation

The following samples were used in this study: (1) recombinant *T. pendens* Cas7 (Csc2) protein incubated with a synthetic $polyU_{(20)}$; (2) recombinant *T. tenax* Cas7 incubated with a synthetic $polyU_{(20)}$; (3) endogenous multi-protein–RNA complex, Type III-A Csm complex from *T. thermophilus* comprising proteins Csm1 to Csm5 assembled around their cognate crRNA.

Recombinant Cas7 proteins from *T. pendens* and *T. tenax* were cloned and expressed as described elsewhere [24,31] and the $polyU_{(20)}$ RNA was synthesized by Purimex. For cross-linking with $polyU_{(20)}$, 1 nmol protein was mixed with 1 nmol $polyU_{(20)}$ in a total volume of 200 µl in a buffer containing 20 mM HEPES (pH 7.5), 10 mM NaCl, 4 mM $MgCl_2$ and 2 mM DTT. This was followed by 15 min incubation at 50 °C. Buffers should not contain reagents that might act as radical scavengers, such as glycerol. Note that DTT acts as a protein–RNA cross-linker under UV irradiation, as it reacts with the uridine base and with cysteine and generates a spacer between cysteine and uridines. This reaction is strictly UV-dependent ([25]; U.Z. and H.U., unpublished data). To avoid this, TCEP can be used instead of DTT. The endogenous Type III-A Csm complex from *T. thermophilus* was purified as described elsewhere [30]. For cross-linking 2 nmol of the complex were used in a total volume of 200 µl buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl with 10 min incubation at 65 °C. The temperatures were based on the previous *in vitro* experiments performed with these complexes.

After complex formation, half of the sample is subjected to UV-induced cross-linking (see Step 2.2). The other half of sample is kept as a non-cross-linked control. All the steps described in Sections 2.3–2.7 are performed with both the cross-linked and the non-cross-linked samples.

2.2. UV-induced protein–RNA cross-linking

An apparatus built in-house was used for UV-induced cross-linking. It was equipped with four 8 W lamps (dimensions 1.5 cm × 28.5 cm; wavelength 254 nm; G8T5, Sankyo Denki, Japan) as described elsewhere [32]. Alternatively, a UV stratalinker 2400 from Stratagene can be used.

The protocol for UV irradiation is as follows:

1. The UV lamp apparatus is placed in a cold-room (4 °C) and switched on at least 30 min before the cross-linking experiment to achieve constant UV intensity.
2. The sample is transferred to a black polypropylene microtiter plate (Greiner Bio-One); aliquots of 100 μl are placed in each well, and the plate is placed on an ice-cold metal (aluminum) block (see [32] for details).
3. The plate is then positioned under the UV lamps at a distance of about 1 cm.
4. The sample is irradiated for 10 min (maximum) and then transferred back into a 1.5 ml reaction tube (Eppendorf Safe Lock Tubes).

The choice of UV irradiation times at 254 nm should be evaluated by incubation of the protein of interest with ^{32}P-labeled RNA and subsequent loading of the cross-linked sample onto SDS–PAGE [33]. A control with UV-irradiated ^{32}P-labeled RNA only is recommended. The radioactively labeled band on the SDS–PAGE will (i) prove the capability of the protein to cross-link to RNA under these conditions; (ii) reveal any protein degradation upon UV-irradiation; and (iii) will reveal the optimal cross-linking yield upon UV irradiation at different times. In general, when working with recombinant proteins and RNA oligonucleotides, we found that UV-irradiation times of 10 min at 254 nm lead to the cross-linking highest yield with no loss of protein by UV-induced hydrolysis [33]. Of note, when one is working with endogenous, i.e. *ex vivo* protein–RNA complexes that are isolated from cells and which contain a larger RNA moiety, irradiation times of max. 2 min are recommended [25,34]. Irradiation at longer wavelength, e.g. at 365 nm when 4-thio-uridine-substituted RNA is used, can be prolonged to 30 min, as no significant damage to the substituted RNA is observed [35].

2.3. Ethanol precipitation

This and all subsequent steps (up to and including 2.7) are carried out on the irradiated and the non-irradiated samples in parallel.

To purify and concentrate the samples before endoprotease and nuclease digestion, they are precipitated with ethanol as follows:

1. Three volumes of chilled (−20 °C) ethanol (Merck, Darmstadt, Germany) and 1/10 volume of 3 M NaOAc at pH 5.2 are added to the sample. Followed by incubation at −20 °C for at least 2 h.
2. The precipitated sample is pelleted by centrifugation (Heraeus Fresco 17 centrifuge, Thermo Fisher Scientific) at 13,000 rpm and 4 °C for 30 min.
3. The supernatant is removed and the pellet is washed with two volumes of ice-cold 80% (v/v) ethanol in water (LiChrosolv, Merck, Darmstadt, Germany) by brief shaking with a Vortex Genie 2 (Scientific Industries).
4. Centrifugation is performed again as above. Finally, the supernatant is carefully removed and the pellet is dried in a SpeedVac (Eppendorf concentrator 5301) for a maximum of 5 min.

2.4. Hydrolysis of protein and RNA

The first step in the isolation of cross-linked peptide–RNA oligonucleotide for subsequent LC–MS/MS analysis is the endoproteolytic and nucleolytic digestion of the protein and RNA moieties under denaturing conditions. The yield of peptide–RNA oligonucleotides depends not only on the UV cross-linking yield [25] but also on the efficiency of the digestion of proteins and RNA. When working with protein–RNA complexes that harbor a relatively short RNA molecule, the conditions (including the buffer) should be chosen such as to allow the digestion of both components in a single step without any change in buffer solutions. When investigating protein–RNA complexes with an RNA that is physically larger than the peptides that are generated by endoproteolytic cleavage of the protein moiety (e.g. (pre)-mRNA, (pre)-rRNA, lncRNA, etc.), the proteins and RNA should be digested successively, in a two-step reaction that includes enrichment of the intact RNA after proteolysis and before nuclease digestion using size exclusion chromatography. However, the latter strategy will not be described here and we refer to references [25,36] for a detailed description.

To achieve digestion of proteins and RNA the precipitated sample is dissolved in buffer containing at least 4 M urea. Note that a higher concentration of urea (maximum 8 M and optionally supplemented with 2 M thiourea) typically achieves a more complete dissociation and denaturation of the protein–RNA complex. However, for the final digestion with endoproteinase trypsin (see below), the urea concentration should be reduced to 1 M, so that the sample volume increases by the factor of at least four. This in turn might result in a relatively high sample volume for the first enrichment steps that remove non-cross-linked RNA oligonucleotides (see Step 2.5.1). The RNA moiety is hydrolyzed by using ribonucleases T1 and A. Neither nuclease cuts double-stranded RNA, so it should be ensured that the RNA moiety is completely denatured and unfolded before digestion. In addition, the nuclease benzonase may be used. Benzonase digests single- and double-stranded RNA as well as DNA in a highly unspecific manner. The advantage of using benzonase is that it generates very short RNA moieties (mainly mono- and dinucleotides) that are still cross-linked to the peptides. We note that, for a mass-spectrometric analysis under the conditions described here, the cross-linked RNA moiety should be as small as possible in order to obtain high-quality MS/MS (fragment spectra) of the cross-linked peptide moiety [34,37].

