

# Single Nucleotide Resolution RNA–Protein Cross-Linking Mass Spectrometry: A Simple Extension of the CLIR-MS Workflow

Michael Götze,\* Chris P. Sarnowski, Tebbe de Vries, Anna Knörlein, Frédéric H.-T. Allain, Jonathan Hall, Ruedi Aebersold, and Alexander Leitner\*



Cite This: *Anal. Chem.* 2021, 93, 14626–14634



Read Online

ACCESS |



Metrics & More

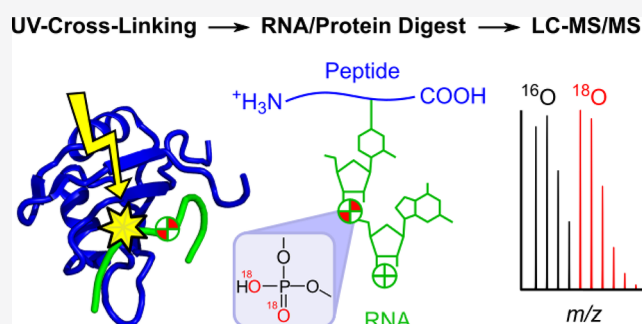


Article Recommendations



Supporting Information

**ABSTRACT:** RNA–protein interactions mediate many intracellular processes. CLIR-MS (cross-linking of isotope-labeled RNA and tandem mass spectrometry) allows the identification of RNA–protein interaction sites at single nucleotide/amino acid resolution in a single experiment. Using isotopically labeled RNA segments for UV-light-induced cross-linking generates characteristic isotope patterns that constrain the sequence database searches, increasing spatial resolution. Whereas the use of segmentally isotopically labeled RNA is effective, it is technically involved and not applicable in some settings, e.g., in cell or tissue samples. Here we introduce an extension of the CLIR-MS workflow that uses unlabeled RNA during cross-linking and subsequently adds an isotopic label during sample preparation for MS analysis. After RNase and protease digests of a cross-linked complex, the nucleic acid part of a peptide–RNA conjugate is labeled using the enzyme T4 polynucleotide kinase and a 1:1 mixture of heavy  $^{18}\text{O}_4$ - $\gamma$ -ATP and light ATP. In this simple, one-step reaction, three heavy oxygen atoms are transferred from the  $\gamma$ -phosphate to the 5'-end of the RNA, introducing an isotopic shift of 6.01 Da that is detectable by mass spectrometry. We applied this approach to the RNA recognition motif (RRM) of the protein FOX1 in complex with its cognate binding substrate, FOX-binding element (FBE) RNA. We also labeled a single phosphate within an RNA and unambiguously determined the cross-linking site of the FOX1-RRM binding to FBE at single residue resolution on the RNA and protein level and used differential ATP labeling for relative quantification based on isotope dilution. Data are available via ProteomeXchange with the identifier PXD024010.



## INTRODUCTION

RNA is a major regulator of biological processes and can be classified into a variety of subtypes.<sup>1</sup> RNAs, in concert with RNA-binding proteins, regulate biological functions, which include but are not limited to transcription, splicing, intracellular transport, translation, stress response, and virulence of pathogens, highlighting the importance of RNA–protein interactions. Thousands of different proteins bind to RNA in human cells,<sup>2</sup> and conversely, RNAs are almost universally bound to proteins. RNAs as well as RNA-binding proteins are involved in different diseases including, for example, cancer (e.g., IGF2BP1<sup>3</sup>) and neurodegenerative diseases (e.g., TDP43 in amyotrophic lateral sclerosis and frontotemporal lobar degeneration<sup>4</sup>). RNA-based or RNA-targeting drugs play an important role in current pharmaceutical development. For example, nusinersen is an antisense-oligonucleotide drug that targets SMN2 in spinal muscular atrophy.<sup>5</sup> The identification of RNA–protein interaction sites at single monomer resolution (nucleotide/amino acid) is therefore of great biological and clinical significance.

Different approaches are available to study RNA–protein interactions from a protein-centric view (to identify novel RNA-

binding proteins and domains) or from an RNA-centric view (to identify novel RNA targets of specific RNA-binding proteins (RBP)). The latter group of methods, e.g., CLIP-Seq (cross-linking immunoprecipitation-high-throughput sequencing), involves next-generation sequencing after affinity purification of an RBP cross-linked to its target RNAs.<sup>6–8</sup> In contrast, protein-centric analyses identify proteins bound to a specific RNA or a subset of RNAs employing, e.g., cross-linking by UV light and mass spectrometry.<sup>2,9–14</sup> Such a method was used to characterize the mRBPome by cross-linking mRNA-bound proteins and subsequent purification of mRNAs by oligo-dT chromatography.<sup>2</sup> The specific amino acid residue that cross-linked to an RNA is identified using mass spectrometry.<sup>11</sup> These methods provide the exact, single monomer interaction site of an RNA–protein complex either from the RNA side, using sequencing

Received: June 6, 2021

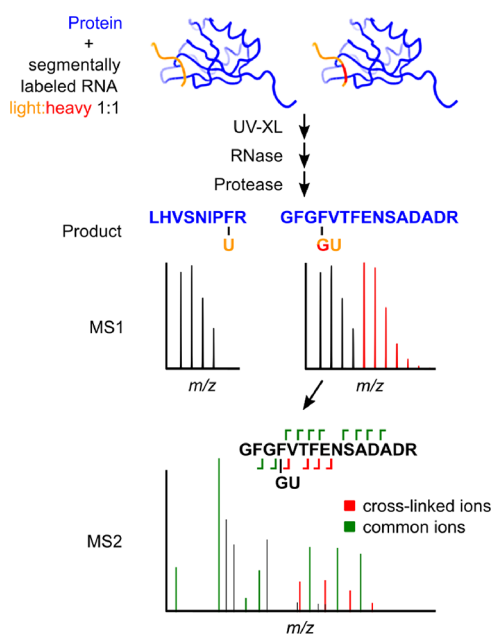
Accepted: October 19, 2021

Published: October 29, 2021



techniques, or from the protein site, employing mass spectrometry. Acquiring the precise, single monomer interaction site on both molecules at the same time is not possible with these strategies.

We recently described CLIR-MS, a method that identifies the interaction sites on photo-chemically cross-linked RNA–protein complexes with single monomer resolution for both the RNA-binding protein and the bound RNA (Figure 1).<sup>12</sup> To



**Figure 1.** Elucidation of cross-linking sites using CLIR-MS. A mixture of unlabeled and segmentally labeled RNA is cross-linked to an interacting protein using UV light at a wavelength of 254 nm. After sample preparation, including RNase and protease digests, mass spectrometric analysis of RNA–peptide adducts is performed. Only peptides that were cross-linked to the labeled RNA segment show a specific isotopic pattern in precursor mass scans (MS1). To localize the cross-link position in the peptide backbone, products are fragmented and tandem mass spectra (MS2) reveal the cross-linking site due to sequence-specific fragmentation, where fragments containing the oligonucleotide adduct will show isotopic shifts when comparing light and heavy MS2 spectra (cross-linked ions). Fragment ions without RNA adducts will share the same isotopic distribution in light and heavy MS2 spectra (common ions).

guide unequivocal assignments of the interaction sites from fragment ion spectra of cross-linked entities, CLIR-MS was designed using segmentally isotopically labeled RNA. The UV-cross-linked RNA–protein sample was digested with RNases and proteases and enriched using TiO<sub>2</sub> chromatography. Cross-links containing the labeled segment will appear as isotopic pairs in MS1 spectra. The spatial resolution of the RNA-cross-linking site therefore depends on the length and the right choice of the labeled stretch within the RNA. The exact peptide-level localization of an RNA–protein cross-link was determined using tandem mass spectrometry analysis (MS2) similar to the localization of a post-translational modification (Figure 1). Additionally, isotopic labeling produces a unique pattern during mass spectrometric analysis that grants two major advantages: (1) Precursor ions that do not show the encoded isotopic pattern can be disregarded from downstream analysis, which ensures that an identified peptide modification is RNA-derived. (2) MS2 spectra of light and heavy species are obtained and can

be combined, which facilitate data analysis. We use xQuest,<sup>15,16</sup> a cross-linking mass spectrometry software designed for isotope-labeled protein–protein cross-links, to analyze CLIR-MS data.

