Nat Methods. Author manuscript; available in PMC 2017 September 27.

Published in final edited form as:
Nat Methods. 2017 May ; 14(5): 487–490. doi:10.1038/nmeth.4235.

Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope labeled RNA and MS/MS

G. Dorn[#1], **A. Leitner**[#2], **J. Boudet**[1], **S. Campagne**[1], **C. von Schroetter**[1], **A. Moursy**[1], **R. Aebersold**[2,3,#], and **F. H.-T. Allain**[1,#]

[1]Department of Biology, Institute for Molecular Biology and Biophysics, ETH Zürich, 8093 Zürich, Switzerland [2]Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland [3]Faculty of Science, University of Zurich, Zürich, Switzerland

[#] These authors contributed equally to this work.

Abstract

Ribonucleoproteins (RNPs) are key regulators of cellular function. We established an efficient approach that combines segmental isotope labeling of RNA with photo-crosslinking and tandem mass spectrometry to localize protein-RNA interactions simultaneously at amino acid and nucleotide resolution. The approach was tested on Polypyrimidine Tract Binding Protein 1 and U1 small nuclear RNP and the results support integrative atomic-scale structural modeling thus providing mechanistic insights into RNP regulated processes.

[#]corresponding authors: F. H.-T. A.: allain@mol.biol.ethz.ch; R. A.: aebersold@imsb.biol.ethz.ch.

Primary accession Codes
PRIDE partner repository dataset identifier: PXD005566
AddGene repository, plasmid ID: #89153, #89154, #89155, #89156, #89157

Referenced accession codes
Protein Data Bank: 2AD9, 2ADB, 2ADC, 2N3O

Data availability
CLIR-MS/MS is a new approach and thus not listed as an accepted experimental method for structure generation by the wwPDB. Therefore, CLIR-MS/MS models are not deposited but coordinate files are available as Supplementary Data (CLIR-MS/MS models: Supplementary Data 1-4, NMR-derived model: Supplementary Data 5). Restraints used for modeling are listed in the Supplementary Table 4. The mass spectrometry data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository26 with the dataset identifier PXD005566 and Source Data File 1 defines the datasets used for Fig. 2, Fig. 3 and the intermolecular CLIR-MS/MS-derived restraints. Plasmids are deposited in AddGene (ID #89153, #89154, #89155, #89156, #89157).

Author Contributions
All authors contributed to the design of experiments, the discussion of the results and the writing of the manuscript. C. v. S. purified isotope labeled nucleotides. G. D. and A. M. optimized crosslinking conditions. G. D. and S. C. prepared U1snRNP samples. G. D. prepared all other samples, performed and analyzed NMR experiments. A. L. and G. D. performed sample preparation for MS analysis, A. L. performed all LC-MS/MS experiments and analyzed the LC-MS/MS results with support from G. D.. J. B. set up the modeling protocol and modeled the RRM2-EMCV^{F} complex based on crosslinking data, G. D. implemented the modeling protocol and modeled the complexes of RRM1, RRM3 and RRM4.

Competing financial interests
The authors declare no competing financial interest

Europe PMC Funders Author Manuscripts

RNPs regulate crucial cellular functions such as gene expression, and even single nucleotide mutations can alter RNA-protein interactions with fatal consequences1. Similarly, single amino acid mutations in RNA-binding proteins (RBPs, e.g. SRSF2) are sufficient to change binding specificity and cause disease (e.g. myelodysplasia2). Deciphering protein-RNA interactions at single amino acid and nucleotide resolution would therefore provide the basis for further functional characterization of RNPs and would support integrated modeling, similar to mass spectrometric (MS) analysis of chemically crosslinked protein-protein complexes3. Photo-crosslinking and liquid chromatography (LC) MS/MS analysis have been used to identify RBPs bound to a specific subset of RNAs, but the exact position of the proteins on the RNA remained inaccessible4–6. Recently, Lelyveld et al.7 specifically mass-labeled single uridines with ^{18}O to demonstrate, by MS/MS, that Lin28A crosslinks to U11 and not U12 in a synthetic 25 nucleotide (nt) let7-pre-miRNA stem-loop. However, this approach is strongly limited as chemical RNA synthesis is restricted to short RNAs, ^{18}O labeled phosphates can detach and the small mass shift overlaps with natural isotope patterns complicating data analysis. Here, we introduce a broadly applicable approach that precisely identifies the RNA interface of an RBP and its localization on the target RNA at a resolution sufficient to support 3D modeling of RNPs. We applied it to an 85 kDa complex of the Polypyrimidine Tract Binding Protein 1 (PTBP1) with a natural RNA target and demonstrated its general applicability on U1 small nuclear RNP (snRNP). PTBP1 is a key alternative splicing factor and a major Internal Ribosomal Entry Site (IRES) *trans*-acting factor of several cellular and viral mRNAs8, 9. This 58 kDa RBP contains four RNA Recognition Motifs (RRM) whose structures in complex with a small single-stranded CUCUCU motif were determined by Nuclear Magnetic Resonance (NMR) spectroscopy10 and revealed base-specific recognition of CU or UC dinucleotides by each RRM. However, the recognition of guanines11 by PTBP1 and cooperative binding of all four RRMs to a large and structured RNA remain unexplained. Here, we have used MS and NMR spectroscopy to study PTBP1 in complex with a structured RNA molecule (88 nt) consisting of domains D-F of the IRES of Encephalomyocarditis Virus (EMCV; referred to as $EMCV^{DElinkF}$ as it includes domains D, E, the linker E-F and domain F; other RNA constructs are referred to accordingly). This IRES part binds all four RRMs of PTBP1 (Fig. 1a) and is essential for the regulatory function of PTBP1 in translation initiation12, 13.

