Mapping protein-RNA interactions with single residue resolution by CLIR-MS/MS

Ruedi Aebersold (aebersold@imsb.biol.ethz.ch)
Allain group
Frédéric H.-T. Allain (allain@mol.biol.ethz.ch)
Allain group
Georg Dorn
Allain group
Alexander Leitner
Julien Boudet
Sébastien Campagne
Christine von Schroetter
Ahmed Moursy

Method Article

Keywords: protein RNA interactions, RNPs, LC-MS/MS, UV-crosslinking

Posted Date: March 31st, 2017

DOI: https://doi.org/10.1038/protex.2017.031

Abstract

Protein-RNA complexes are key regulators involved in all steps of RNA metabolism and are thus crucial for cellular function. The precise knowledge of the RNA/protein interaction sites in such complexes provides essential information for understanding ribonucleoprotein \(RNP) regulated processes in health and disease. We established an efficient technique that combines segmental isotope labeling of RNA with photo-crosslinking and tandem mass spectrometry. It resolves protein-RNA interactions in RNPs assembled _in vitro_, at single amino acid and single nucleotide level resolution in the same analysis. Information on these zero-length crosslinks can be used to guide further functional assays and as intermolecular distance restraints, thus supporting atomic scale modeling of protein-RNA complexes in an integrative structural biology approach. Optimally, the mapping of protein-RNA interactions can be achieved for a given complex within three weeks. This protocol accompanies Dorn et al, Nature Methods, published online 27 March 2017 \(doi:10.1038/nmeth.4235).

Introduction

Protein-RNA complexes regulate gene expression by controlling maturation, localization, translation and degradation of RNA. Mutation of even a single nucleotide or amino acid is sufficient to alter the recognition of the binding partners and leads to disease[1, 2]. Deciphering protein-RNA interactions lays the basis for the further functional and structural characterization of cognate protein-RNA complexes and sheds light on how specific intermolecular interactions regulate associated cellular processes. UV-induced protein-RNA crosslinking has been widely used to detect target transcripts bound by selected bait proteins by crosslinking and immunoprecipitation \(CLIP) techniques with single nucleotide resolution[3, 4]. However, these techniques do not identify the nature of the interaction, the exact binding site of the protein or other functionally important proteins engaged in the RNP. Complementary to CLIP-techniques, UV-crosslinking and liquid chromatography/tandem mass spectrometry \(LC-MS/MS) have been used to identify and discover RNA binding proteins \(RBPs) bound to a specific subset of RNAs, but the exact position of the RBPs on the RNA has remained inaccessible[5-8]. Thus, the detection of protein-RNA interactions at single residue resolution has only been possible by classical de-novo structure determination methods. Here, we provide a protocol \(see Fig. 1 for an overview) that combines segmental $^{13}C^{15}N$-isotope labeling of RNA[9] with UV-crosslinking of protein-RNA complexes and LC-MS/MS. We term the method “CLIR-MS/MS” for CrossLinking of segmentally Isotope labeled RNA and MS/MS. It has the ability to decipher protein-RNA interactions at single amino acid and single nucleotide resolution, requires only small sample amounts and is in principle not limited by the size or flexibility of the investigated RNP. We successfully applied the method on the U1 small nuclear RNP \(U1snRNP) and the 58 kDa RBP Polypyrimidine Tract Binding Protein 1 \(PTBP1) in complex with a large \(88 nucleotides), structured RNA sequence derived from the Internal Ribosomal Entry Site \(IRES) of Encephalomyocarditis Virus. U1snRNP consists of 10 different proteins bound to a large, heavily structured RNA and is an essential component of the spliceosome[10]. The multifunctional PTBP1 consists of four RNA Recognition Motifs that bind primarily CU-rich RNA elements and acts e.g. as alternative