CLIR-MS was developed using *in vitro* transcribed RNA using metabolically labeled (<sup>13</sup>C/<sup>15</sup>N) nucleotide triphosphates (NTPs).<sup>17</sup> Labeled as well as unlabeled RNAs have to be prepared and purified individually. Segmental labeling of RNAs is achieved by the consecutive ligation of three *in vitro* transcribed and purified RNAs, limiting the yield as well as the minimum length of the labeled segment within the RNA. Further, RBPs can only be analyzed with this approach if the binding sequence is known and available as a transcription template; CLIR-MS cannot be used, for instance, to probe RNA–protein interactions in cells or tissue lysates. Finally, the workflow also requires that labeled NTPs are prepared from *Escherichia coli* cultures grown in expensive, heavy labeled media.

Here we describe two labeling methods that can be used (1) to introduce a heavy label during sample preparation after protease and RNase digests (“post-digest labeling”) and (2) to site-specifically introduce a heavy label in the RNA backbone prior to UV cross-linking, reducing the length of the labeled segment within the RNA to a single phosphate (“single-phosphate labeling”). Both methods make use of the T4-PNK and commercially available <sup>18</sup>O<sub>4</sub>-γ-ATP to introduce an isotope-labeled phosphate into RNA. We envision that these labeling techniques enable the use of CLIR-MS (1) for more complex samples, such as immunoprecipitates of RNA-binding proteins, and (2) for simplified in-depth analysis of an RNA–protein complex with single nucleotide resolution. As a proof of principle, we applied CLIR-MS with two novel labeling techniques to the RRM of the FOX1 protein bound to its cognate RNA substrate, referred to as FBE-RNA.<sup>18</sup> FOX1 is a nuclear protein that acts as a splicing regulator by binding to RNAs at splicing–enhancing sequence elements with high sequence specificity (5′-UGCAUG-3′). Besides mRNAs, more than 100 pre-miRNAs contain high affinity FOX1 binding sites. FOX proteins have been shown to be involved in the miRNA biogenesis of several of those FBE-containing pre-miRNAs.<sup>19</sup> We could identify multiple cross-links in the RRM of FOX1 to several RNA adducts up to a length of four nucleotides. Using site-specific labeling, we could show that F160 cross-links to U6/G7 within the splicing enhancer sequence, a result that agrees with previously published structures.<sup>18</sup> The two labeling approaches complement the available toolset for mass spectrometry-based RNA–protein interaction studies.

## EXPERIMENTAL SECTION

**Protein Preparation.** The FOX1-RRM protein was expressed in transformed BL21 Codon+ *Escherichia coli* at 37 °C in an LB medium with kanamycin and chloramphenicol and purified as previously described.<sup>18</sup> Cells were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at an OD<sub>600</sub> of 0.6. After 4 h, the cells were harvested by centrifugation. Cells were lysed in buffer A (50 mM Na<sub>2</sub>HPO<sub>4</sub> and 1 M NaCl (pH 8)) using a cell homogenizer (Microfluidizer LM10, Instrumat). The cell lysate was centrifuged at 20,000g at 4 °C for 30 min, and the supernatant was subjected to Ni-NTA affinity chromatography (Ni-NTA agarose, Qiagen). After washing with buffer B (50 mM Na<sub>2</sub>HPO<sub>4</sub> and 3 M NaCl (pH 8)), the protein was eluted with a step gradient of imidazole (40–500 mM). The fractions with the highest purity as judged by 5–20% SDS-PAGE were combined, and the column purification was repeated. Pure fractions were dialyzed against 5 L of buffer C

(20 mM NaCl and 10 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 6.5)) for 18 h. The protein was concentrated to ~4 mM by centrifugation at 4 °C using a 3 kDa molecular mass cutoff membrane. The identity of the protein was confirmed by ESI-MS (Bruker maXis, mass calculated = 13562.18 Da, mass found = 13562.84 Da). No other bands were visible when the purified protein was analyzed by SDS-PAGE.

**Preparation of Single-Phosphate Labeled RNA.** Chemically synthesized RNA (UUGUCA; 20 nmol), blocked at the 3'-end with a C3-spacer (C<sub>3</sub>H<sub>7</sub>O) (Integrated DNA Technologies), was labeled with a 1:1 mixture of light ATP and <sup>18</sup>O<sub>4</sub>-γ-ATP (Cambridge Isotope Laboratories Inc.) using 40 U of T4 polynucleotide kinase (T4-PNK, New England Biolabs, NEB) in 1× PNK buffer for 1 h at 37 °C in a 200 μL reaction. T4-PNK was inactivated for 20 min at 65 °C. The reaction was supplemented with 20 μL of the T4 RNA ligase buffer (NEB), 20 nmol chemically synthesized RNA (UAAGUUGCAUG, Integrated DNA technologies), 152 μL of PEG-8000, and 40 U of T4 RNA ligase (NEB). The reaction was carried out for 1 h at 25 °C followed by an overnight incubation at 16 °C. T4 RNA ligase was inactivated for 15 min at 65 °C. The ligated RNA was purified by ethanol precipitation twice by adding 1/10 vol of 3 M sodium acetate (pH 5.2) and 3 vol of ice-cold ethanol. The RNA was incubated at -20 °C for at least 30 min and centrifuged at 4 °C for 30 min at 16,000g. The pellet was washed with ice-cold 80% ethanol, dried, and dissolved in HPLC-grade water.

**UV Cross-Linking.** The FOX1-RRM and RNA were mixed in 10 mM sodium phosphate buffer (pH 6.9) and 10 mM NaCl at equimolar concentrations (10 μM). For cross-linking, the sample was dispensed in a 96-well plate in 50 μL aliquots and cross-linking was performed on ice at a distance of 2 cm to the UV lamps in a Spectrolinker XL-1500 UV Crosslinker (Spectronics Corporation) at an energy of 3.2 J/cm<sup>2</sup> and 254 nm wavelength. The same distance to UV lamps was maintained by using a 3D-printed ice bucket holding a 96-well plate (Figure S1).