Larger RNA cross-linked oligonucleotides generated by digestion with e.g. only RNase T1 (which cuts exclusively 3′ to G) lead to very intense RNA product ions in gas-phase fragmentation in the mass spectrometer. These suppress the fragment ions derived from the cross-linked peptide, so that the peptide sequence can hardly be determined under these conditions in the mass spectrometer [38].

Of the endoproteinases, trypsin is the most widely used in MS-based proteomics. Some proteomic studies use a first endoproteolytic cleavage step with the enzyme Lys-C, which is still active at higher urea concentrations such as 4 M [39] followed by a second digestion step with trypsin.

The steps for RNA and protein hydrolysis of cross-linked protein–RNA complexes are as follows:

1. The pellet obtained after the ethanol precipitation (Step 2.3.4) is dissolved in 50 μl 4 M urea in 50 mM Tris–HCl, pH 7.9.
2. After resuspension, the urea concentration is adjusted to 1 M by addition of 150 μl 50 mM Tris–HCl, pH 7.9.
3. The RNA is digested by using 1 μl each of RNase A (1 μg/μl) and T1 (1 U/μl) (both from Ambion), followed by incubation at 52 °C for 2 h.

4. Alternatively, digestion is performed with benzonase instead of – or in addition to – RNases A and T1. For this, the sample is supplemented with 2 µl 100 mM $MgCl_2$ to a concentration of 1 mM $MgCl_2$; thereafter 1 µl benzonase (25 U/µl) (Novagen, Merck, Darmstadt, Germany) is added and the sample is incubated for 1 h at 37 °C.
5. After digestion of the RNA moiety, trypsin (Promega) is added in a protein-to-enzyme ratio of 20:1 (w/w) followed by overnight incubation at 37 °C. The calculation of the protein–enzyme ratio is based on the starting amount of recombinant protein or protein–RNA complex (see Step 2.1).
6. After digestion, 10 µl 100% acetonitrile (ACN) and 2 µl 10% (v/v) formic acid (FA) in water are added to the sample to give a final concentration of 5% (v/v) ACN and 0.1% (v/v) FA. The sample is dissolved by brief vortexing and sonication for 1 min.

2.5. Enrichment of cross-linked peptide–RNA oligonucleotides

UV-induced cross-linking between proteins and RNA is a radical-induced reaction with relatively low yields [40]. Therefore, one essential step is enrichment of cross-linked species, i.e., cross-linked peptide–RNA oligonucleotides from the complex mixture obtained after digestion of protein–RNA complexes; this mixture consists mainly of non-cross-linked peptides and RNA oligonucleotides. In an LC-coupled MS analysis such non-cross-linked species will interfere drastically with the detection of the (much less abundant) cross-linked species. Consequently, two purification steps are needed to remove non-cross-linked oligonucleotides and peptides, to enrich the cross-linked species to a level above that of any residual non-cross-linked species.

2.5.1. Removal of non-cross-linked RNA oligonucleotides by C18 reversed-phase chromatography

Non-cross-linked RNA oligonucleotides are removed from the mixture by C18 reversed-phase chromatography (Fig. 1C). Small RNA oligonucleotides present in the sample after RNA hydrolysis do not bind to the C18 material, whereas the peptides (both cross-linked and non-cross-linked) have a strong affinity towards the C18 material. For this purpose a C18 column (AQ 120 Å 5 µM, Dr. Maisch GmbH) packed in-house is used. The column consists of a pipette tip (epT.I.P.S 0.5–10 µl; Eppendorf) in which a 2 mm^2 piece of standard coffee filter is fitted into the very end of the tip. The filter paper serves a permeable plug that retains the column material, but not the sample, during loading and elution. To prepare slurry, 20 mg C18 matrix is suspended in 100 µl 100% (v/v) methanol (LiChrosolv, Merck, Darmstadt, Germany) and the slurry is filled into the pipette tip to a height of 3–5 mm. The pipette tip is then inserted into a punched hole of a lid of a 2.0 ml reaction tube (Eppendorf Safe Lock Tubes) as shown in Fig. 1B. A regular screwdriver or a similar device can be used to punch a hole in the lid of the reaction tube to fit the spin column.

Column equilibration, sample loading, washing and elution are performed with centrifugation steps at 5000 rpm (Heraeus Biofuge pico, Thermo Fisher Scientific) for 5 min each. Closing the lid of the rotor does not physically interfere with the spin column. Nonetheless, the lid of the rotor might be removed in this low-speed centrifugation step. For Steps 1–4 below the flow-throughs are collected in separate 2.0 ml reaction tubes.

The details for the individual steps are as follows:

1. The packed column is equilibrated successively with 60 µl of 95% (v/v) ACN, 0.1% (v/v) FA in water (ACN and water, LiChrosolv, Merck, Darmstadt, Germany; FA, Sigma–Aldrich), 60 µl of 80% (v/v) ACN, 0.1% (v/v) FA in water, 60 µl of 50% (v/v) ACN, 0.1% (v/v) FA in water, and 60 µl of 0.1% (v/v) FA in water.
2. The hydrolyzed sample (see Step 2.4.5 above) is then loaded onto the column in 60 µl aliquots.
3. After centrifugation, the sample retained on the column is washed twice with 60 µl 0.1% (v/v) FA in water.
4. Elution of the sample is performed in three steps, first twice with 60 µl 50% (v/v) ACN, 0.1% (v/v) FA in water and the third time with 60 µl 80% (v/v) ACN, 0.1% (v/v) FA in water. The eluates from all three steps are pooled in a single 1.5 ml reaction tube (Eppendorf Safe Lock Tube).
5. The eluted sample is dried in a SpeedVac until all the solvent has been removed.

2.5.2. Removal of non-cross-linked peptides using TiO_2 enrichment

After removal of the non-cross-linked RNA oligonucleotides, the dried sample consists mainly of non-cross-linked peptides, cross-linked peptide–RNA oligonucleotides and residual non-cross-linked RNA oligonucleotides.

To remove non-cross-linked peptides and enrich peptide–RNA oligonucleotides, a matrix is required that makes use of the physicochemical properties of the cross-linked RNA moiety. Titanium dioxide (TiO_2) chromatography has been established as a method for enrichment of phosphopeptides in MS-based proteomics [41,42]. The underlying principle can also be applied for enrichment of cross-linked peptide–RNA oligonucleotides over the majority of non-cross-linked and cross-linked peptides (Fig. 1C). The TiO_2 (Titansphere 5 µM, GL Sciences) columns are packed in pipette tips similar to the C18 columns (Step 2.5.1) with a coffee filter plug at the very end of the tip. 20 mg TiO_2 material is suspended in 100 µl 80% (v/v) ACN in water containing 0.1% (v/v) trifluoroacetic acid (TFA) (Roth) in water and added to the column as described for Step 2.5.1. All the centrifugation steps for column equilibration, sample loading, washing and elution are performed with centrifugation at 3000 rpm (Heraeus Biofuge pico, Thermo Fischer Scientific) for 5 min each. For steps 1–5 below the flow-throughs are collected in separate 2.0 ml reaction tubes.