To determine precisely the binding interface by MS, we UV-crosslinked a PTBP1-$EMCV^{DElinkF}$ complex containing equimolar ratios of unlabeled and fully $^{13}C^{15}N$-labeled RNA. Thus, protein-RNA crosslinks appear in the precursor ion mass spectrum as doublets separated by a mass shift that corresponds to the attached, differentially labeled nucleotide(s) (e.g. 11 Da for uracil, Fig. 1b and Supplementary Table 1). Modified peptides and the modifications themselves are unambiguously identified by the software xQuest14, which uses the isotope labeling information to reduce false positive assignments and to improve the identification process (Supplementary Note 1). Tandem mass spectrometry reveals the sequence of the crosslinked peptide, the modification site and the composition of the nucleotide adduct (Fig. 1c). Because long RNA adducts complicate peptide sequencing due to unfavorable fragmentation properties4, we treated the crosslinked RNP with a specific protease (trypsin) and unspecific nucleases to generate peptides with short

nucleotide chains as adduct. Peptide-nucleotide adducts were enriched prior to LC-MS/MS analysis (Online Methods and Supplementary Note 2).

We identified 22 U- and UU-modified amino acids representing 12 different peptides that belong to all four PTBP1 RRMs (Supplementary Table 2). 19 modification sites were found in close proximity (within 5 Å) to the RNA according to previously reported structures of the individual RRMs10, 3 modified residues correspond to a region that was affected upon binding of a long single stranded RNA to RRM3415. Non-irradiated control samples exhibited considerably fewer peaks in the ion chromatogram and no detectable nucleotide adducts (Fig. 1d). The multitude of UU-dinucleotides in $EMCV^{DElinkF}$ rendered the localization of the RRMs on the RNA impossible. However, the same analysis conducted on a smaller complex consisting only of RRM1 and the shorter sequence of $EMCV^{E}$ led to a unique localization of the domain at nucleotides 324-326 of the loop (Supplementary Fig. 1). To reduce the mapping possibilities for the full length PTBP1 in complex with $EMCV^{DElinkF}$, we combined the use of heavy isotopes for MS3,16–19 with the established method of segmental labeling of RNA20. We prepared four segmentally isotope labeled $EMCV^{DElinkF}$. Each RNA contained either stem-loop (SL) D (called “D”), SLE (“E”), the linker (“Link”) or SLF (“F”) in $^{13}C^{15}N$-labeled form, while the other parts of the RNA remained unlabeled (Fig. 1a, Supplementary Fig. 2 and Supplementary Note 3). These RNAs were then mixed with equimolar amounts of fully unlabeled RNA, complexed with PTBP1, UV-crosslinked, enriched and analyzed by LC-MS/MS (Fig. 1e, Online Methods and Supplementary Protocol). This way, crosslinks detected by their split isotope patterns in the precursor ion spectrum can only reside in the segmentally isotope labeled part. We named this new approach CLIR-MS/MS for **C**ross**L**inking of segmentally **I**sotope labeled **R**NA and tandem **M**ass **S**pectrometry.