**RNase/Protease Digest.** The cross-linked sample (100 μg of FOX1-RRM) was precipitated using 1/10 vol of 3 M sodium acetate (pH 5.2) and 3 vol of ice-cold ethanol, stored at -20 °C for at least 2 h, and centrifuged at 4 °C for 30 min at 16,000g. After washing the pellet with 80% ethanol, the sample was dissolved in 50 mM Tris-HCl (pH 8) and 4 M urea. After resuspension, the sample was diluted to 1 M urea with 50 mM Tris-HCl (pH 8). Five units of RNase T1 (Thermo Fisher Scientific) and 5 μg of RNase A (Roche) were added per mg of the cross-linked complex and incubated at 52 °C for 2 h.<sup>11,12</sup> Sequencing-grade trypsin (Promega) was added at a 24:1 substrate-to-enzyme ratio and incubated overnight at 37 °C. Trypsin was inactivated at 70 °C for 10 min. Another 4 U of RNase T1 and 4 μg of RNase A were added per mg of the cross-linked complex and incubated at 37 °C for 1 h. Solid-phase extraction using Waters C<sub>18</sub> SepPak columns (50 mg) was performed, and the samples were evaporated to dryness in a vacuum centrifuge.

**Post-digest Labeling.** The dried sample was dissolved in 86 μL of water, 10 μL of 10× T4-PNK buffer (NEB), 1 μL of 100 mM ATP, and 1 μL of 100 mM <sup>18</sup>O<sub>4</sub>-γ-ATP. Labeling was performed with 20 U of T4-PNK (NEB) for 1 h at 37 °C. Another solid-phase extraction step was performed as above.

**Metal Oxide Affinity Enrichment.** The dried sample was dissolved in 60 μL of the loading buffer (50% acetonitrile, 300 mg/mL lactic acid, and 0.1% trifluoroacetic acid (TFA)) and incubated for 30 min with 3 mg of equilibrated TiO<sub>2</sub> beads

(Titansphere PhosTiO 10 μm, GL Sciences). The beads were washed sequentially with 60 μL of the loading buffer and washing buffer (50% acetonitrile and 0.1% TFA) each, cross-linked peptides were eluted twice with 50 μL of the elution buffer (50 mM (NH<sub>4</sub>)<sub>2</sub>HPO<sub>4</sub> (pH 10.5)), and the solution was immediately acidified by adding TFA. Cross-linked peptides were further purified by C<sub>18</sub> StageTip solid-phase extraction. Two layers of C<sub>18</sub> membranes (3M Empore) were washed sequentially with (1) 100% acetonitrile (ACN), (2) 80% ACN with 0.1% formic acid (FA), and (3) two times 5% ACN with 0.1% FA. After applying the sample, the tips were washed three times using 5% ACN with 0.1% FA and finally eluted three times using 50% ACN with 0.1% FA. The eluate was collected in prewashed LoBind tubes (Eppendorf), and samples were evaporated to dryness.

**Liquid Chromatography–Tandem Mass Spectrometry.** Samples were resuspended in 20 μL of water/acetonitrile/formic acid (95:5:0.1, v/v/v), and 5 μL was used for analysis in technical duplicates. LC–MS/MS analysis was performed with an Easy nLC 1200 HPLC system (ThermoFisher Scientific) connected to an Orbitrap Fusion Lumos mass spectrometer (ThermoFisher Scientific) equipped with a Nanoflex electrospray source. Peptides were separated on a PepMap RSLC column (250 mm × 75 μm, 2 μm particle size, ThermoFisher Scientific) using a gradient of 6–32% mobile phase B within 60 min, where A = water/acetonitrile/formic acid (98:2:0.15, v/v/v) and B = acetonitrile/water/formic acid (80:20:0.15, v/v/v); the flow rate was set to 300 nL/min.

The Orbitrap Fusion Lumos was operated in data-dependent acquisition mode. Precursor ion spectra were acquired in the Orbitrap analyzer at a resolution of 120,000 in 3 s cycles (top speed mode). During each cycle, precursor ions were selected for fragmentation using stepped higher energy collision-induced dissociation (stepped HCD, normalized collision energy, 23 ± 5%) and detection of the fragment ions in the linear ion trap in rapid scan mode or in the Orbitrap mass analyzer at a resolution of 30,000. Additional fragmentation settings were as follows: isolation width, 1.2 *m/z*; dynamic exclusion (30 s after one sequencing event) was activated; and selected charge states = 2–7+.

**Data Analysis.** For data analysis, files were converted from the native Thermo raw format into centroided mzXML using msconvert (ProteoWizard version 3.0.9393)<sup>20</sup> and searched against the target protein sequence and its reversed sequence using xQuest (version 2.1.5, available from [https://gitlab.ethz.ch/leitner\\_lab/xquest\\_xprophet](https://gitlab.ethz.ch/leitner_lab/xquest_xprophet)).<sup>15</sup> No contaminant proteins that might have originated from the protein preparation were observed at relevant abundances. To adapt xQuest to the search of different types of nucleotide adducts on arbitrary amino acid residues, all amino acid residues were specified as possible modification sites. Based on the target RNA sequence UGCAUGU, 15 different nucleotide adducts up to a length of four nucleotides, combined with different neutral losses, were considered. In total, 75 different mass modifications between ~300 and 1400 Da were included in the search (Table S1). Mass shifts (isotopeshift/cp\_isotopediff) of 6.012735 and 4.008490 Da were defined in the xquest.def and xmm.def configuration files for post-digest labeling and single-phosphate labeling, respectively. A mass tolerance of 15 ppm and a retention time tolerance of 60 s were used to match heavy and light spectra.

Additional search settings were as follows: enzyme = trypsin, maximum number of missed cleavages = 2, MS1 mass tolerance = 10 ppm, and MS2 mass tolerance = 0.2 Da for "common"-type

fragment ions and 0.3 Da for "xlink"-type fragment ions. The scoring scheme of xQuest as presented in Walzthoeni et al. (2012)<sup>16</sup> was used, and only identifications with a score  $\geq 20$  were considered for later analysis.

**Relative Quantification of RNA Cross-Links.** FOX1-RRM and FBE-RNA were cross-linked and labeled individually with light and heavy ATP. Light/heavy ratios from 25:1 to 1:25 were prepared by mixing directly after labeling, and samples were analyzed by LC-MS/MS as above. For relative quantification, the raw data were analyzed using Skyline (version 20.1.0.76). Sixty-one different precursor mass-to-charge ratios of RNA-peptide adducts with different neutral losses were quantified in light and heavy states (6.01 Da mass difference) in MS1 spectra (Table S2).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository<sup>21</sup> with the data set identifier PXD024010. All identified cross-links are listed in Table S3.