The details for the individual steps are as follows:

1. The dried pellet from the C18 chromatography (Step 2.5.1.5) is dissolved in 100 µl buffer A by vortexing and 1 min sonication. Buffer A consists of 200 mg/ml 2,5-dihydroxybenzoic acid (Sigma–Aldrich) in buffer B (80% (v/v) ACN, 5% (v/v) TFA in water).
2. The TiO_2 column is washed twice with 60 µl buffer B.
3. The sample is loaded onto the column in 50 µl aliquots.
4. The column is washed three times with 60 µl buffer A to eliminate non-cross-linked peptides and five times with 60 µl buffer B to remove any residual DHB.
5. The sample is eluted by applying 40 µl 0.3 N NH_4OH, pH 10.5, three times. The eluates are pooled in a 1.5 ml reaction tube (see Step 2.5.1.4).
6. The eluate is dried in a SpeedVac until the solvent has been entirely removed.

2.6. Mass spectrometry analysis

The last practical step in the analysis of protein–RNA binding sites after UV-induced cross-linking of protein–RNA complexes is the MS analysis of the purified peptide–RNA oligonucleotide cross-links. This analysis allows sequencing the cross-linked peptide and RNA moieties in the gas phase of the mass spectrometer. In this way, not only the amino-acid sequence of the cross-linked oligopeptide is determined, but also the cross-linked amino acid is identified. The cross-linked nucleotide is determined by calculating the mass difference between the entire mass of the cross-linked species and the mass of the cross-linked peptide. In addition, marker ions of the cross-linked nucleotides in the lower

m/z regime of the fragment spectrum (C = 306.0491, U = 307.0331, A = 330.0603, G = 346.0553 and bases (C′ 112.0511, U′ = 113.0351, A′ = 136.0623, G′ = 152.0572) are taken into account. Some of these marker ions are indicated in Fig. 2. Previous analyses have made use of matrix-assisted laser desorption/ionization (MALDI) mass spectrometry to analyze peptide–RNA oligonucleotide cross-links

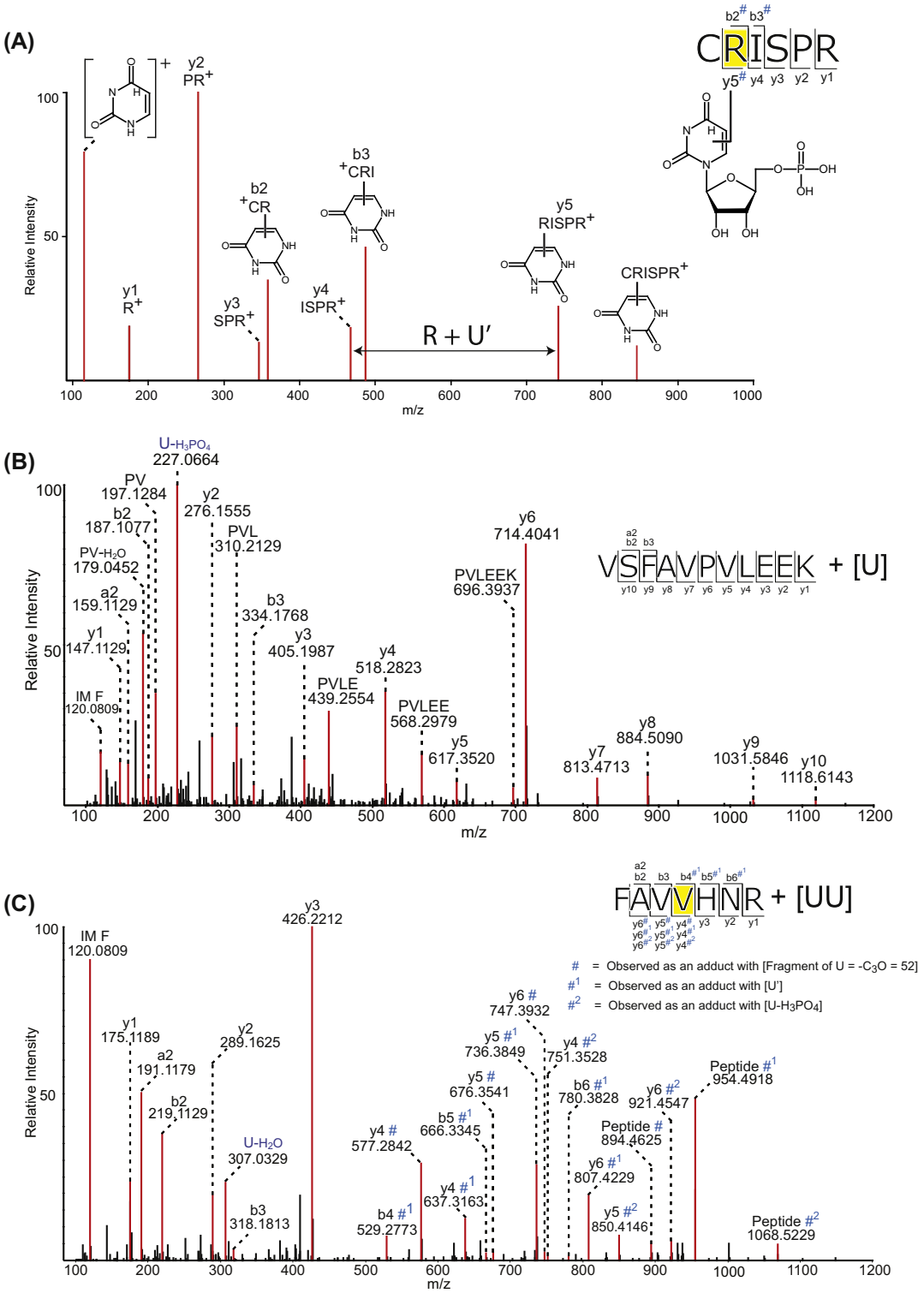

Fig. 2. MS/MS spectra of cross-linked species. (A) Sample spectrum of a peptide 'CRISPR' cross-linked to a uracil nucleotide to indicate the characteristic peaks and shifts observed during fragmentation of a peptide–RNA cross-link. Distinct fragment ions containing nucleic acid base and peptide fragment are indicated. All fragments containing the cross-linked residue are shifted by the mass of uracil when compared to regular peptide fragments. (B) MS/MS fragmentation of the *T. tenax* Cas7 peptide ^{127}VSFAVPVLEEK^{137} cross-linked to a uracil nucleotide, lacking any b- and y-ions with a mass-shift that could indicate exactly which amino acid is cross-linked. (C) MS/MS fragmentation of the *T. tenax* Cas7 peptide ^{145}FAVVHNR^{151} cross-linked to UU dinucleotide, with a clear mass-shift indicating the V^{148} as the cross-linked amino acid. The peptide sequence and fragment ions are indicated on the top and cross-linked residues are highlighted in yellow. Annotated fragment ions of the cross-linked peptide are shown in red. Some of the b- and y- ions were observed with a mass shift of #, #1, and #2 corresponding to –C3O (a fragment of uracil), U′ and U-H_3PO_4 respectively. IM: immonium ions, U′: U marker ion adduct of 112.0273 Da.