We extracted (semi-)quantitative information from the MS data by spectral counting21. RRM2 and RRM4 crosslinked exclusively to “F” and “Link”, respectively, and RRM1 and RRM3 crosslinked preferentially to “E” and “D”, respectively (Supplementary Note 4). We reproduced the results for “D” in an independent experiment (Supplementary Table 2). Tyr127 (RRM1), Tyr267 (RRM2), His411 (RRM3) and His457 (RRM4) were the most frequent modification sites (Fig. 2a and Supplementary Table 2). Based on all detectable di- and trinucleotide modifications, we could precisely map RRM3 to nucleotides 5´-$U_{303}U_{304}$-3´ of SLD and RRM4 to 5´-$U_{341}UCC_{344}$-3´ of the linker E-F (Fig. 2b and Supplementary Note 5). Contrary to a previous low-resolution model12, our data suggest that RRM1 binds to 5´-$G_{323}UUUGU_{328}$-3´ of SLE and RRM2 to 5´-$C_{358}UUUUG_{363}$-3´ of SLF, which we confirmed independently by NMR experiments. Isolated RRM1 and RRM2 can both bind $EMCV^{E}$ and $EMCV^{F}$ (Supplementary Fig. 3) but in presence of both RRMs, RRM2 occupies the loop of $EMCV^{F}$ (Supplementary Fig. 4) as indicated by the overlapping chemical shifts. When superimposing the NMR spectra of RRM1-$EMCV^{E}$, RRM2-$EMCV^{F}$ and RRM34-$EMCV^{DElink}$, we can reproduce the spectra of full length PTBP1 in complex with $EMCV^{DElinkF}$ (^{1}H-^{13}C HMQC with ILV-methyl group labeling and ^{1}H-^{15}N-TROSYs, see Fig. 2c and Supplementary Fig. 5), demonstrating identical binding of the RRMs in the subcomplexes. Lastly, adding RRM1 to RRM34-$EMCV^{DElink}$ leaves the signals of RRM34

unchanged while those of RRM1 correspond to the RRM1-EMCV^{E} complex (Supplementary Fig. 6).

Taking advantage of the high-resolution protein-RNA interaction mapping, we used the identified crosslinks as intermolecular distance restraints and combined them for structural modeling with restraints derived from available structural data of PTBP1-RRMs10 (pdb: 2N3O for RRM1) and from RNA structure predictions (Supplementary Fig. 7, Supplementary Tables 3 and 4 and Supplementary Note 5). Interestingly, all except one (Ile128-AU) CLIR-MS/MS distance restraints were fulfilled by a single conformation for each RRM (Supplementary Fig. 8, coordinates are provided in Supplementary Data 1-4) with RRM1, RRM2, RRM3 and RRM4 recognizing $G_{329}UC_{331}$, $C_{358}UUU_{361}$, $U_{302}UG_{304}$ and $C_{343}C_{344}$ of EMCV^{DElinkF}, respectively (Supplementary Fig. 9). The novel recognition of a G in *syn*-conformation instead of C by RRM1 is the only possibility that is in agreement with the detected UU-adducts on Tyr 127. Recognition of Gs by PTBP1 had been suggested previously11 and earlier work on SRSF2-RRM has shown that a *syn*-G can effectively replace an *anti*-C22 with almost identical interactions. The accommodation of a U instead of a C by RRM3 is based on direct experimental evidence (see Supplementary Note 5). These binding registers indicate that the secondary structure context influences the location of the RRMs because CU-motifs reside in close proximity within SLD and SLE. Independently, we determined a high-resolution model of the RRM2-EMCV^{F} complex using classical NMR structure determination (coordinates in Supplementary Data 5). Strikingly, the binding register found in both models is the same (Supplementary Fig. 9 and 10) demonstrating the great precision and accuracy of CLIR-MS/MS based modeling.

To demonstrate the applicability of CLIR-MS/MS to larger RNPs, we reconstituted U1snRNP with either SL12 or SL34 segmentally isotope labeled (Fig. 3a). U1snRNP consists of a structured RNA bound by 10 proteins and initiates splicing by recognizing the 5’ splice site of a pre-messenger RNA23. We detected crosslinks with the zinc finger of SNRPC (also known as U1-C), the RRMs of SNRPA and SNRP70 (also known as U1-A and U1-70K, respectively) and with SNRPD2 and SNRPG (also known as Sm-D2 and Sm-G, respectively) that are all compatible with previously published structures24, 25(Fig. 3b).

In summary, CLIR-MS/MS revealed the first precise structural arrangement of PTBP1 with one of its natural RNA targets, the exact binding registers of its RRMs and the recognition of single stranded guanine-containing pyrimidine tracts embedded in stem-loops. CLIR-MS/MS reports on direct contacts and provides valuable intermolecular restraints for integrated structural biology. It requires no chemical modifications and thus minimizes the risk of artefacts. This approach is not restricted to RRM-containing proteins and not limited by size, solubility or crystallizability (see also Supplementary Note 6). Thus, it is applicable to any RNP of interest to elucidate protein-RNA interactions and to generate and refine precise structural models of such RNPs. This approach extends the application range of crosslinking-MS derived data in hybrid 3D structure determination from protein-protein complexes to protein-RNA complexes. We expect a wide-range application of the method to more complex systems such as *in vitro* reconstituted multicomponent RNPs.