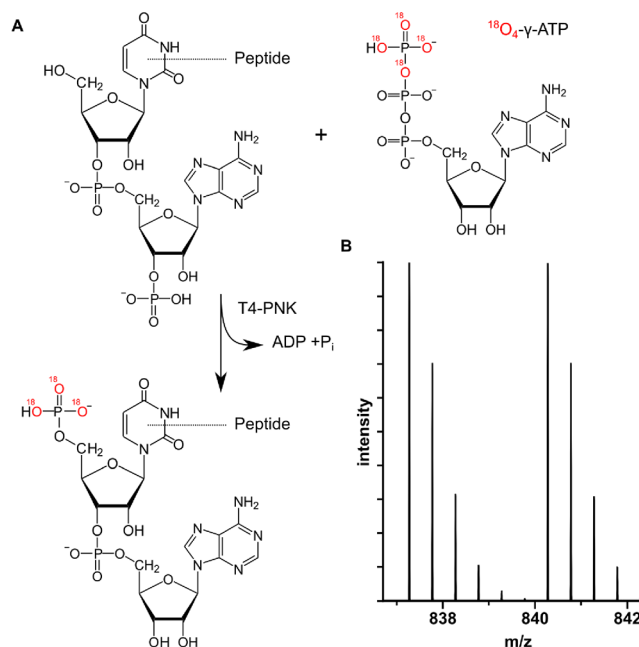
## RESULTS

### Labeling of Cross-Linked RNA–Peptide Adducts.

Using CLIR-MS, the nucleotide composition of RNA adducts can be resolved but not the RNA sequence itself. For short substrate RNA sequences, identifying the composition of a di- or trinucleotide segment cross-linked to a peptide might be specific enough to precisely pinpoint this cross-link to a single region within the RNA. For longer RNA substrates, due to the limited information contained in short RNA segments, the composition of a cross-linked nucleic acid is ambiguous with regard to its localization within the RNA. To enable new applications of the CLIR-MS workflow and to make it compatible with unlabeled RNAs and more complex samples, we devised a straightforward extension of the sample preparation procedure whereby the oligonucleotide–peptide adduct is labeled after cross-linking and protease/RNase digestion. It is illustrated in Scheme 1A. We used commercially available  $^{18}\text{O}_4\text{-}\gamma\text{-ATP}$  and T4-PNK to label the 5'-hydroxy groups of cross-linked oligonucleotides with a phosphate group containing three heavy oxygen atoms. An RNA–peptide adduct after RNase digest contains one 5'-hydroxy group so that only a single label is introduced irrespective of the length of the RNA. When using ATP and  $^{18}\text{O}_4\text{-}\gamma\text{-ATP}$  in a 1:1 mixture, an isotopic signature is introduced that is subsequently detected by mass spectrometry. Heavy/light doublet signals will be observed for the intact RNA–peptide adducts similarly to the use of pre-labeled RNA in the original CLIR-MS workflow (Scheme 1B).

**Application of the Method to the Complex of FOX1-RRM and FBE-RNA.** To benchmark our labeling strategies, we chose to analyze the RRM of the FOX1 protein in complex with an FBE-containing RNA (UGCAUGU). This complex has a high affinity with a dissociation constant of  $<1$  nM.<sup>18</sup> We used the recombinantly expressed His-tagged RNA-recognition domain of FOX1 (residues 109–208), mixed it with RNA in equimolar amounts, and cross-linked the complex using UV light at a wavelength of 254 nm. The sample was processed as described in Dorn et al. (2017)<sup>12</sup> (for details, see the Experimental Section), and  $^{18}\text{O}$ -labeling was applied after protease and RNase digests. After  $\text{TiO}_2$  enrichment, the sample was analyzed by LC-MS/MS. Mass spectrometric analysis revealed multiple isotopic pairs introduced by  $^{18}\text{O}$ -labeling (Figure 2A), and data analysis using xQuest revealed two major cross-linking sites within the FOX1-RRM: F126 and F160

### Scheme 1. Post-digest Labeling Reaction Scheme<sup>44</sup>

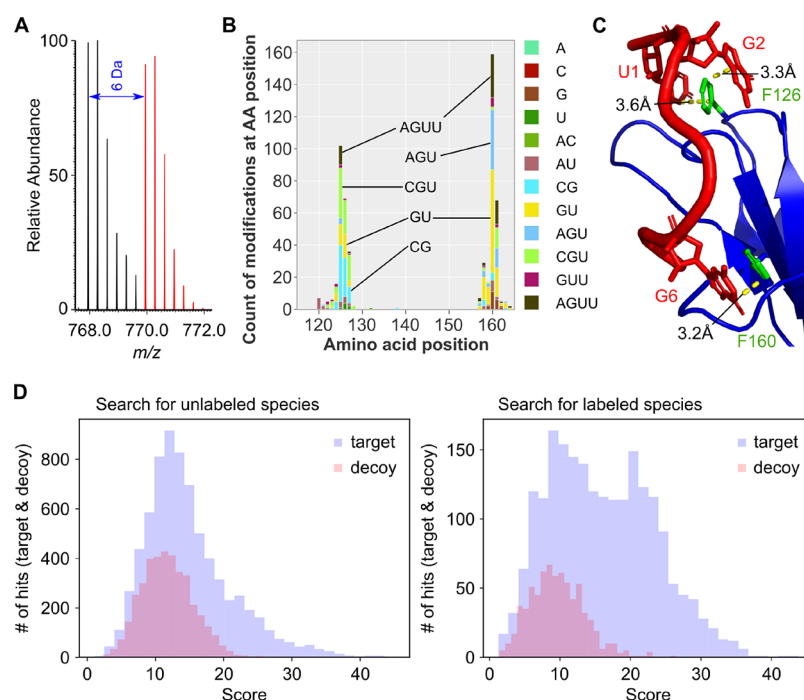


<sup>44</sup>(A) The  $\gamma$ -phosphate of  $^{18}\text{O}_4\text{-}\gamma\text{-ATP}$  is transferred to the 5'-hydroxy group of a cross-linked RNA oligonucleotide. T4-PNK also catalyzes a 3'-dephosphorylation reaction. (B) Simulation of the isotopic distribution<sup>33</sup> of a cross-linked peptide (PEPTIDER cross-linked to a GU dinucleotide, doubly charged) for a 1:1 mixture of light and heavy ATP using single-phosphate labeling.

(Figure 2B). Both residues are in very close proximity to RNA in the NMR spectroscopy-derived structure of the FOX1-RRM in complex with FBE-RNA (PDB ID 2ERR), validating the results obtained by post-digest labeling (Figure 2C). Because only the subset of the MS2 spectra characterized by light/heavy pairs at the label-defined delta mass needs to be searched and the merging of the spectra removes noise peaks, the identification of cross-links from fragment ion spectra is improved and we observed an enhanced separation of target and decoy hits (Figure 2D) compared to an exhaustive search that ignores isotopic labels. (Decoy hits were obtained by reversing the amino acid sequence of FOX1, and matches to this sequence must therefore be false positives.) Nevertheless, the estimation of the false discovery rate for RNA–peptide cross-links is more complex than that for peptides due to the heterogeneous nature of the cross-linked adducts. For the presented data, we chose to use a fixed Id-score threshold of 20, as only few peptide decoy hits can be identified above this threshold.

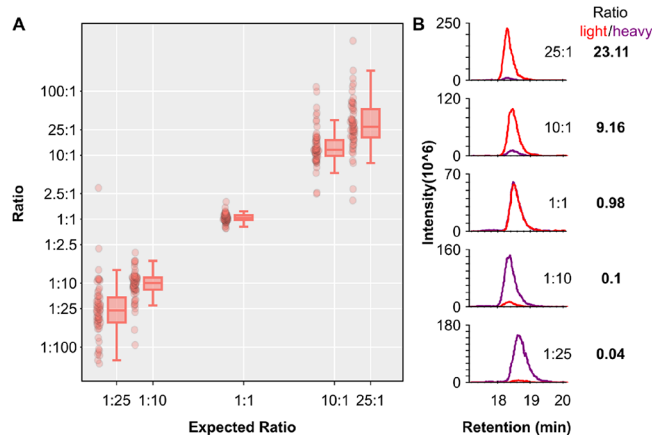
### Relative Quantification of Light and Heavy Signals.