splicing factor[11] and IRES trans-acting factor[12, 13]. CLIR-MS/MS takes advantage of the use of segmentally isotope labeled RNA that is combined with equimolar amounts of unlabeled RNA of the same sequence prior to complex formation. The combined sample is then photo-crosslinked and the product is enzymatically treated with a protease to generate MS-compatible peptides and with nucleases to trim the RNA adducts crosslinked to specific peptide sequences. The resulting peptides are analyzed by LC-MS/MS. The modified peptide is sequenced by tandem mass spectrometry and the identity of the peptide as well as the modified amino acid and the composition of the RNA adduct can be deduced from the fragment ion spectra. Crosslinks that occurred at the differentially labeled segment appear as peak doublets of approximately 1:1 ratio in the precursor ion spectrum. The peak splitting corresponds to the mass difference of the RNA adduct and precursor ions corresponding to peptides crosslinked to the differentially labeled RNA segment can be identified. The minimal mass difference of 11 Da \($^{13}C^{15}N$-labeled vs. unlabeled uracil) is large enough to avoid interference with natural isotope patterns and the differential isotope labeling adds an additional selection criterion for mass spectrometric and subsequent computational fragment ion analysis, thereby reducing the false positive detection rate. Furthermore, the focus on differentially labeled RNA adducts allows to limit the range of possible interaction sites to the respective $^{13}C^{15}N$-labeled segment. In combination with the detection of mono-, di- and trinucleotide modifications occurring at the same amino acid, up to single nucleotide resolution on the RNA side can be achieved. Segmental isotope labeling of RNA is _per se_ not restricted by size and any type of labeled RNA-segment can be used. Thus, combinations of different labeling schemes like deuterated RNA segments or chemically synthesized 4-thio-uridinylated RNA are compatible with the method[9].

Reagents

A: RNA-transcription - DNA-template coding for the target RNA including a T7-promotor sequence - 1 M $MgCl_2$ solution - 80 mM nucleotide triphosphate \(NTP) solution \(for each NTP, Applichem, ATP: A1348.0005; GTP: A1803.0001; CTP: A2145.0500; UTP: A2237.0001) - 80 mM $^{13}C^{15}N$-labeled NTP solution \(commercially available or purified in-house) - 20x transcription buffer \(TB): 800 mM Tris-HCl, pH 8.1, 20 mM spermidine, 0.2% Triton X-100, 100 mM dithiothreitol \(DTT) - HPLC buffer A: 6 M urea, 12.5 mM Tris-HCl, pH 7.4 - HPLC buffer B: 6 M urea, 12.5 mM Tris-HCl, pH 7.4, 500 mM $NaClO_4$ - n-butanol **B:** Segmental isotope labeling - RNase H \(ThermoScientific, 18021071 or purified in-house) - T4-DNA-Ligase \(Fermentas, EL0014; or purified in-house) - 10x RNase H buffer: 500 mM Tris-HCl, pH 7.5, 1 M NaCl, 100 mM $MgCl_2$ - 10x T4-DNA-Ligase buffer: 400 mM Tris-HCl, pH 7.8, 5 mM ATP, 100 mM $MgCl_2$, 100 mM DTT - 50 % \(w/v) PEG-4000 - suitable 2´-O-methyl-RNA/DNA chimeras **C:** UV-crosslinking - 96-well-plate \(U-bottom; e. g. Greiner bio one, 650901) - NuPage Bis-Tris-Acrylamide gels \(Thermo Scientific, NP0322BOX) - NuPage MES-SDS running buffer \(Thermo Scientific) - NuPage 4x LDS loading buffer \(ThermoScientific, NP0007) - PageRuler Prestained Plus protein ladder \(Thermo Scientific, 26619) - γ-^{32}P-ATP \(Hartmann Analytic, SCP-401) - Amersham Protran 0.2 nitrocellulose membrane \(GE, 10600001) - ethanol **D:** Digestion and sample enrichment - 50 mM Tris-HCl, pH 7.9, 4 M urea - 50 mM Tris-HCl, pH 7.9 - 1 M $MgCl_2$ solution - RNase T1 \(ThermoFisher, EN0541) - RNase A \(Ambion, AM2271)

- Benzonase \(Sigma-Aldrich, E1014) - Sequencing Grade Modified Trypsin, Frozen \(Promega, V5113) - acetonitrile - TiO_2 beads \(5 µm Titansphere, GL Sciences) - trifluoroacetic acid - lactic acid - formic acid

Equipment

- semipreparative HPLC system with column oven \(85 °C) and water bath \(88 °C) for buffer preheating - semipreparative anion exchange column \(DNAPac PA-100, Thermo Scientific, SP2091) including guard column \(SP4513) - lyophilization device - XCell SureLock Mini-Cell Electrophoresis System \(Thermo Scientific, EI0001) - TransBlot SD Semi Dry Electrophoretic Transfer Cell \(BioRad 170-3940) - SepPak tC18 cartridges \(Waters) - Thermomixer/shaker \(e.g. Eppenforf ThermoMixer) - benchtop centrifuge - vacuum concentrator - LC-MS/MS system, e.g. Easy nLC 1000 HPLC system \(ThermoFisher Scientific) connected to an Orbitrap Elite mass spectrometer \(ThermoFisher Scientific) equipped with a Nanoflex electrospray source - xQuest software, available from proteomics.ethz.ch