Online Methods

A detailed step-by-step instruction for CLIR-MS/MS is provided as Supplementary Protocol and is accessible from Nature Protocol Exchange.

Protein expression and purification

The coding sequences of PTBP1-RRM1 (residues 41 to 163 of PTBP1) and PTBP1-RRM2 (residues 178 to 317) were cloned in pTYB11 (New England Biolabs, NEB), those of PTBP1-RRM12 (residues 41 to 317), PTBP1-RRM34 (residues 324 to 531) and full-length PTBP1 in pET28a (Novagen). Cys250 and Cys251 were mutated to Ser in all constructs. All plasmids were sequenced and transformed in BL21-Codon Plus (DE3)-RIL cells (Agilent Technologies) for protein expression. All proteins were expressed overnight at 20 °C after induction at an OD_{600} of 0.6-0.8 with 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG). For crosslinking and LC-MS/MS-analysis, cells were grown in LB-medium (DIFCO TM LB-Broth, MILLER, Fisher Scientific). For NMR-studies, we expressed proteins in M9-minimal medium containing $^{15}NH_4Cl$ and either (i) 99% D_2O (Sigma-Aldrich) and D-glucose or (ii) 99% D_2O (Sigma-Aldrich) and D-glucose supplemented with (U-$^{13}C_5$; 3-D_1) α-ketoisovaleric acid (100 mg/L) and (U-$^{13}C_4$; 3,3-D_2) α-ketobutyric acid (60 mg/L; both Cambridge Isotope Laboratories, CIL) 1 h prior to induction27, 28 or (iii) 90-99 % D_2O and D-glucose-$^{13}C_6$ (CIL) or (iv) H_2O and D-glucose-$^{13}C_6$ leading to either (i) ^{15}N-labeled, deuterated or (ii) uniformly ^{15}N-labeled, deuterated, U-^{13}C-Ile-δ_1-Leu-δ_1-δ_2-Val-γ_1-γ_2-^1H protein or (iii) $^{13}C^{15}N$-labeled, partially deuterated or (iv) $^{13}C^{15}N$-labeled, protonated proteins. Cells were lysed using a microfluidizer (Microfluidics). Clarified lysates containing PTBP1-RRM1 or PTBP1-RRM2 were purified using chitin resin (NEB) with 100 mM sodium phosphate, pH 8, 1 mM ethylene diamine tetra-acetic acid (EDTA), protease inhibitor cocktail (cOmplete EDTA-free, Roche) and 1 M NaCl as lysis buffer, or 1.5 M NaCl as wash buffer, or 200 mM NaCl and additionally 50 mM dithiothreitol (DTT) as cleavage buffer. On-column intein cleavage was performed for 48 h at 4 °C. Eluates were concentrated and applied to size exclusion chromatography (SEC, HiLoad 16/60 Superdex 75 pg, GE) with 200 mM NaCl, 50 mM sodium phosphate, pH 6.5. Fractions were tested by SDS-PAGE for purity, pooled and buffer exchanged in centrifugal concentrators (Vivaspin, Sartorius) to a final NMR buffer containing 10 mM sodium phosphate, pH 6.5, 20 mM NaCl.

Clarified lysates containing PTBP1-RRM12, PTBP1-RRM34 or PTBP1 were loaded on Ni-NTA-agarose beads (Qiagen) using 50 mM sodium phosphate, pH 8, 1 M NaCl, 7 mM imidazole and protease inhibitor cocktail (cOmplete EDTA-free, Roche) as lysis buffer and step-wise washed (10, 20 and 40 mM imidazole) and eluted (60, 80, 100 and 200 mM imidazole). Proteins were dialyzed overnight in NMR buffer, concentrated and cleaved with thrombin (0.3-1 NIH units per mg) overnight at 4 °C to remove the hexa-His-tag. Samples were further purified by cation exchange chromatography (CEX; HiTrap SP HP 5 mL, GE) and finally by SEC (HiLoad 16/60 Superdex 200 pg for PTBP1, else HiLoad 16/60 Superdex 75 pg, both GE) using 10 mM sodium phosphate, pH 6.0, 20 mM NaCl for SEC and binding to the CEX column. Proteins were eluted from the CEX by a linear gradient (0-100 %) with 10 mM sodium phosphate, pH 6.0, 1 M NaCl. Fractions were tested for

purity by SDS-PAGE, pooled and concentrated. All protein concentrations were determined by their absorption at 280 nm and their theoretical extinction coefficients calculated using the ExPASy tool ProtParam29. PTBP1 samples were always kept with 1 mM DTT except for thrombin digestion. U1A, U1-70K, U1-C and SmB, SmD1, SmD2, SmD3, SmE, SmF and SmG were expressed and purified as described previously25.