Differential stable isotope labeling enables the precise relative quantification of different states that are labeled with either light or heavy isotopes based on the principles of stable isotope dilution. A simple labeling strategy enables similar approaches to be applied for ribonucleoprotein analyses with CLIR-MS. Using the post-digest labeling technique described above, a mass shift of  $\sim 6$  Da is introduced that is sufficient to separate the isotopic signals from naturally occurring  $^{13}\text{C}$ -isotopes and perform quantification of the MS1 signals. To illustrate this, we labeled aliquots of a cross-linked sample of the FOX1-RRM/FBE-complex with ATP and  $^{18}\text{O}_4\text{-}\gamma\text{-ATP}$ , respectively, and subsequently mixed the samples at different ratios ranging from 25:1 to 1:25. MS1 traces of 61 individual mass-to-charge



**Figure 2.** Cross-linking of FOX1-RRM to FBE-RNA (UGCAUGU). (A) Isotopic pattern ( $\Delta m = 6.01$  Da) in the precursor mass spectrum (MS1) of the cross-linked RNA–peptide adduct GFGFVTFENSADADR (1631.72 Da) cross-linked to the GU dinucleotide with phosphate loss at the 3′-end (669.08 Da) ( $[M + 3H]^{3+} = 767.94$   $m/z$ ). Heavy isotopic signals are highlighted in red. (B) Identified RNA–protein cross-links represented by adduct type and sequence position. The adduct type represents the nucleotide composition of the RNA adduct. Counts represent cross-link spectrum matches. Two major cross-linking sites are identified at F126 and F160. (C) Identified cross-links match the NMR structure of the RNA–protein complex (PDB ID 2ERR) with distances of less than 5 Å to the closest atom in the RNA base. The RNA with the sequence UGCAUGU is shown in red, and the FOX1-RRM is represented in blue. (D) CLIR-MS data of duplicate runs were analyzed twice: once disregarding the introduced label and once considering the isotopic label. The distribution of target and decoy hits (to the reversed protein sequence) is shown. A score threshold of 20 was used to filter all results in this manuscript. When including the isotopic labeling information, target and decoy hits clearly separate better (right panel).

ratios were quantified using Skyline (Figure 3A). These traces represent 27 unique RNA–peptide cross-link combinations, aggregating different charge states and types of neutral losses. The MS1 traces for a single precursor at the different concentrations are shown in Figure 3B. The relative



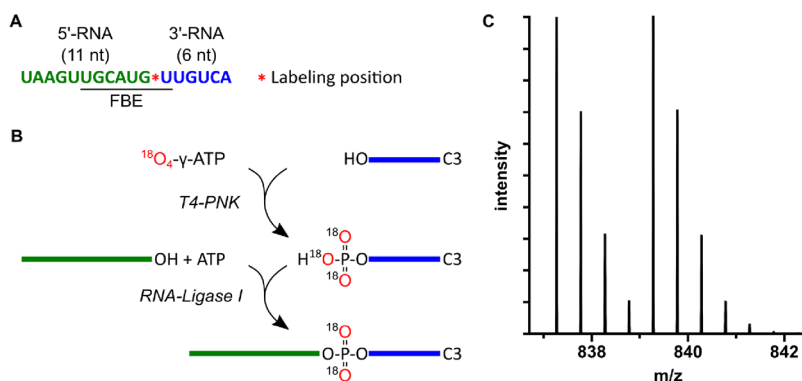
**Figure 3.** Quantification of RNA–peptide adducts. (A) Relative quantification of 61 precursor ions (different mass-to-charge ratios) of light/heavy cross-linked RNA–peptide adducts in set-ratio mixtures (25:1, 10:1, 1:1, 1:10, and 1:25) presented on a double logarithmic plot. The quantification of these ions is linear over almost 3 orders of magnitude ( $R^2 = 0.93$ ). (B) Example MS1 traces ( $m/z$  1028.36,  $[M + 2H]^{2+}$ ) of the peptide LHVSNIPFR (1081.60 Da) cross-linked to a UGU trinucleotide (973.09 Da).

quantification of MS1 signals is linear over almost 3 orders of magnitude. This technique can therefore also be applied to detect changes in RNA–protein cross-linking in different conditions using two differentially labeled biological states or an in vitro generated heavy reference standard for normalization.

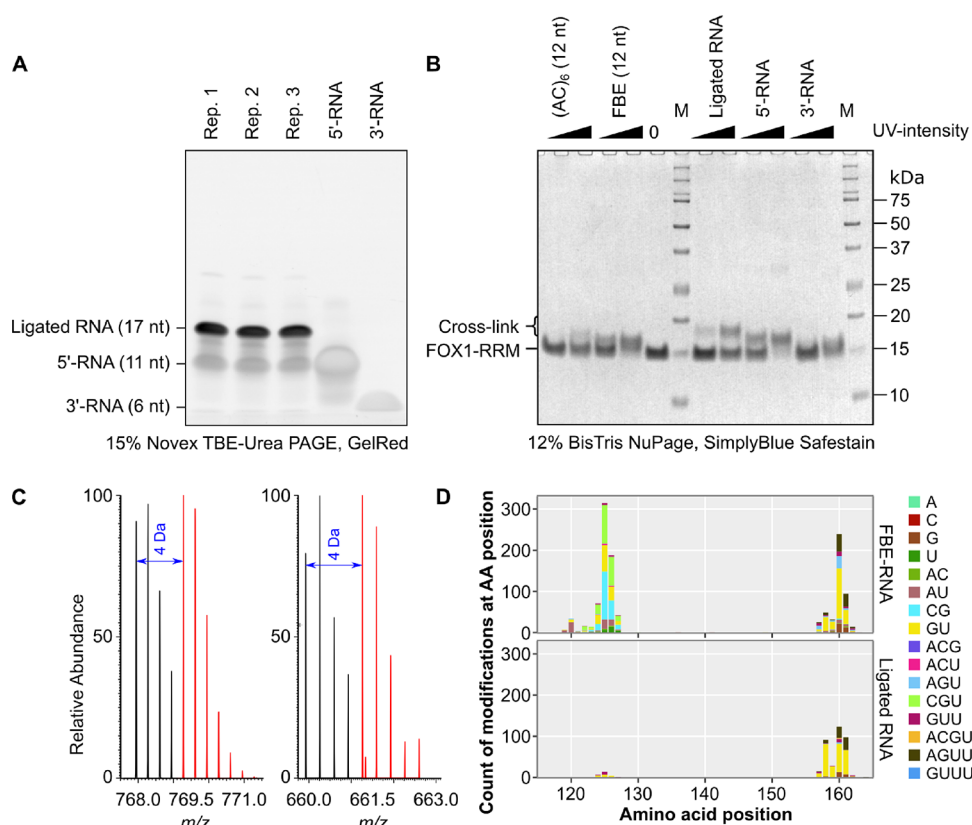
#### Single Nucleotide Resolution CLIR-MS by $^{18}\text{O}$ -Labeling.