Procedure

A: RNA transcription \(3 days) 1) Optimize in-vitro transcription \($MgCl_2$, T7-RNA-polymerase) in 50 µL test reactions for unlabeled and $^{13}C^{15}N$-labeled NTPs, monitor transcription efficiency using urea-PAGE, a typical reaction condition contains in 1x TB: 1.7 µM T7-RNA-polymerase, 33 µg/mL linearized plasmid as DNA template, 4.5 mM of each NTP, 42.5 mM $MgCl_2$ \(day 1) 2) _in-vitro_ transcription of RNA using T7-RNA Polymerase, $^{13}C^{15}N$ labeled nucleotide triphosphates \(NTPs) and unlabeled NTPs \(day 2) 3) Purification of both RNAs \($^{13}C^{15}N$ labeled and unlabeled) by denaturing anion exchange chromatography followed by n-butanol extraction, snap cooling for refolding of secondary structures, lyophilization \(day 2) 4) Resuspension of RNA in RNase H buffer \(50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$), optionally final purification of RNA using size exclusion chromatography \(day 3) **B: Segmental isotope labeling of RNA \(3 days)** 1) Optimization of RNase H cleavage using 15 µL of 33 µM RNA, 800 nM RNase H and various amounts of 2´-O-methyl-RNA/DNA chimera \(e.g. chimera:RNA = 1:50, 1:10, 1:5, 1:2, 1:1), analysis on urea-PAGE. In case of suboptimal cleavage, try annealing of the chimera \(day 4) _Critical step: depending on the RNA sequence, a chimera:RNA ratio of 1:1 including an annealing step prior to the RNase H cleavage reactions can be required. Monitor a broad range of chimera:RNA ratios_ 2) Upscaling of optimal condition to 750 µL reactions \(RNase H amount scales approximately 8x less, final condition: 100 nM RNase H) \(day 5) _Critical step: as RNase H levels do not scale linearly with the increase of volume rather prepare several 750 µL reactions to achieve desired amounts_ 3) Purification of RNA fragments by denaturing anion exchange chromatography, n-butanol extraction, lyophilization and resuspension of fragments in double distilled water \(day 5) 4) Optimization of the ligation reaction for T4-DNA-Ligase concentration \(10-50 U per nmol RNA, Weiss-units; 0.1 -0.5 mg/mL of in-house produced T4-DNA-Ligase), reaction time \(2-6 hours) and crowding reagents \(0-20% PEG-4000) in 20 µL test reactions containing 10 µM of each fragment annealed to 1 to 1.2-fold excess of DNA splint in 1x T4-DNA Ligase Buffer \(40 mM Tris-HCl, pH 7.8, 0.5 mM ATP, 10 mM $MgCl_2$, 10 mM DTT) \(day 6) 5) Large-scale \(20-50 nmol RNA-fragments) ligation followed by