[37,43,44]. Currently, electrospray ionization (ESI) MS coupled to a nano-liquid chromatography (LC) is the method of choice for analyzing such cross-links. The advantages over MALDI is that: (i) it can be directly coupled to chromatography systems, which results in a significant shorter analysis time, (ii) the fragment-ion-based sequence information obtained from the cross-linked peptide (e.g. y-type and/or b-type product ions) is more comprehensive than the information from a similar MALDI-MS/MS analysis, so that the cross-linked peptide moiety is more readily identified in a subsequent database search, (iii) the data analysis software has been developed for ESI-MS data and helpful filtering steps are based on comparisons of chromatographic peaks and would not be available for MALDI without extensive redesign.

In the experiments described here, the UV-cross-linked peptide–RNA heteroconjugates were analyzed by LC–MS/MS with an LTQ Orbitrap Velos instrument (Thermo Fisher Scientific) coupled to a nano-LC system (Agilent 1100 series, Agilent Technologies) equipped with C18 trapping column of ~2 cm length and 150 μm inner diameter, in-line with a C18 analytical column of ~15 cm length and 75 μm inner diameter. Both columns were packed in-house, with C18 AQ 120 Å 5 μm material (Dr. Maisch GmbH).

We note that any nano-LC–ESI-MS setup (independent of vendors) can be used for the analysis of peptide–RNA oligonucleotide cross-links. It should be kept in mind that the more accurate the mass spectrometric analysis is – i.e. in determination of the precise masses of the intact cross-linked species (the so-called precursor) and the product ions (e.g. y- and b-type ions derived from the sequencing of the cross-linked peptide and nucleotide moieties) – the better the data analysis in terms of fewer false positive hits in the subsequent database search. We therefore recommend performing analysis only on high-resolution MS instruments that deliver a high mass accuracy (⩽10 ppm). Since the MS analysis is performed in the so-called data-dependent acquisition (DDA) mode, the data acquisition speed of the mass spectrometer is a critical factor as well. In DDA an initial MS scan over a specific mass range detects m/z of all species eluting at that particular point, of which the precursor ions with the most intense signals are selected for fragmentation in subsequent MS/MS scans. Accordingly, the more precursors are selected and sequenced within a certain time, the more comprehensive is the analysis, species with lower intensities are also selected and sequenced. We further note that gas-phase fragmentation of the cross-link in 3D or linear ion traps is not recommended, as the fragment spectra do not have sufficient quality to assign marker ions in the lower m/z range as well as to unambiguously correlate production ion peaks with theoretical (e.g., b- and y-type) ions of the sequence. Fragmentation should be performed in the quadrupole or hexapole of the mass spectrometer.

The following MS instruments are suitable for such an analysis: qQ-TOF instruments from AB Sciex, Agilent technologies, Bruker and Waters companies and Orbitrap instruments from Thermo Fisher Scientific company that work in HCD mode with sufficient sensitivity (Orbitrap Velos and Elite, Q-Exactive instruments, Orbitrap Fusion instrument). The ESI 3D and linear iontrap mass spectrometers are not adequate.

The LC system should (i) allow in-line (i.e., in a row) set-up of the pre-column and the analytical column, (ii) allow the generation of a stable nano-flow, i.e. 100–300 nL/min, and (iii) leave enough freedom for the operator to program various sample-loading times on the trapping column, washing times and elution times, so that a system consisting of a loading pump (for higher flow rates) and two gradient pumps (nano-flow) is beneficial. In principal all nano-LC systems used for MS-based proteomic approaches are suitable.

In summary, the LC–ESI-MS/MS protocol is as follows:

1. The dried samples obtained after the TiO_2 enrichment (Step 2.5.2) are dissolved in 2 μl 50% (v/v) ACN, 0.1% (v/v) FA in water and diluted to a final concentration of 10% (v/v) ACN, 0.1% (v/v) FA in water by the addition of 10 μl 0.1% (v/v) FA in water.
2. 5 μl of sample is loaded on the trapping column over 5 min at a flow rate of 10 μl/min in buffer A (0.1% (v/v) FA in water).
3. The sample is eluted and separated on the analytical column with a gradient of 7–38% buffer B (95% (v/v) ACN in water, 0.1% (v/v) FA in water) over 33 min (0.87%/min) at a flow rate of 300 nL/min.
4. The mass spectrometer (LTQ Orbitrap Velos) is operated in a data-dependent acquisition mode using TOP 10 method. MS1 is recorded in the m/z range of 350–1600 at a resolution of 30,000 and for subsequent MS/MS the ten most intense ions are selected. Fragment ions are generated by HCD activation (high energy collision dissociation, normalized collision energy = 40), and recorded with a fixed first mass of m/z = 100 and a resolution of 7500. Both precursor ions and fragment ions are scanned in the orbitrap analyzer and the resulting spectra are measured with high accuracy in both the MS and the MS/MS level.

2.7. Data analysis

The experimental workflow described above, with its last step of the mass-spectrometric analysis, results in two mass spectrometric data files (.raw) per experiment, one for the UV-cross-linked sample and the other for the non-cross-linked control.

The mass-spectrometric data analysis is automated and implemented into a workflow that is based on OpenMS software [45,46] in combination with the freely available search engine OMSSA [47]. An extended description of the database analysis is available in [25], and a step-by-step tutorial is available in the supplementary files of that reference; the tutorial explains in detail how to prepare raw mass-spectrometric files for the dedicated database search.

In brief, the principle of the workflow is as follows: When state-of-the-art MS instruments are used, the DDA mode selects a very large number of precursors, from which corresponding fragment spectra are generated. The mass information from the precursor and fragment ions is stored in the raw data. However, the large number of spectra cannot be evaluated manually, and we

Table 1
List of RNA contacting regions and cross-linked amino acids identified in the four Cas7 proteins.

Protein	Peptide sequence	Cross-linked amino acid
T. tenax Cas7	^{3}VAPPYVR^{9}	Y^{7}
	^{14}FEAQLSVLTGAGNMGNYNMHAVAK^{37}	G^{28}
	^{127}VSFAVPVLEEK^{137}[a]	–
	^{145}FAVVHNR^{151}[a]	V^{148}
	^{152}VDPFKR^{157}	F^{155}
	^{163}SKEEQEGTEMMVFK^{176}	M^{172}
T. pendens Cas7	^{82}LMAVTR^{87}	M^{83}
	^{124}KVSEEWNCTIQPPLAEFGEK^{143}	C^{131}
	^{346}WVEELKGGGQK^{356}	W^{346}
T. thermophilus Csm3	^{21}IGMSRDQMAIGDLDNPVVR^{39}	–
	^{40}NPLTDEPYIPGSSLK^{54}	^{49}P–K^{54}
	^{91}IFGLAPENDER^{101}	P^{96}
	^{136}GGLYTEIKQEVFIPR^{150}	Q^{144}
	^{151}LGGNANPR^{158}	C^{153}
	^{159}TTERVPAGAR^{168}	R^{162}