With the post-digest labeling approach, the oligonucleotide composition of the RNA adduct after RNase digestion can be determined, but neither the nucleotide sequence nor the cross-linking position within the RNA sequence is apparent. For short RNAs, the nucleotide composition might still be sufficiently specific to a single position within the original RNA sequence, but for longer RNAs, this is unlikely to be the case. Using standard molecular biology techniques, it is possible to introduce a single  $^{18}\text{O}_2$ -labeled phosphate in almost any position within an RNA (Scheme 2). Two RNA oligonucleotides are ligated using T4 RNA ligase, where the 3′ segment is labeled at the 5′-end using ATP (light/heavy mixture) and T4-PNK. With this approach, an isotopic shift of 4 Da can be observed if the analyzed RNA–peptide adduct contains the labeled phosphate (Scheme 2C). Single-phosphate labeling can be used to generate an array of synthetic RNAs with a single labeled phosphate in different positions of a longer RNA sequence to explore the exact RNA-binding properties of an RBP.

We identified multiple cross-links to GU dinucleotides, which occur in different positions within the short RNA-sequence UGCAUGU. To distinguish those sites within the RNA, we chose to label the phosphate between G6 and U7 of the FOX1-binding element in an extended FBE-containing sequence. This

Scheme 2. Single-Phosphate Labeling<sup>a</sup>

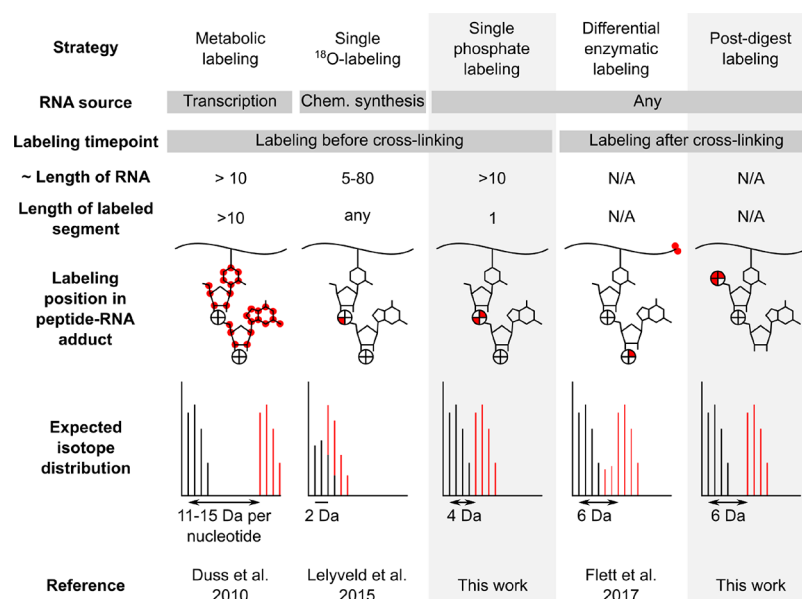
<sup>a</sup>(A) Desired product with a labeled phosphate between G6 and U7 of the extended FBE sequence. (B) A single heavy-labeled phosphate is introduced into an RNA by labeling the 5'-end of the 3'-segment (blue) with <sup>18</sup>O<sub>4</sub>-γ-ATP and subsequent ligation to the 5'-segment (green) using T4 RNA ligase. The 3'-RNA is blocked at the 3'-end with a C3-spacer to prevent undesired ligation side products. (C) Simulation of isotopic distribution<sup>33</sup> of a cross-linked peptide (PEPTIDER cross-linked to a GU dinucleotide, doubly charged) for a 1:1 mixture of light and heavy ATP using single-phosphate labeling.



**Figure 4.** Single-phosphate labeling. (A) Yield of the ligated RNA product UAAGUUGCAUGUUGUCA (see also Scheme 2). Substrate RNAs and products of three replicates were separated on a 15% Tris borate EDTA (TBE)-urea PAGE. (B) Different RNAs were cross-linked to FOX1-RRM using high and low energy (3.2 or 0.8 J/cm<sup>2</sup>). FOX1-RRM cross-links well to FBE-RNA, the 5'-RNA, and the ligated RNA. Unspecific cross-linking can be observed at higher energies for an unspecific (AC)<sub>6</sub>-oligomer and the 3'-RNA. (C) MS1 isotopic patterns for cross-links of the peptide GFGFVTFENSADADR to a GU dinucleotide (left spectrum, [M + 3H]<sup>3+</sup> = 767.94 m/z) and a G mononucleotide (right spectrum, [M + 3H]<sup>3+</sup> = 659.93 m/z). (D) Cross-link positions within the FOX1-RRM protein sequence with post-digest labeling (upper panel) and the single-phosphate labeled RNA (lower panel).

way, it is possible to locate RNA–peptide adducts to a labeled nucleotide within the RNA. For the synthesis of this RNA, we labeled the 5'-hydroxy group of the 3'-segment of the RNA (3'-RNA) with heavy/light-ATP mix and subsequently ligated the RNA to the free 3'-end of the 5'-segment of the RNA (5'-RNA) using T4 RNA ligase I. The ligation product is a 17 nt long RNA

(ligated RNA, Figure 4A). Cross-linking this RNA to the FOX1-RRM introduces a higher mass shift than the FBE-RNA or any of the two ligation substrates (Figure 4B). Two labeled oxygen atoms remain in the final product. We observed multiple distinct isotope shifts of 4.01 Da in the precursor mass spectra (Figure 4C). The identified, labeled cross-links exclusively map to the



**Figure 5.** Comparison of labeling strategies for RNA–protein cross-linking analyses.<sup>22–24</sup> Different atoms within the RNA–peptide adduct can be labeled. Positions are indicated on a schematic representation of a peptide (curved line) cross-linked to a hypothetical dinucleotide. The labeled atoms are highlighted in the schematic structure in red. Phosphates with four oxygen atoms are represented as circles with four slices. Single-phosphate labeling and post-digest labeling lead to a higher isotopic shift, which is unique to RNA–peptide adducts (as compared to single  $^{18}\text{O}$ -labeling and differential enzymatic labeling). Both techniques can be implemented for any purified RNA.

F160 region of the FOX1 protein (Figure 4D). In a search for all unlabeled adducts, both sites within the protein can be identified (Figure S2). Only cross-links at F160 show the specific 4.01 Da isotopic shifts, demonstrating that phenylalanine 160 is in contact with G6/U7 within the binding sequence and that, consequently, cross-links at phenylalanine 126 occur to the U1/G2 site within the FBE-RNA. This agrees with previous data and conclusively shows the precision of single-phosphate labeling in combination with the CLIR-MS technique.

## DISCUSSION

RNA–protein interactions play a crucial role in many biological contexts and have been studied from RNA-centric or protein-centric views for some time. With CLIR-MS, it is possible to retrieve information about the protein interaction site as well as the interacting RNA sequence at single monomer resolution from one experiment. This is achieved by differential stable isotope labeling of the RNA, whereby the labeled segment could be as long as the entire RNA or as short as a single nucleotide.

Various methods for isotope labeling of RNAs are available (see Figure 5 and Table S4). The preparation of labeled and especially segmentally labeled RNA requires special expertise. Longer RNAs can be synthesized by *in vitro* transcription using  $^{13}\text{C}/^{15}\text{N}$ -labeled NTPs.<sup>22</sup> Other techniques utilize  $^{18}\text{O}$  isotopes during chemical RNA synthesis<sup>23</sup> or sample processing.<sup>24</sup> Segmental labeling can be achieved by consecutive ligation of unlabeled and labeled RNAs.<sup>12</sup> The work presented here provides a simplified workflow for the incorporation of isotope labels into RNA, removing the requirement for specialist expertise and therefore opening the CLIR-MS approach to a broader community.