RNA *in vitro* transcription and purification

Large RNAs like EMCV^{DElinkF} (nucleotides 287-371, 5´-GGAUACUGGC CGAAGCCGCU UGGAAUAAGG CCGGUGUGCG UUUGUCUAUA UGUUAUUUUC CACCAUAUUG CCGUCUUUUG GCAAUGUG-3´), EMCV^{DElink} (nucleotides 287-346, 5´-GGAUACUGGC CGAAGCCGCU UGGAAUAAGG CCGGUGUGCG UUUGUCUAUA UGUUAUUUUC CAC-3´), HH-U1-SL12 (5´-gggaucaggu aaguauccug aaguauccug augaguccgu gaggacgaaa cgguacccgg uaccgucGAU ACUUACCUGC AGGGGAGAUA CCAUGAUCAC GAAGGUGGUU UUCCCAGGGC GAGGCUUAUC CAUUGCACUC CGGAUGUGCU GACCCCUGCG AUUUCCCGUC GA-3´), U1-SL34 (5´-GGGAUCGCU GACCCCUGC GAUUUCCCC AAAUGUGGG AAACUCGAC UGCAUAAUUU GUGGUAGUGG GGGACUGCGU UCGCGCUUUC CCCU-3´) and HH-U1 (5´-gggaucaggu aaguauccug aaguauccug augaguccgu gaggacgaaa cgguacccgg uaccgucGAU ACUUACCUGG CAGGGGAGAU ACCAUGAUCA CGAAGGUGGU UUUCCCAGGG CGAGGCUUAU CCAUUGCACU CCGGAUGUGC UGACCCCUGC GAUUUCCCCA AAUGUGGGAA ACUCGACUGC AUAAUUUGUG GUAGUGGGGG ACUGCGUUCG CGCUUUCCCC UGUCGA -3´) were transcribed from linearized plasmids, all other RNA-sequences from short DNA-templates (Microsynth), namely EMCV^{EmutF}(5´-GAGCG UUUGUCUAUA UGUgaaaaaggagCAUAUUG CCGUCUUUUG GCAAUGUG-3´), EMCV^{E} (5´- GGAGCG UUUGUCUAUA UGUUCC-3´) and EMCV^{F} (5´-GGAUAUUG CCGUCUUUUG GCAAUGUCC-3´). We used T7 RNA polymerase and unlabeled (Applichem) or $^{13}C^{15}N$-labeled nucleotide triphosphates (NTPs, produced in-house)30 for transcription. $MgCl_2$ concentrations were optimized in 50 µL test reactions for each construct. All transcribed EMCV derived RNA contained an artificial 5´-GGA or 5´-GAG sequence to enhance transcription initiation. HH-U1 and HH-U1-SL12 contained a hammerhead ribozyme (above shown in small letters) at the 5´-end that was cotranscriptionally cleaved. Names of secondary structure elements of EMCV-IRES constructs correspond to Kaminski et al.31, their nucleotide numbers to Duke et al.32. Transcripts were purified by denaturing anion exchange chromatography followed by butanol extraction as described33. RNA pellets were resuspended in boiling water, incubated at 98 °C for one minute and snap-cooled in liquid nitrogen for refolding. EMCV^{DElinkF} and EMCV^{DElink}, which exceed the size-range of optimal resolution of the denaturing anion exchange chromatography were further purified by SEC (HiLoad 16/60 Superdex 200 pg, GE) performed with NMR buffer or 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$ (1X RNase H buffer) as eluent. Purity of all transcripts was tested by urea-PAGE34.

Segmental labeling of RNA

We performed RNase H cleavage and DNA-splinted RNA ligation as described by Duss and Diarra dit Konte et al.20 using three 2´-O-methyl-RNA-DNA-chimeras to direct RNase H to perform cleavage of EMCV^{DElinkF} after nucleotide 319 (chimSLD: 5´- A^m A^m C^m G^m C^m

A^m dC dA dC dC G^m G^m C^m C^m U^m U^m -3′), 337 (chimSLE: 5′- A^m A^m A^m A^m U^m A^m dA dC dA dT A^m U^m A^m G^m A^m C^m -3′) and 347 (chimLinkF: 5′- A^m A^m U^m A^m U^m G^m dG dT dG dG A^m A^m A^m A^m U^m -3′; see also Fig. 1) and to cleave U1-SL12 and U1-SL34 after nucleotide 92 and 16 (both chimSL23: 5′-G^m A^m A^m A^m U^m C^m G^m dC dA dG dG G^m G^m U^m C^m A^m G^m C^m-3′), respectively, to generate fragments for RNA ligation. Optimal RNA:chimera-ratios for cleavage were 50:1 (chimSLE, chimLinkF), 5:1 (chimSLD) and 2:1 (chimSL23) as tested in 15 μL small-scale reactions. Optimal RNase H concentration did not scale up linearly and large scale cleavage was performed in aliquots of 33 μM RNA and 100 nM RNase H in 750 μL 1X RNaseH buffer and above mentioned RNA:chimera-ratios. $^{13}C^{15}N$-labeled $EMCV^{DElinkF}$-RNA was triple-digested with all three chimeras at the same time to produce the isotope labeled fragments embedding nucleotides 284-319, 320-336, 337-347 and 348-372 corresponding to stem-loop (SL) D, SLE, the linker between SLE and SLF (Link) and SLF, respectively. Cleavage efficiency reached almost 100% for all digests after 2 h at 37 °C and cleaved products were purified by denaturing anion exchange chromatography followed by butanol extraction33.