[a] The MS/MS fragment spectra for the peptides ^{127}VSFAVPVLEEK^{137} cross-linked to a single uracil nucleotide and ^{145}FAVVHNR^{151} cross-linked to UU dinucleotide are given in Fig. 2B and C respectively. The cross-linking results for *T. pendens* Cas7 and *T. thermophilus* Csm3 have also been described earlier in [24,30]. All the cross-linked amino acids identified have been mapped on the Cas7 protein models and are illustrated in Fig. 3.

therefore developed a workflow that filters the data in order to keep only those MS spectra that are most likely to be derived from true positive peptide–RNA oligonucleotides. The filtered rawdata is finally used for database search. The steps for data analysis are based on our previously published work [25]; OpenMS software is used, with OMSSA as search engine. To apply this workflow with subsequent database search on the raw data that is generated by the manufacturer's instrument software, the raw files are converted into .mzML format [48] by using msconvert of ProteoWizard software [49]. The first step in the workflow is a conventional database search to identify residual non-cross-linked peptides and also non-cross-linked oligonucleotides. The latter can be deduced from the fractional masses which differ from those of peptides and cross-linked peptides [40]. Once these precursors have been discarded from the MS data, the precursor masses of putative cross-links between the non-UV-irradiated control and the UV-irradiated sample are compared. Precursors with identical masses and the same retention times in LC–MS/MS are deleted, as these cannot represent UV-induced peptide–RNA oligonucleotide cross-links (i.e., because they are present in both the irradiated and the non-irradiated sample). In the remaining set of spectra, a so-called precursor variant approach is applied [40]. Basically, as the composition and hence the mass of the cross-linked RNA are not known, several combinations of theoretical masses of putative RNA oligonucleotides are subtracted from each precursor mass. The result is an altered peaklist that contains the original and the newly calculated precursor masses plus the corresponding mass information of the fragment ions – the latter remains unaltered. The database search compares the precursor masses and their corresponding fragment masses with all theoretical precursor and fragment masses within the protein database. A hit for a cross-link will only appear when an altered precursor mass in the peaklist with its unaltered fragment masses (which contain the information of the peptide sequence, see above) matches up with a certain peptide precursor in the database. Consequently, the composition of the cross-linked RNA oligonucleotide is known, since the altered precursor masses were generated by subtracting the theoretical mass corresponding to a defined RNA composition. Any MS/MS fragment ion spectra of matched cross-links should be annotated manually for final

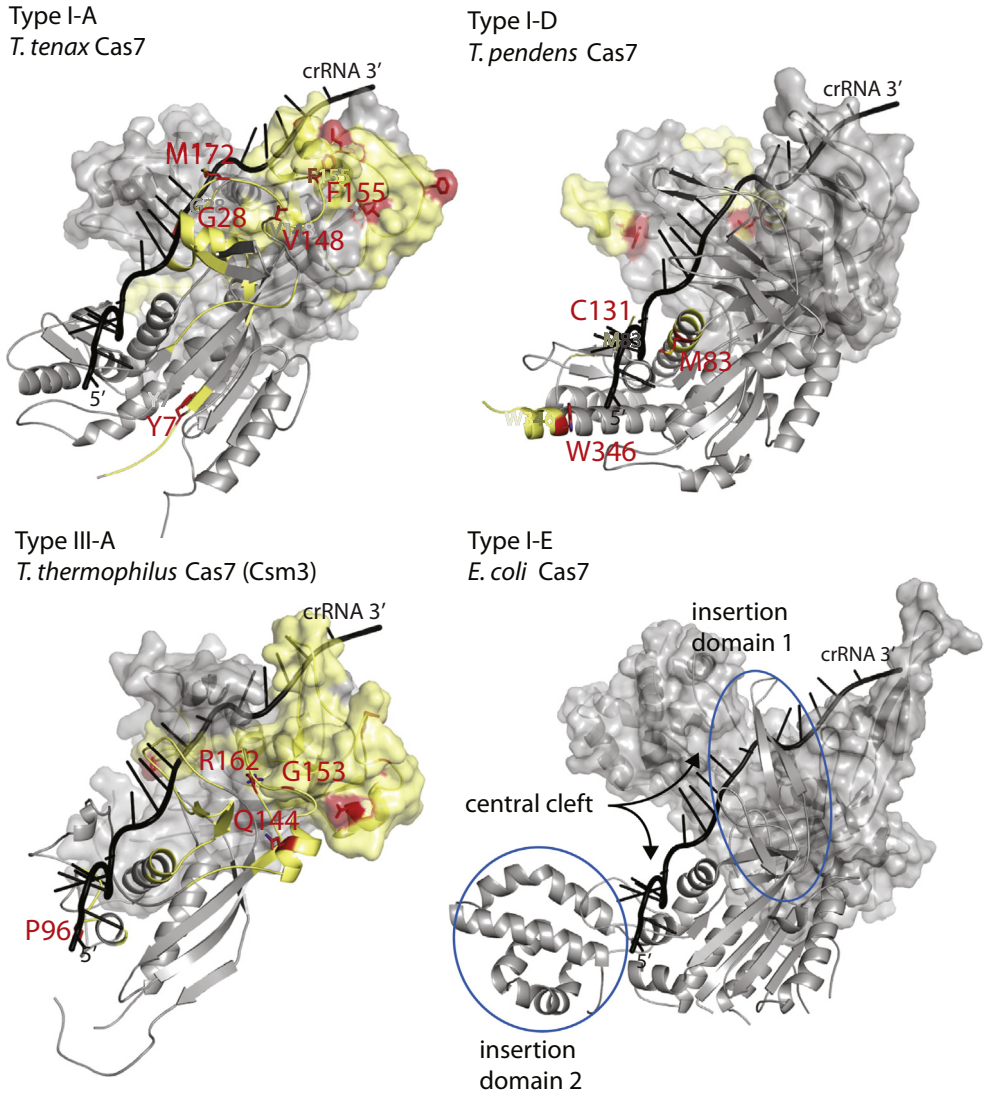

Fig. 3. Mapping the protein–RNA cross-link data to crystal and modeled structures. Predicted Cas7 3D-models (Type I-A and III-A) and the Type I-D Cas7 crystal structure (PDB ID: 4TXD) have been superposed to two copies of crRNA-bound Type I-E Cas7 (PDB ID: 1VY8). The front copy is shown in cartoon, the back copy additionally in surface representation. The crRNA is colored black, the protein structures are colored gray, the cross-linked peptide yellow and cross-linked sites are highlighted in red. The universally present central cleft, defined by the RRM and insertion domain 2 (blue circle), and the insertion domain 1 (blue circle) are labeled in Type I-E Cas7 for clarity.

confirmation of the peptide sequence and RNA composition. The tutorial on the use of the software [25] also includes some guidelines for the manual evaluation of the fragment spectra.