purification using denaturing anion exchange chromatography, n-butanol extraction, snap cooling for refolding of secondary structures, lyophilisation and resuspension in suitable buffer \(day 6) _Critical step: The exact determination of the RNA-fragment concentrations and of the DNA-splint concentration are crucial as any deviation from the optimal 1:1:1 \(fragment:fragment:splint) ratio reduces the amount of productively annealed reactants._ **C: Complex formation and UV-crosslinking \(2 days)** 1) Mixing of unlabeled RNA and segmentally $^{13}C^{15}N$-labeled RNA at 1:1 ratio, complex formation and complex purification \(day 7) 2) Optimization of UV-Dose \(device: Stratalinker 1800; UV-dose: 2.4, 3.2, 4.0, 4.8, 5.6, 6.4, 7.2 J/cm^2 as monitored by the built-in detector, sample to bottom distance: 12 cm) with 50 µL per condition of 1 mg/mL unlabeled protein-RNA complex in one well of a 96-well-plate \(PS, U-bottom, non-binding, clear; Greiner bio one). Sample should be kept on ice and should be cooled for 1 min after the application of a UV dose of 0.8 J/cm^2. Analysis by NuPage Bis-Tris-Acrylamide gels \(Coomassie and silver staining, supershift of the protein band indicates crosslinking, monitor UV-induced damage) and Western blotting on nitrocellulose membrane after digestion and end-labeling of RNA with γ-^{32}P-ATP \(in parallel to day 6, no segmentally labeled RNA needed for this step) 3) Large-scale irradiation \(250 µg of each sample), ethanol precipitation \(day 8) **D: Digestion and sample enrichment \(2 days)** 1) Resuspension of pellet in 50 µL of 50 mM Tris-HCl, pH 7.9, 4 M urea, dilution with 150 µL 50 mM Tris-HCl, pH 7.9 \(day 9) 2) Addition of 5 U RNase T1 and 5 µg RNase A per mg of crosslinked sample, incubation at 52 °C for 2 h \(day 9) 3) cool down of digest on ice, addition of 2 µL 1 M $MgCl_2$ solution and 125 U benzonase per mg of crosslinked sample, digestion continued at 37 °C for 1.5 h \(day 9) 4) Addition of 24:1 protein:enzyme-ratio \(w/w) of trypsin, incubation overnight at 650 rpm and 37 °C \(day 9) 5) Trypsin inactivation at 70 °C for 10 min and addition of 100 U benzonase, 4 U RNase T1 and 4 µg RNase A per mg of crosslinked sample for completion of the RNA digestion for 1 h at 37 °C \(optional) \(day 10) 6) Purify/desalt samples by solid-phase extraction using Waters SepPak tC18 cartridges \(day 10) _Enrichment of peptide-RNA adducts:_ 7) Preequilibrate 5 mg of TiO_2 beads in resuspension buffer \(50 % ACN, 0.1 % TFA, 300 mM lactic acid), spin down at 16 100 g for 2 min and discard the supernatant \(day 10) 8) Resuspend dried samples in 100 µL resuspension buffer, add to the TiO_2-beads, incubate for 10 min on a thermomixer \(1 400 rpm) and spin down as above \(day 10) 9) Wash beads with resuspension buffer as described above \(day 10) 10) Wash beads as described above with 50 % acetonitrile \(ACN), 0.1 % trifluoroacetic acid \(TFA) \(day 10) 11) Elute protein-RNA crosslinks with 50 mM ammonium phosphate, pH 10.5 and repeat solid phase extraction \(steps 6) \(day 10) **E: LC-MS/MS analysis \(at least 1 day, depending on the number of samples)** 1) Resuspend samples in water/acetonitrile/formic acid \(95:5:0.1, v/v/v) \(day 11) 2) Analyze an aliquot by LC-MS/MS. Recommended acquisition settings depend on the actual instrument used, but follow a typical data-dependent acquisition method for unmodified peptides or phosphopeptides. \(day 11) **F: Data analysis \(at least 1 day, depending on the number of samples)** 1) Convert raw files from the mass spectrometers into mzXML format and search against target protein sequence using xQuest. Detailed instructions for the general use of xQuest are provided by Leitner et al.[14] \(day 12) 2) Repeat xQuest search for all possible nucleotide adducts by specifying the appropriate mass shifts in the xmm.def and xquest.def configuration files of xQuest \(day 12). Different RNA modifications detected on the same

amino acid localize the crosslink within a larger RNA sequence and allow single nucleotide resolution of the interaction site \(see Figure 2).

Timing

The time required for the successful application of CLIR-MS/MS crucially depends on the target RNA, the number of segments to be labeled and the efficiency of the RNP reconstitution. For the complex of PTBP1-EMCV^{DElinkF} \(four segments), we established the segmental isotope labeling including the crosslinking in about one month. The measurement time corresponds to the number of labeled segments and the desired number of technical replicates and ranges typically from several hours to one day. The data analysis procedure is currently only partially automated and requires some expertise in the manual assessment of MS/MS spectra.

Troubleshooting

Degradation of RNA: In-house purified reagents and enzymes should be tested for RNase activity before use. Splints that cover single-stranded RNA parts during ligation reduce degradation at this site[9]. **Inefficient RNase H cleavage:** Fragments can be produced from any precursor RNA that includes the correct sequence or could even be chemically synthesized. Generally, a successful RNA ligation depends crucially on the presence of a 5´-phosphate and a 3´-OH at the reactants and should be taken into account if alternatives for fragment generation are used. **Inefficient RNA ligation:** Other enzymes such as T4-RNA-Ligase could be tested. Unreacted fragments and the DNA splint can be re-purified by denaturing anion exchange chromatography. **Poor crosslinking efficiency:** Crosslinking efficiency depends on the amount of complex present in solution. Thus, crosslinking of low-affinity complexes could benefit from increased concentrations.