Lelyveld et al. developed a technique to specifically label a single oxygen atom in an RNA.<sup>23</sup> During solid-phase RNA synthesis with the phosphoramidite method, the backbone phosphates are oxidized. In this iodine-mediated step, water serves as the oxygen source and can therefore be provided as

either light or heavy ( $\text{H}_2^{18}\text{O}$ ) water. Using two different oxidation mixes (either light or heavy) during automated synthesis, highly site-specific labeling, similar to single-phosphate labeling, can be achieved but only by introducing a mass shift of 2 Da. For data analysis based on isotopic pairs, a 2 Da mass shift is too small as the two isotopic species largely overlap with the natural isotope distribution of carbon (Figure 5). Another approach, introduced by Flett et al., introduces heavy oxygen atoms during protease and RNase digest.<sup>24</sup> The hydrolysis of the peptide bond and phosphodiester is performed either in light or heavy-oxygen ( $\text{H}_2^{18}\text{O}$ ) water, and one heavy oxygen per cleavage site is introduced. Non-cross-linked peptides will therefore be labeled with two heavy oxygen atoms (4.01 Da) at the C-terminus and cross-linked peptides will be labeled with an additional heavy oxygen atom at the 3'-end of the RNA (6.01 Da). For both approaches, the sample must be split and labeled individually with heavy and light oxygen. Simple and affordable techniques that clearly distinguish between cross-linked and non-cross-linked peptides and allow site-specific labeling have the potential to advance RBP analysis techniques that utilize mass spectrometry.

Here, we introduce two simple labeling techniques that employ  $^{18}\text{O}_4$ - $\gamma$ -ATP to introduce labeled phosphates into the RNA. In the first technique, a single labeled phosphate is introduced in a peptide–oligonucleotide adduct after cross-linking in a one-step reaction. This is highly advantageous if the RNA cannot be labeled in advance, e.g., in cell lysates or affinity purifications. In the second labeling technique, a single phosphate within an RNA is labeled in a simple two-step enzymatic reaction prior to cross-linking, yielding a segmentally labeled RNA with a single labeled phosphate, using standard molecular biology techniques. In both cases, the label is introduced by the T4-PNK that transfers the  $\gamma$ -phosphate from ATP to the free 5'-hydroxy group of a nucleic acid. T4-PNK exhibits two reactions: (1) phosphorylation of the 5'-hydroxy group and (2) dephosphorylation of the 3'-phosphate

group (Scheme 1).<sup>25</sup> The second reaction is not required for the protocols presented here but should be considered during the data analysis step. Alternatively, T4-PNK is also commercially available without 3'-phosphatase activity. For the post-digest labeling technique, it is important to consider the choice of RNase. Only RNases or other RNA cleavage methods that result in a 5'-hydroxy product (e.g., RNase A,<sup>26</sup> RNase T1,<sup>27</sup> RNase I,<sup>28</sup> micrococcal nuclease,<sup>29</sup> or alkaline hydrolysis<sup>30</sup>) are compatible with the method. Other nucleases, like benzonase, result in a 5'-phosphate,<sup>31</sup> which cannot be labeled quantitatively without additional measures. The native 5'-ends of naturally occurring RNAs are most frequently a 5'-cap structure or a 5'-triphosphate. These modifications would interfere with post-digest labeling and need to be removed to access the very 5'-nucleotide of a natural RNA.

We used the post-digest labeling approach and single-phosphate labeling to analyze the human FOX1-RRM domain and its interaction with its substrate RNA (UGCAUGU). Among other adducts, we could identify multiple cross-links to GU dinucleotides from two different positions within the protein (F126 and F160). Currently, it is only possible to determine the composition of the RNA adduct, but not its sequence, by CLIR-MS. A GU dinucleotide therefore also corresponds to the sequence UG. Within the short 7-mer sequence UGCAUGU, the GU/UG dinucleotide occurs in three different positions. To distinguish these cross-linking sites on the RNA, we labeled the phosphate between G6 and U7 in an FBE-containing sequence and could show that the labeled GU dinucleotide cross-links to phenylalanine 160 exclusively, therefore practically achieving single nucleotide and single amino acid resolution for an RNA–protein cross-link.

Post-digest labeling is a simple method to separate RNA-cross-linked peptides from highly abundant unmodified peptides from any sample, paving the way to applying CLIR-MS to more complex samples. Like all methods that rely on protein–RNA cross-linking by UV light, it is dependent on the yield from the cross-linking step that may depend on several factors, including the affinity of the complex and specific steric/geometric properties of the interaction, such as  $\pi$ – $\pi$  stacking of amino acid side chains and nucleobases. The advantages of the post-digest labeling technique have the potential to boost RNA–protein cross-link identifications in large-scale interaction studies,<sup>11,32</sup> although with this strategy, only RNA-binding sites on the protein sequence can be deduced. With the single-phosphate labeling approach, RNA–protein interactions can be mapped to a single amino acid and nucleotide within the protein and RNA. Using multiple pairs of synthetic oligonucleotides, it would be possible to generate, with moderate effort, a set of RNAs with labels covering different binding sites to narrow down the binding interface in novel RNA–protein interactions to single nucleotide and single amino acid resolution.

The techniques presented here use the versatile nucleotide kinase T4-PNK that can also be used to phosphorylate DNA oligonucleotides. The use of T4-PNK enables post-digest labeling and single-phosphate labeling for DNA-binding proteins, using the same enzymatic reaction. Moreover, ATP-based labeling techniques could be used with other kinases (e.g., protein kinases) to specifically label substrates of this specific kinase in vitro. The two ATP-based RNA labeling techniques provide all the benefits of previously described labeling techniques without the need for specialized equipment or expertise.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.1c02384>.

3D-printable 96-well plate cooled holder for Spectro-linker XL-1500 (Figure S1); analysis of a single-phosphate-labeled RNA sample without heavy–light spectrum matching (Figure S2); and labeling strategies for RNA–peptide cross-link characterization (Table S4) (PDF)

Table of all RNA sequences and neutral losses and corresponding mass modifications that were used for xQuest analyses (Table S1) (XLSX)

Cross-links used for relative quantification and quantification values (Table S2) (XLSX)

Table of all identified cross-links extracted from xQuest result files for Figures 2, 3, and 4 (Table S3) (XLSX)

3D-printable 96-well plate cooled holder for Spectro-linker XL-1500 (ZIP)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Michael Götz** – Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zürich 8093, Switzerland; Present Address: Department of Biology, Chemistry, Pharmacy, Institute of Chemistry and Biochemistry, Freie Universität Berlin, 14195 Berlin, Germany; Email: [m.goetze@fu-berlin.de](mailto:m.goetze@fu-berlin.de)

**Alexander Leitner** – Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zürich 8093, Switzerland; [orcid.org/0000-0003-4126-0725](https://orcid.org/0000-0003-4126-0725); Email: [leitner@imsb.biol.ethz.ch](mailto:leitner@imsb.biol.ethz.ch)

### Authors

**Chris P. Sarnowski** – Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zürich 8093, Switzerland

**Tebbe de Vries** – Department of Biology, Institute of Biochemistry, ETH Zürich, Zürich 8093, Switzerland

**Anna Knörlein** – Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, ETH Zürich, Zürich 8093, Switzerland

**Frédéric H.-T. Allain** – Department of Biology, Institute of Biochemistry, ETH Zürich, Zürich 8093, Switzerland

**Jonathan Hall** – Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, ETH Zürich, Zürich 8093, Switzerland; [orcid.org/0000-0003-4160-7135](https://orcid.org/0000-0003-4160-7135)

**Ruedi Aebersold** – Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zürich 8093, Switzerland

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.1c02384>

### Author Contributions

Experimental design, M.G. and A.L.; protein production and purification, A.K. and T.d.V.; experimental work, M.G.; software and data analysis, M.G. and C.S.; writing – original draft, M.G. and A.L.; writing – review and editing, all authors.