3. Mapping the RNA binding interface in Cas7 proteins

We applied the biochemical, mass spectrometric and computational workflow to map the RNA-binding sites within homologous Cas7 family proteins – *T. tenax* Cas7, *T. pendens* Cas7 and *T. thermophilus* Csm3 – bound to polyU and to crRNA. *In vivo*, several copies of Cas7 proteins are wrapped around crRNA in a sequence-unspecific helical fashion [5,30,50,51]. Crystal structures from single and complex-bound Cas7 proteins show two composite RNA-binding surfaces: a central cleft and a structurally variable insertion domain [5,23,24,52]. In all Cas7 proteins characterized to date both these domains are defined by insertions within the secondary structure elements of the central RRM domain (insertion domain 1 is β1-α1, β2-β3, α2-β4, and insertion domain 2 is α1-β2). Using polyU and crRNA substrates, we were able to point to the potential RNA interacting regions of three Cas7 family proteins: a to date uncharacterized Cas7 protein from *T. tenax* [31] and two structurally and functionally characterized homologs from *T. pendens* and *T. thermophilus* [24,30].

3.1. Identification of RNA interaction sites in the Cas7 family proteins of the CRISPR-Cas system

All three Cas7 homologs RNA interaction sites were identified by mass spectrometry with single amino-acid and single nucleotide resolution after UV cross-linking. The cross-links identified, with their cross-linked peptide sequence, are summarized in Table 1 for the three Cas7 homologs. Fig. 2A shows an example of an annotated HCD spectrum of a peptide 'CRISPR' with arginine as the amino-acid cross-linked to a uracil nucleotide. The characteristic feature of peptide–RNA oligonucleotide crosslinks are indicated in the annotated spectrum i.e., the b- and y- ion fragment series of the peptide, marker ion of uracil (base) and shifts in some of the b- and y- ions corresponding to the mass of an arginine residue identifying arginine as the cross-linked amino acid. In all cross-linked peptides the cross-linked amino acid could be determined in this manner, with the exception of two peptides (V^{127}–K^{137} in Cas7 from *T. tenax* and I^{21}–R^{39} in Csm3 from *T. thermophilus*). Here, no mass-shift in the b- or y- type fragment ions series of the cross-linked peptide could be identified. Fig. 2B and C show the fragment spectra of the two cross-linked peptides identified in *T. tenax* Cas7 when bound to polyU RNA. In the peptide encompassing positions F^{145}–R^{151}, V^{148} could be identified as a cross-linked amino acid, whereas in the peptide encompassing positions V^{127}–K^{137} the cross-linked amino acid could not be identified.

3.2. Cross-link sites on the structural model of Cas7 proteins

The identified cross-linked peptides together with their cross-linked amino acids were mapped to the crystal structure of Type I-D *T. pendens* Cas7 (PDB ID: 4TXD) and to predicted 3D-structure models of Type I-A *T. tenax* Cas7 and type III-A *T. thermophilus* Csm3 that were generated using the Phyre2 server [53]. We compared our results with the crRNA-binding surface of Type I-E *Escherichia coli* Cas7, which was crystallized in context of the fully assembled crRNP complex from *E. coli* [5]. For this, the crystal structure of *T. pendens* Cas7 and the homology models (*T. tenax* Cas7 and *T. thermophilus* Csm3) were superposed onto two copies of *E. coli* Cas7 (PDB ID: 1VY8) using secondary-structure matching (SSM) superposition in COOT [54]. In addition, the structure of *E. coli* Cas7 bound to crRNA was also used for superpositioning (Fig. 3). In all superimposed models of the Cas7 homologs, the crRNA uniformly contacts secondary structure elements of the peripheral insertion domain 1 as well as the central cleft defined by the core RRM and insertion domain 2. The cross-linking sites within *T. pendens* Cas7 encircle a positively charged groove and biochemical analysis demonstrated that conserved residues in this groove contribute significantly to RNA binding [24]. Moreover, the location of the cross-linked residues within the predicted insertion domain 1 of the proteins *T. tenax* Cas7 and *T. thermophilus* Csm3 are in full agreement with previous studies on the respective Type I-A and III-A homologs, *Sulfolobus solfataricus* Cas7 and *Methanopyrus kandleri* Csm3 [23,52].

4. Conclusions

We have established a general workflow of UV-induced cross-linking and mass spectrometry for the identification of proteins with their respective peptides and amino acids in contact with RNA. The workflow outlined here proves especially useful when crystal structures or structural models of RNA-binding proteins are available without their cognate RNA. In this case, the cross-linking sites help map the RNA on to the structure of its binding proteins. The given examples of the Cas7 protein homologs illustrate how in the absence of a conserved primary RNA binding motif a structurally conserved interface of this protein family contribute to a similar mode of RNA-binding. Cross-linking sites identified, in particular in those proteins and their motifs that have not previously been associated with RNA-binding, should be investigated in more detail e.g. by mutation studies and/or binding assays. Mutation studies should include not only the cross-linked amino acid but also the adjacent protein regions. Mutation of a specific cross-linking site might not completely abolish RNA-binding, as the RNA-binding region is larger than a single amino-acid residue. Such investigations had been performed on *T. pendens* protein Cas7 [24] or on the NHL domain (WD40 domain) of BRAT bound to RNA [55]. The protein–RNA cross-linking approach described here and in related studies [25] also addresses changes in binding of RNA to proteins in dependence upon different cellular environments and identifies transient interactions of the RNA with the proteins. In these cases several cross-linking sites in one and the same protein can be identified *in vivo* and *in vitro* [25,33] depending on the RNA and/or the cellular conditions. We note that non-specific cross-linking of proteins to RNA is barely observed, and is only found in studies of recombinant proteins when these are partially unfolded or denatured, or when they lack other components/proteins for their specific RNA-binding.

In the Type III-A Csm complex, the Csm proteins were cross-linked to the endogenous crRNA assembled in the complex. In all protein–RNA cross-links identified, the cross-linked nucleotide was found to be uridine. Uridine has been observed to be the most reactive nucleotide upon irradiation at 254 nm [56,57]. Cross-links also occur between C and G and proteins in other systems [25], but are less frequent. Under the conditions used here, and in other work described [25], cross-links of amino acids to adenosine have never been identified so far. Accordingly, we assume that protein cross-links to RNA that contain exclusively poly A stretches are difficult to obtain, and thus mapping of RNA-binding regions in the respective proteins by the method described here is expected to be difficult to achieve. UV irradiation at 254 nm also produces protein–DNA cross-links, but with much less efficiency as DNA exists mainly in its Watson–Crick base-paired form, in which the bases are unreactive towards amino acids (similarly to double-stranded RNA).

The approach of UV cross-linking at 254 nm wavelength, as described here, has also been applied to entire cells, such as the

whole yeast cells metabolically labeled with 4-thiouridine (4-S-U) as described in [25]. In the study of Kramer et al., the entire poly(A) mRNA population was isolated after UV cross-linking at 365 nm (which is the UV-irradiation wavelength for 4-S-U) for 30 min and cross-linked peptide–RNA oligonucleotide heteroconjugates were isolated to the current protocol. A difference between the approach for identification of protein–RNA cross-linking sites derived from reconstituted proteins with RNA oligonucleotides (as described here) and from RNA isolated from cells or derived from a cellular extract lies in the removal of non-cross-linked peptides. In the latter case, as described in detail by Kramer et al., the purified RNPs are first subjected to endoprotease digestion and then non-cross-linked peptides are removed by size-exclusion chromatography (SEC). This step is absolutely necessary as (i) phosphorylated peptides in the endogenous sample that interfere with the detection of cross-linked peptides are removed and (ii) the RNA population (e.g. premature and mature mRNAs) is usually larger, as short RNA oligonucleotides (e.g. 10–20-mers) so that an intact RNA population is isolated by SEC that contains cross-linked peptide. After endonuclease digestion, the non-cross-linked RNA oligonucleotides are removed by C18 chromatography exactly as described here, and enriched peptide–RNA oligonucleotides can be either directly subjected to LC–MS/MS or, additionally, can be further enriched by TiO_2 chromatography.