Anticipated Results

CLIR-MS/MS is a very flexible approach and can be used for a crude mapping in a screening experiment that will result in the localization of the interacting RNA segment and the crosslinked amino acids. This would allow further refinement of segmental isotope labeling \(smaller segments) or for mutagenesis studies. Single nucleotide and single amino acid resolution can be generally expected if the composition of the detected RNA adducts is unique throughout the labeled RNA segment. This in turn means that the design of the RNA segments should take the RNA primary sequence into account and will benefit from prior knowledge about potential binding preferences of the protein. For all PTBP1-RRMs, we could detect mono- and di-nucleotide adducts and in favorable cases even tri-nucleotide modifications. Thus, the sequence information from the CLIR-MS/MS data covers usually up to four nucleotides and in favorable cases up to six nucleotides. CLIR-MS/MS derived protein-RNA restraints can provide crucial information for the refinement of low-resolution structures and can be combined with other restraints in an integrative structural biology approach.

References

1. Cooper, T.A., Wan, L. & Dreyfuss, G. RNA and disease. _Cell_ **136**, 777-793 \(2009). 2. Kim, E. et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. _Cancer Cell_ **27**, 617-630 \(2015). 3. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. _Cell_ **141**, 129-141 \(2010). 4. Konig, J. et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. _Nat Struct Mol Biol_ **17**, 909-915 \(2010). 5. Kramer, K. et al. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. _Nat Methods_ **11**, 1064-1070 \(2014). 6. Kwon, S.C. et al. The RNA-binding protein repertoire of embryonic stem cells. _Nat Struct Mol Biol_ **20**, 1122-1130 \(2013). 7. Beckmann, B.M. et al. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. _Nat Commun_ **6**, 10127 \(2015). 8. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. _Cell_ **149**, 1393-1406 \(2012). 9. Duss, O., Diarra Dit Konte, N. & Allain, F.H. Cut and paste RNA for nuclear magnetic resonance, paramagnetic resonance enhancement, and electron paramagnetic resonance structural studies. _Methods Enzymol_ **565**, 537-562 \(2015). 10. Will, C.L. & Luhrmann, R. Spliceosome structure and function. _Cold Spring Harb Perspect Biol_ **3** \(2011). 11. Kafasla, P. et al. Defining the roles and interactions of PTB. _Biochem Soc Trans_ **40**, 815-820 \(2012). 12. Spriggs, K.A., Bushell, M., Mitchell, S.A. & Willis, A.E. Internal ribosome entry segment-mediated translation during apoptosis: the role of IRES-trans-acting factors. _Cell Death Differ_ **12**, 585-591 \(2005). 13. Romanelli, M.G., Diani, E. & Lievens, P.M. New insights into functional roles of the polypyrimidine tract-binding protein. _Int J Mol Sci_ **14**, 22906-22932 \(2013). 14. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. _Nat Protoc_ **9**, 120-137 \(2014).

Acknowledgements

We thank Dr. V. Panse for using his UV-device for crosslinking, Dr. N. Diarra dit Konte and Dr. O. Duss for advice concerning segmental isotope labeling of RNA. G. D. was supported by a PhD fellowship of the Boehringer Ingelheim Fond. This work was supported by the SNF project \(31003A_149921) and the SNF-NCCR RNA and Disease to F. A. and by ERC grants Proteomics v3.0 \(AdvG grant #233226) and Proteomics4D \(AdvG grant #670821) to R. A.