Fragments for segmental isotope labeling (Fig. 1, Fig. 3 and Supplementary Fig. 3) were annealed to a DNA-splint which is reverse complementary to nucleotides 305 to 361 of $EMCV^{DElinkF}$(RNA:splint = 1:1.2) or nucleotides 69 to 119 of U1-snRNA. We ligated 10 μM RNA in 1X T4-DNA-Ligase buffer, 10 % PEG 4000 at 37 °C for either 6 h using 500 U/mL of T4-DNA-Ligase (Fermentas, Weiss-units) or 3 h using 0.24 mg/mL in-house produced T4-DNA-Ligase. After ligation, we digested the EMCV-DNA-splint for 15 min at 37 °C by adding RNase-free DNase I and RDD-buffer (RNase-free DNase Set, Qiagen) to a final concentration of 10 U/mL (Kunitz-units). Ligation products were purified using denaturing anion-exchange-chromatography followed by butanol-extraction and refolding. All steps were monitored using urea-PAGE34.

RNA-protein complex formation

Complexes of single PTBP1-RRMs and of PTBP1-RRM12 with their cognate RNA were prepared by mixing both components in equimolar ratios at desired concentrations. To reduce aggregation upon complex formation of PTBP1-RRM34 and PTBP1 with multivalent RNA-targets, we mixed appropriate volumes of concentrated protein (0.5-1 mM) rapidly with dilute (5-10 μM), ice-cold RNA. Samples for NMR were further concentrated and purified by SEC using NMR-buffer as running buffer. Fractions were tested by native-gel electrophoresis, pooled and concentrated. PTBP1-$EMCV^{DElinkF}$ complexes for UV-crosslinking are made of equimolar mixtures of unlabeled and segmentally or uniformly isotope labeled RNA. Samples were named “D”, “E”, “Link” and “F” according to the isotope labeled RNA-segment or “U” for uniformly labeled RNA. For crosslinking of RRM1, we mixed unlabeled and uniformly isotope labeled $EMCV^{E}$ at equimolar ratios and added purified RRM1. U1snRNP was prepared as described previously25 after annealing of the 5′ splice site (5′-GGAGUAAGUCU-3′) of the SMN1 exon 7.

UV-induced RNA-protein crosslinking

We irradiated one half of each PTBP1-$EMCV^{DElinkF}$ sample corresponding to 500 μg for sample “U” or 250 μg for “D”, “E”; “Link” and “F” samples at a concentration of 0.8-1.0

mg/mL using a UV stratalinker 1800 (Stratagene), the other half of each sample was kept as control. For U1snRNP, we irradiated 180 µg of U1snRNP "12" and U1snRNP "34" with a concentration of 0.6 mg/mL in 10 mM sodium phosphate, pH 6.8, 50 mM NaCl, 5 mM DTT. For UV treatment, we loaded 50 µL sample/well on a 96-well-plate (PS, U-bottom, non-binding, clear; Greiner bio one), placed it on ice into the UV-device with a distance of the sample from the bottom of the device of 12 cm and irradiated 5 times with 800 mJ/cm^2 as monitored by the build-in detector. Each irradiation step was separated by 1 min for sample cooling. Irradiated and control samples were precipitated with ethanol as described previously35. We optimized the irradiation energy in steps of 800 mJ/cm^2 in the range of 2400-7200 mJ/cm^2 total energy on PTBP1-EMCV^{DElinkF} complex using free EMCV^{DElinkF} and free PTBP1 and non-irradiated samples as control.