Possible limitations of the workflow are comparable to those for mass spectrometry based detection of post-translation modifications. Cross-link detection and localization is difficult, if the cross-linking site is located within a region of the protein that is not accessible for tryptic digestion (i.e. large tryptic peptides) or contains too many arginine and lysine residues. Therefore, the use of different endoproteinases is recommended to achieve optimum sequence coverage. However this in turn may influence the identification of the cross-linked peptide and the cross-linked amino acid as the peptides that do not harbor a basic amino acid at their C- or N-terminus can show poor fragment-ion series. In addition a sufficient amount of starting material can be challenging to obtain. Although cross-links are enriched, the chances of identifying all cross-linking sites within a protein (as most proteins have various sites of cross-linking) are higher, the greater the starting amount used and the larger the protein of interest.

The method described here could in principle also deliver sequence information about the cross-linked RNA moiety under conditions where protein–RNA cross-links with larger stretches of RNA (e.g. 10–20 mers or larger) are isolated and sequenced in the gas phase in the mass spectrometer. However, this is mainly hampered by the fact that the two components (peptides and RNA oligonucleotides) have different physico-chemical properties (similar to glyco-peptides). Gas-phase sequencing of cross-links with larger RNA oligonucleotides therefore results in fragmentation and sequencing of the RNA, but no sequence information about the cross-linked peptide part is obtained. In the case of glyco-peptides, electron-transfer dissociation (ETD) has been successfully applied to obtain sequence information about the peptide and the larger glycol moiety, as the modification remains on the amino-acid residue upon ETD fragmentation [58]. A similar analysis might be performed on peptide–RNA cross-links with larger RNA moieties.

UV-induced cross-linking followed by mass spectrometry has been proven to be highly useful for the identification of cross-linked amino acids, and thus of the RNA-binding site(s) in RNA-binding proteins. This approach is complementary to other UV-induced cross-linking approaches such as PAR-CLIP [59] and CRAC [60], in which next-generation sequencing techniques are applied in order to identify the nucleotides cross-linked to the proteins of interest. Combining of both these approaches in future studies promises an unprecedented insight into RBP biology at both the protein and the RNA level.

Author contributions

K.S. carried out the protein–RNA crosslinking experiments and data analysis in the lab of H.U. A.H. performed the expression and purification of *T. pendens* and *T. Tenax* Cas7 proteins in the lab of E.C, using the plasmid constructs provided by A.M. and L.R. respectively. A.H. performed the modeling and superposition for Fig. 3. R.S purified the endogenous *T. thermophilus* Type III-A Csm complex in the lab of J.v.d.O. K.K., T.S and O.K. established the data analysis workflow and provided useful suggestions and inputs for the manuscript. K.S. and H.U. wrote the manuscript.

Acknowledgements

The authors thank M. Raabe and U. Pleßmann for technical assistance, all the members of Urlaub laboratory and members of Forschergruppe 1680 for helpful discussions. This work was supported by the Deutsche Forschungsgemeinschaft [DFG, FOR 1680].

References

[1] B.M. Lunde, C. Moore, G. Varani, Nat. Rev. Mol. Cell Biol. 8 (2007) 479–490.
[2] S. Gerstberger, M. Hafner, T. Tuschl, Nat. Rev. Genet. 15 (2014) 829–845.
[3] C.G. Burd, G. Dreyfuss, Science 265 (1994) 615–621.
[4] H. Nishimasu, F.A. Ran, P.D. Hsu, S. Konermann, S.I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, O. Nureki, Cell 156 (2014) 935–949.
[5] R.N. Jackson, S.M. Golden, P.B. van Erp, J. Carter, E.R. Westra, S.J. Brouns, J. van der Oost, T.C. Terwilliger, R.J. Read, B. Wiedenheft, Science 345 (2014) 1473–1479.
[6] O. Duss, E. Michel, M. Yulikov, M. Schubert, G. Jeschke, F.H. Allain, Nature 509 (2014) 588–592.
[7] A.M. Anger, J.P. Armache, O. Berninghausen, M. Habeck, M. Subklewe, D.N. Wilson, R. Beckmann, Nature 497 (2013) 80–85.
[8] J.R. Stagno, A.S. Altieri, M. Bubunenko, S.G. Tarasov, J. Li, D.L. Court, R.A. Byrd, X. Ji, Nucleic Acids Res. 39 (2011) 7803–7815.
[9] H. Christian, R.V. Hofele, H. Urlaub, R. Ficner, Nucleic Acids Res. 42 (2014) 1162–1179.
[10] H. Urlaub, E. Kuhn-Holsken, R. Luhrmann, Methods Mol. Biol. 488 (2008) 221–245.
[11] A. Castello, B. Fischer, K. Eichelbaum, R. Horos, B.M. Beckmann, C. Strein, N.E. Davey, D.T. Humphreys, T. Preiss, L.M. Steinmetz, J. Krijgsveld, M.W. Hentze, Cell 149 (2012) 1393–1406.
[12] A.G. Baltz, M. Munschauer, B. Schwanhausser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, M. Landthaler, Mol. Cell 46 (2012) 674–690.
[13] M. Scheibe, F. Butter, M. Hafner, T. Tuschl, M. Mann, Nucleic Acids Res. 40 (2012) 9897–9902.
[14] C. Maris, C. Dominguez, F.H. Allain, FEBS J. 272 (2005) 2118–2131.
[15] R. Valverde, L. Edwards, L. Regan, FEBS J. 275 (2008) 2712–2726.
[16] S.M. Quintal, Q.A. dePaula, N.P. Farrell, Metall. Integr. Biomet. Sci. 3 (2011) 121–139.
[17] C.P. Ponting, Trends Biochem. Sci. 22 (1997) 51–52.
[18] S.D. Jayasena, B.H. Johnston, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 3526–3530.
[19] L. Aravind, E.V. Koonin, Trends Biochem. Sci. 24 (1999) 342–344.
[20] H. Hermann, P. Fabrizio, V.A. Raker, K. Foulaki, H. Hornig, H. Brahms, R. Luhrmann, EMBO J. 14 (1995) 2076–2088.
[21] M.A. Schumacher, R.F. Pearson, T. Moller, P. Valentin-Hansen, R.G. Brennan, EMBO J. 21 (2002) 3546–3556.
[22] C. Sauter, J. Basquin, D. Suck, Nucleic Acids Res. 31 (2003) 4091–4098.
[23] A. Hrle, A.A. Su, J. Ebert, C. Benda, L. Randau, E. Conti, RNA Biol. 10 (2013) 1670–1678.
[24] A. Hrle, L.K. Maier, K. Sharma, J. Ebert, C. Basquin, H. Urlaub, A. Marchfelder, E. Conti, RNA Biol. 11 (2014).
[25] K. Kramer, T. Sachsenberg, B.M. Beckmann, S. Qamar, K.L. Boon, M.W. Hentze, O. Kohlbacher, H. Urlaub, Nat. Methods 11 (2014) 1064–1070.
[26] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero, P. Horvath, Science 315 (2007) 1709–1712.
[27] J. van der Oost, M.M. Jore, E.R. Westra, M. Lundgren, S.J. Brouns, Trends Biochem. Sci. 34 (2009) 401–407.
[28] K.S. Makarova, D.H. Haft, R. Barrangou, S.J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F.J. Mojica, Y.I. Wolf, A.F. Yakunin, J. van der Oost, E.V. Koonin, Nat. Rev. Microbiol. 9 (2011) 467–477.
[29] R. Jansen, J.D. Embden, W. Gaastra, L.M. Schouls, Mol. Microbiol. 43 (2002) 1565–1575.