Figures

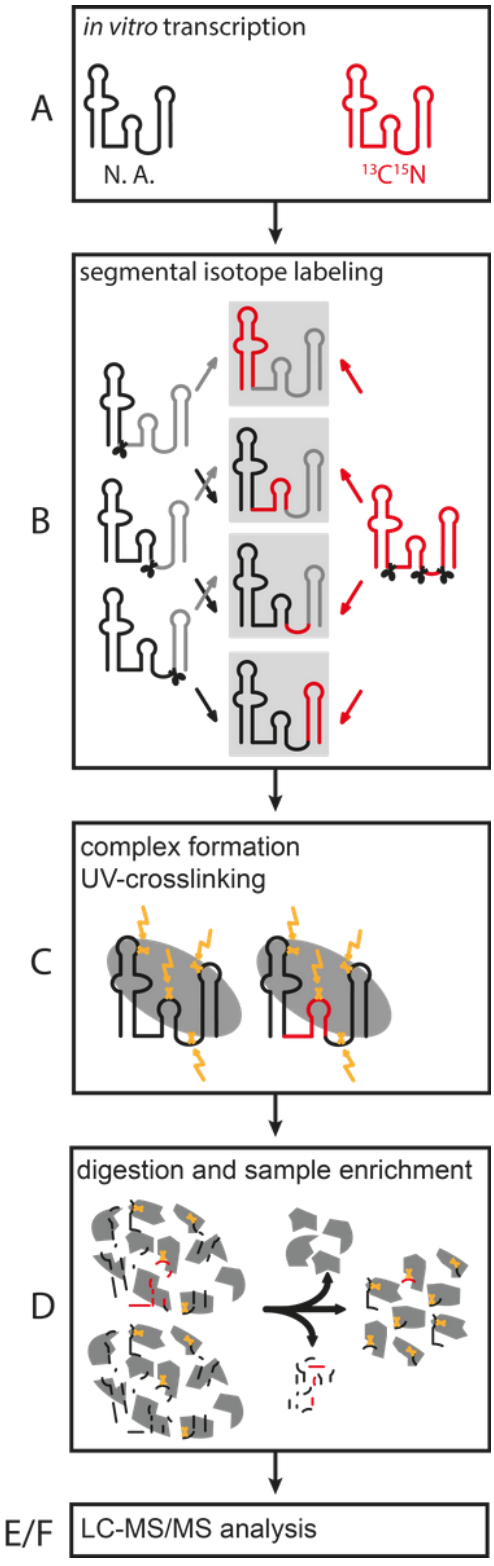

Figure 1

Flowchart describing the CLIR-MS/MS procedure The individual steps in the protocol are indicated, N.A. = natural abundance.

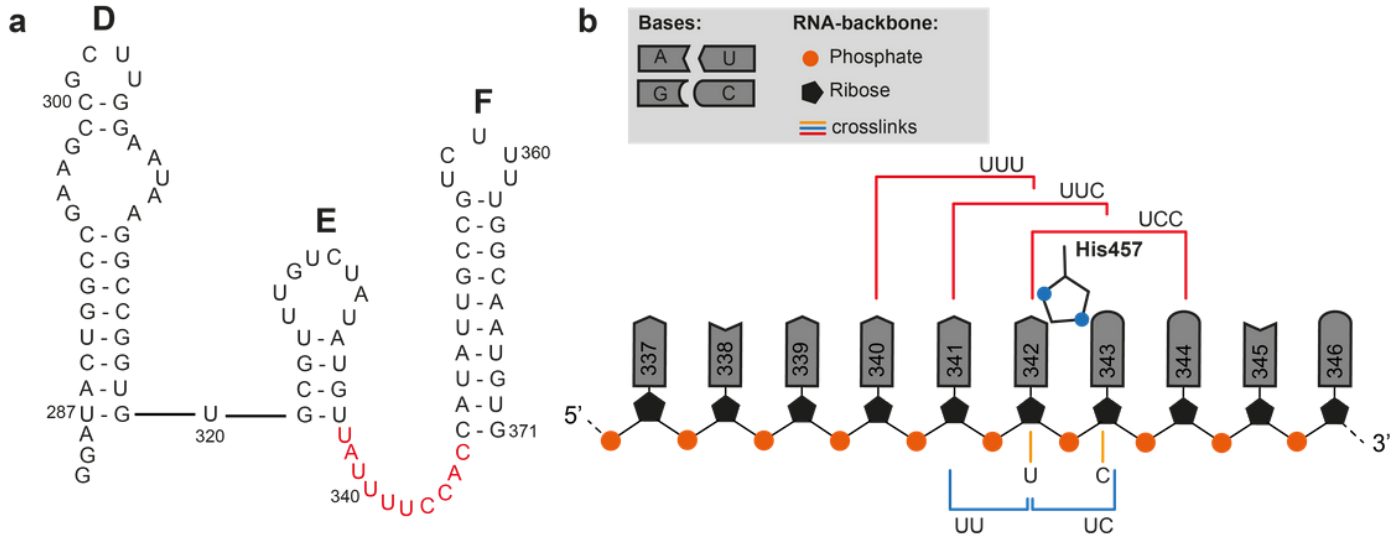

Figure 2

Mapping of the protein binding site on the RNA *a,* overview of the IRES-RNA sequence, the isotopically labeled segment “Link” depicted in red. *b,* schematic representation of the isotopically labeled segment “Link” to which PTBP1-RRM4 was exclusively crosslinked. His 457 was modified with U, C, UU, UC, UUU, UUC, and UCC which allows to position this amino acid precisely between U342 and C343 as only this position explains all detected crosslinks.