Digestion, clean up and enrichment of RNA-protein crosslinks

Ethanol precipitates were resuspended and hydrolyzed according to Sharma et al.35. In brief, pellets were resuspended in 50 µL 50 mM Tris-HCl, pH 7.9, 4 M urea, diluted with 150 µL 50 mM Tris-HCl, pH 7.9, to a urea concentration of 1 M, and incubated at 52 °C after addition of 1.25 U RNase T1 (ThermoFisher) and 1.25 µg RNase A (Ambion), which corresponds to 5 U and 5 µg enzyme per mg of RNA-protein complex, respectively. After 2 h, samples were cooled on ice, supplemented with $MgCl_2$ to a concentration of 1 mM and digested with 31.25 U benzonase (Sigma-Aldrich; 125 U per mg complex) at 37 °C for 1.5 h. After RNA digestion, we added 7 µg trypsin (Promega) yielding a 24:1 protein:enzyme-ratio (w/w) and incubated the samples overnight on a thermomixer (Eppendorf) at 650 rpm and 37 °C, inactivated trypsin for 10 min at 70 °C, replenished 25 U benzonase, 1 U RNase T1 and 1 µg RNase A and completed the RNA digestion for 1 h at 37 °C.

Digestions were purified by solid-phase extraction (SPE, Waters SepPak tC18 cartridges) and RNA-protein crosslinks were enriched by titanium dioxide affinity chromatography according to Leitner et al.36. SPE eluates were dried and resuspended in 100 µL of 50 % acetonitrile (ACN), 0.1 % trifluoroacetic acid (TFA), 300 mM lactic acid. The samples were then incubated with 5 mg of pre-equilibrated TiO_2 beads (5 µm Titansphere, GL Sciences). We used the same buffer for equilibration, incubation and the first washing step. A second washing step was performed with 50 % ACN, 0.1 % TFA, followed by elution with 50 mM ammonium phosphate, pH 10.5. For each step, we incubated the beads for 10 min at 1400 rpm on a mixer and pelleted them by centrifugation at 16 100 g for 2 min. All eluates were immediately acidified to pH 2 with concentrated TFA and purified by SPE as above.

LC-MS/MS and MS-data-analysis

For mass spectrometry analysis, samples were resuspended in 16 µL of water/acetonitrile/formic acid (95:5:0.1, v/v/v), and 4 µL of each sample were used for duplicate injections. LC-MS/MS analysis was performed with an Easy nLC 1000 HPLC system (ThermoFisher Scientific) connected to an Orbitrap Elite mass spectrometer (ThermoFisher Scientific) equipped with a Nanoflex electrospray source. For the PTBP1 samples, peptides were separated on a PepMap RSLC column (150 mm × 75 µm, 2 µm particle size, ThermoFisher Scientific) using a gradient of 5-30% 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 (98:2:0.15,

v/v/v); the flow rate was set to 300 nL/min. For the U1 snRNP sample, an extended gradient from 5-25% mobile phase B within 90 min was used.

The Orbitrap Elite was operated in the data dependent acquisition mode. Precursor ion spectra were acquired in the Orbitrap analyzer at a resolution of 120 000. For each cycle, the top 15 precursor ions were selected for fragmentation using collision-induced dissociation and detection of the fragment ions in the linear ion trap at normal scan rate. Additional fragmentation settings were: Isolation width, 2 m/z; normalized collision energy, 35; activation time = 10 ms. Dynamic exclusion (30 s after one sequencing event) was activated.

For data analysis, files were converted from the native Thermo raw format into mzXML using msConvert37 and searched against the target protein sequences using xQuest (version 2.1.3)38. 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 preliminary data analysis, 15 different nucleotides were considered as potential modifications and specified as monolink adducts (parameter “monolinkmw” in xquest.def), along with their water loss products: C, U, AU, CU, GU, UU, GC, AC, AG, AA, AUU, CUU, GUU, CCU, and UUU (listed in Supplementary Table 1). The respective mass shifts between all-$^{12}C/^{14}N$ (“light”) and all-$^{13}C/^{15}N$ (“heavy”) nucleotides were specified to find MS/MS spectral pairs with a mass tolerance of 10 ppm and a retention time tolerance of 30 s. Because xQuest does not allow the simultaneous search against all possible adducts with different light/heavy mass shifts, independent searches were carried out for all adduct types (an intact modification and its water loss product were searched together).

Additional search settings were as follows: Enzyme = trypsin, maximum number of missed cleavages = 2, MS mass tolerance = 5 ppm, MS/MS mass tolerance = 0.2 Da for “common”-type fragment ions and 0.3 Da for “xlink”-type fragment ions. The original scoring scheme of xQuest14 was used and only identifications with a score ≥ 20 (for PTBP1) and ≥ 16 (for U1snRNP) were considered.