[30] R.H. Staals, Y. Zhu, D.W. Taylor, J.E. Kornfeld, K. Sharma, A. Barendregt, J.J. Koehorst, M. Vlot, N. Neupane, K. Varossieau, K. Sakamoto, T. Suzuki, N. Dohmae, S. Yokoyama, P.J. Schaap, H. Urlaub, A.J. Heck, E. Nogales, J.A. Doudna, A. Shinkai, J. van der Oost, Mol. Cell 56 (2014) 518–530.
[31] A. Plagens, V. Tripp, M. Daume, K. Sharma, A. Klingl, A. Hrle, E. Conti, H. Urlaub, L. Randau, Nucleic Acids Res. 42 (2014) 5125–5138.
[32] X. Luo, H.H. Hsiao, M. Bubunenko, G. Weber, D.L. Court, M.E. Gottesman, H. Urlaub, M.C. Wahl, Mol. Cell 32 (2008) 791–802.
[33] J. Schmitzova, N. Rasche, O. Dybkov, K. Kramer, P. Fabrizio, H. Urlaub, R. Luhrmann, V. Pena, EMBO J. 31 (2012) 2222–2234.
[34] H. Urlaub, V.A. Raker, S. Kostka, R. Luhrmann, EMBO J. 20 (2001) 187–196.
[35] A. Castello, R. Horos, C. Strein, B. Fischer, K. Eichelbaum, L.M. Steinmetz, J. Krijgsveld, M.W. Hentze, Nat. Protoc. 8 (2013) 491–500.
[36] H. Urlaub, V. Kruft, O. Bischof, E.C. Muller, B. Wittmann-Liebold, EMBO J. 14 (1995) 4578–4588.
[37] H. Urlaub, K. Hartmuth, S. Kostka, G. Grelle, R. Luhrmann, The Journal of biological chemistry 275 (2000) 41458–41468.
[38] S. Nottrott, H. Urlaub, R. Luhrmann, EMBO J. 21 (2002) 5527–5538.
[39] T. Glatter, C. Ludwig, E. Ahrne, R. Aebersold, A.J. Heck, A. Schmidt, J. Proteome Res. 11 (2012) 5145–5156.
[40] K. Kramer, P. Hummel, H.H. Hsiao, X. Luo, M. Wahl, H. Urlaub, Int. J. Mass Spectrom. 304 (2011) 184–194.
[41] M.R. Larsen, T.E. Thingholm, O.N. Jensen, P. Roepstorff, T.J. Jorgensen, Mol. Cell. Proteomics: MCP 4 (2005) 873–886.
[42] M.W. Pinkse, S. Lemeer, A.J. Heck, Methods Mol. Biol. 753 (2011) 215–228.
[43] E. Kuhn-Holsken, C. Lenz, B. Sander, R. Luhrmann, H. Urlaub, RNA 11 (2005) 1915–1930.
[44] E. Kuhn-Holsken, O. Dybkov, B. Sander, R. Luhrmann, H. Urlaub, Nucleic Acids Res. 35 (2007) e95.
[45] A. Bertsch, C. Gropl, K. Reinert, O. Kohlbacher, Methods Mol. Biol. 696 (2011) 353–367.
[46] M. Sturm, A. Bertsch, C. Gropl, A. Hildebrandt, R. Hussong, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, A. Zerck, K. Reinert, O. Kohlbacher, BMC Bioinformatics 9 (2008) 163.
[47] L.Y. Geer, S.P. Markey, J.A. Kowalak, L. Wagner, M. Xu, D.M. Maynard, X. Yang, W. Shi, S.H. Bryant, J. Proteome Res. 3 (2004) 958–964.
[48] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W.H. Tang, A. Rompp, S. Neumann, A.D. Pizarro, L. Montecchi-Palazzi, N. Tasman, M. Coleman, F. Reisinger, P. Souda, H. Hermjakob, P.A. Binz, E.W. Deutsch, Mol. Cell. Proteomics: MCP 10 (R110) (2011) 000133.
[49] D. Kessner, M. Chambers, R. Burke, D. Agus, P. Mallick, Bioinformatics 24 (2008) 2534–2536.
[50] C. Rouillon, M. Zhou, J. Zhang, A. Politis, V. Beilsten-Edmands, G. Cannone, S. Graham, C.V. Robinson, L. Spagnolo, M.F. White, Mol. Cell 52 (2013) 124–134.
[51] M. Spilman, A. Cocozaki, C. Hale, Y. Shao, N. Ramia, R. Terns, M. Terns, H. Li, S. Stagg, Mol. Cell 52 (2013) 146–152.
[52] N.G. Lintner, M. Kerou, S.K. Brumfield, S. Graham, H. Liu, J.H. Naismith, M. Sdano, N. Peng, Q. She, V. Copie, M.J. Young, M.F. White, C.M. Lawrence, J. Biol. Chem. 286 (2011) 21643–21656.
[53] L.A. Kelley, M.J. Sternberg, Nat. Protoc. 4 (2009) 363–371.
[54] E. Krissinel, K. Henrick, Acta Crystallogr. Sect. D, Biol. Crystallogr. 60 (2004) 2256–2268.
[55] I. Loedige, M. Stotz, S. Qamar, K. Kramer, J. Hennig, T. Schubert, P. Loffler, G. Langst, R. Merkl, H. Urlaub, G. Meister, Genes Dev. 28 (2014) 749–764.
[56] M.D. Shetlar, K. Home, J. Carbone, D. Moy, E. Steady, M. Watanabe, Photochem. Photobiol. 39 (1984) 135–140.
[57] M.D. Shetlar, J. Carbone, E. Steady, K. Hom, Photochem. Photobiol. 39 (1984) 141–144.
[58] Y. Mechref, Current protocols in protein science/editorial board, John E. Coligan ... et al., Chapter 12 (2012) Unit 12 11 11–11.
[59] 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, Cell 141 (2010) 129–141.
[60] S. Granneman, G. Kudla, E. Petfalski, D. Tollervey, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 9613–9618.