### Notes

The authors declare no competing financial interest.



## ACKNOWLEDGMENTS

We thank Paola Picotti for access to mass spectrometry and laboratory infrastructure and Natalie de Souza for critical comments on the manuscript. This work was supported by funding from ETH Zurich (Research Grant ETH-24 16-2 to A.L., F.H.-T.A., J.H., and R.A.); the ETH Domain Strategic Focus Area "Personalized Health and Related Technologies" (TechTransfer Project PHRT-503 to A.L. and F.H.-T.A.); the European Research Council (Advanced Grant ERC-20140 AdG 670821 to R.A.); and the National Center of Competence in Research, RNA & Disease (NCCR RNA & Disease). The Orbitrap Fusion Lumos mass spectrometer used in this work was purchased using funding from the ETH Scientific Equipment program and the European Union Grant ULTRA-DD (FP7-JTI 115766).

## REFERENCES

- (1) Cech, T. R.; Steitz, J. A. *Cell* **2014**, *157*, 77–94.
- (2) Castello, A.; Fischer, B.; Frese, C. K.; Horos, R.; Alleaume, A.-M.; Foehr, S.; Curk, T.; Krijgsveld, J.; Hentze, M. W. *Mol. Cell* **2016**, *63*, 696–710.
- (3) Müller, S.; Bley, N.; Glaß, M.; Busch, B.; Rousseau, V.; Misiak, D.; Fuchs, T.; Lederer, M.; Hüttelmaier, S. *Nucleic Acids Res.* **2018**, *46*, 6285–6303.
- (4) Neumann, M.; Sampathu, D. M.; Kwong, L. K.; Truax, A. C.; Micsenyi, M. C.; Chou, T. T.; Bruce, J.; Schuck, T.; Grossman, M.; Clark, C. M.; McCluskey, L. F.; Miller, B. L.; Masliah, E.; Mackenzie, I. R.; Feldman, H.; Feiden, W.; Kretschmar, H. A.; Trojanowski, J. Q.; Lee, V. M. Y. *Science* **2006**, *314*, 130–133.
- (5) Ottesen, E. W. *Transl. Neurosci.* **2017**, *8*, 1–6.
- (6) Li, X.; Song, J.; Yi, C. *Genomics, Proteomics Bioinf.* **2014**, *12*, 72–78.
- (7) Lee, F. C. Y.; Ule, J. *Mol. Cell* **2018**, *69*, 354–369.
- (8) Nechay, M.; Kleiner, R. E. *Curr. Opin. Chem. Biol.* **2020**, *54*, 37–44.
- (9) Steen, H.; Jensen, O. N. *Mass Spectrom. Rev.* **2002**, *21*, 163–182.
- (10) Schmidt, C.; Kramer, K.; Urlaub, H. *J. Proteomics* **2012**, *75*, 3478–3494.
- (11) Kramer, K.; Sachsenberg, T.; Beckmann, B. M.; Qamar, S.; Boon, K.-L.; Hentze, M. W.; Kohlbacher, O.; Urlaub, H. *Nat. Methods* **2014**, *11*, 1064–1070.
- (12) Dorn, G.; Leitner, A.; Boudet, J.; Campagne, S.; von Schroetter, C.; Moursy, A.; Aebersold, R.; Allain, F. H.-T. *Nat. Methods* **2017**, *14*, 487–490.
- (13) Shchepachev, V.; Bresson, S.; Spanos, C.; Petfalski, E.; Fischer, L.; Rappsilber, J.; Tollervy, D. *Mol. Syst. Biol.* **2019**, *15*, No. e8689.
- (14) Urdaneta, E. C.; Vieira-Vieira, C. H.; Hick, T.; Wessels, H.-H.; Figini, D.; Moschall, R.; Medenbach, J.; Ohler, U.; Granneman, S.; Selbach, M.; Beckmann, B. M. *Nat. Commun.* **2019**, *10*, 990.
- (15) Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. *Nat. Methods* **2008**, *5*, 315–318.
- (16) Walzthoeni, T.; Claassen, M.; Leitner, A.; Herzog, F.; Bohn, S.; Förster, F.; Beck, M.; Aebersold, R. *Nat. Methods* **2012**, *9*, 901–903.
- (17) Asadi-Atoi, P.; Barraud, P.; Tisne, C.; Kellner, S. *Biol. Chem.* **2019**, *400*, 847–865.
- (18) Auweter, S. D.; Fasan, R.; Reymond, L.; Underwood, J. G.; Black, D. L.; Pitsch, S.; Allain, F. H.-T. *EMBO J.* **2006**, *25*, 163–173.
- (19) Chen, Y.; Zubovic, L.; Yang, F.; Godin, K.; Pavelitz, T.; Castellanos, J.; Macchi, P.; Varani, G. *Nucleic Acids Res.* **2016**, *44*, 4381–4395.
- (20) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egerton, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.

Deusch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A. *Nat. Biotechnol.* **2012**, *30*, 918–920.

(21) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. *Nucleic Acids Res.* **2019**, *47*, D442–D450.

(22) Duss, O.; Maris, C.; von Schroetter, C.; Allain, F. H.-T. *Nucleic Acids Res.* **2010**, *38*, No. e188.

(23) Lelyveld, V. S.; Björkbohm, A.; Ransey, E. M.; Sliz, P.; Szostak, J. W. *J. Am. Chem. Soc.* **2015**, *137*, 15378–15381.

(24) Flett, F. J.; Sachsenberg, T.; Kohlbacher, O.; Mackay, C. L.; Interthal, H. *Anal. Chem.* **2017**, *89*, 11208–11213.

(25) Cameron, V.; Uhlenbeck, O. C. *Biochemistry* **1977**, *16*, 5120–5126.

(26) Findly, D.; Herries, D. G.; Mathias, A. P.; Rabin, B. R.; Ross, C. A. *Nature* **1961**, *190*, 781–784.

(27) Takahashi, K. *J. Biochem.* **1970**, *67*, 833–839.

(28) Spahr, P. F.; Hollingworth, B. R. *J. Biol. Chem.* **1961**, *236*, 823–831.

(29) Cunningham, L.; Catlin, B. W.; De Garille, M. P. *J. Am. Chem. Soc.* **1956**, *78*, 4642–4645.

(30) Lipkin, D.; Talbert, P. T.; Cohn, M. *J. Am. Chem. Soc.* **1954**, *76*, 2871–2872.

(31) Nestle, M.; Roberts, W. K. *J. Biol. Chem.* **1969**, *244*, 5213–5218.

(32) Bae, J. W.; Kwon, S. C.; Na, Y.; Kim, V. N.; Kim, J.-S. *Nat. Struct. Mol. Biol.* **2020**, *27*, 678–682.

(33) Loos, M.; Gerber, C.; Corona, F.; Hollender, J.; Singer, H. *Anal. Chem.* **2015**, *87*, 5738–5744.