NMR spectroscopy and NMR data analysis

NMR spectra were acquired at 303 K and 313 K for PTBP1-RRM1-EMCV^{E} complexes or 313 K for all other PTBP1-EMCV complexes on Bruker Avance III 500, 600, 700 or 900 MHz spectrometers equipped with cryoprobes and on a Bruker Avance III 750 MHz spectrometer with a room temperature probe. We processed spectra with Topspin 2.1 or Topspin 3.0 and analyzed in Sparky 3.0. ^{1}H, ^{13}C and ^{15}N assignments of RNA and protein were achieved by standard methods39. For modelling of the PTBP1-RRM2-EMCV^{F} complex, we used intramolecular distance restraints derived from HHC- and HHN- 3D-NOESY experiments as well as residual dipolar couplings measured for backbone amides and RNA-C1´-H1´, C5-H5, C6-H6, C8-H8 and C2-H2 bonds. Intermolecular distance restraints were extracted from 3D ^{13}C-F_1-edited, F_3-filtered-NOESY-HSQCs40 and a 2D ^{1}H-^{1}H F_1-^{13}C-filtered, F_2-^{13}C-edited NOESY-spectra41 recorded on complexes reconstituted either from $^{13}C^{15}N$-labeled protein and unlabeled RNA or from ^{15}N-labeled protein and $^{13}C^{15}N$-labeled RNA.

Europe PMC Funders Author Manuscripts

Modelling

Modelling of the PTBP1-RRM2-EMCV^{F} complex was established with a combination of different software classically required for structure prediction and determination of protein-RNA complexes. We used the Atnos/Candid-program suite42, 43 and artificial RRM NOESY matrices to generate peak lists corresponding to intramolecular NOESY patterns typical for the RRM fold. CYANA 3.044 and more particularly the CYANA noeassign command were used to integrate distance and angle restraints and to calculate models. For modelling, CLIR-MS/MS-data were inserted as ambiguous distance restraints because crosslinking sites define various distances between base rings of nucleic acids and side chains of amino acids, respectively. Intramolecular restraints were derived from published protein structures10 (PTBP1-RRM1-pdb: 2N3O; PTBP1-RRM2-pdb: 2ADB; PTBP1-RRM34-pdb: 2ADC) and RNA-structures predicted by MC-FOLD and MC-SYM45. Additional specific protein-RNA contacts extracted from available complex structures were integrated as unambiguous distance restraints. For all models, we calculated 200 structures per cycle and selected the 20 of lowest energy as starting ensemble for the next cycle. For modelling PTBP1-RRM1-EMCV^{E} we initiated the CYANA noeassign calculation with the average RRM1 structure (pdb: 2N3O) in cycle 1 excluding the RNA-moiety. The final 20 lowest energy models obtained with CYANA noeassign were refined with amber 1246 force-field to avoid steric clashes and to improve electrostatic and hydrophobic protein/RNA contacts. CLIR-MS/MS derived intermolecular distance restraints are listed in Supplementary Table 2, other restraints in Supplementary Table 4. Atomic coordinates of the CLIR-MS/MS-derived models of PTBP1-RRM1, RRM2, RRM3 and RRM4 as well as the NMR-derived model of PTBP1-RRM2 are available as Supplementary Data 1, 2, 3, 4 and 5, respectively.

Statistics

All MS/MS measurements were performed as technical replicates and all numbers displayed in Supplementary Table 2 are the sum of the spectral counts of the two technical replicates. Only identifications with a score ≥ 20 (for PTBP1) and ≥ 16 (for U1snRNP) were considered for quantification. For validation, we repeated sample “D” in an independent experiment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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.

References

1. Cooper TA, Wan L, Dreyfuss G. RNA and disease. Cell. 2009; 136:777–793. [PubMed: 19239895]

Europe PMC Funders Author Manuscripts

2. Kim E, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell. 2015; 27:617–630. [PubMed: 25965569]
3. Leitner A, Faini M, Stengel F, Aebersold R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. Trends Biochem Sci. 2016; 41:20–32. [PubMed: 26654279]
4. Kramer K, et al. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat Methods. 2014; 11:1064–1070. [PubMed: 25173706]
5. Kwon SC, et al. The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol. 2013; 20:1122–1130. [PubMed: 23912277]
6. Beckmann BM, et al. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat Commun. 2015; 6:10127. [PubMed: 26632259]
7. Lelyveld VS, Bjorkbom A, Ransey EM, Sliz P, Szostak JW. Pinpointing RNA-Protein Cross-Links with Site-Specific Stable Isotope-Labeled Oligonucleotides. J Am Chem Soc. 2015; 137:15378–15381. [PubMed: 26583201]
8. Spriggs KA, Bushell M, Mitchell SA, Willis AE. Internal ribosome entry segment-mediated translation during apoptosis: the role of IRES-trans-acting factors. Cell Death Differ. 2005; 12:585–591. [PubMed: 15900315]
9. Romanelli MG, Diani E, Lievens PM. New insights into functional roles of the polypyrimidine tract-binding protein. Int J Mol Sci. 2013; 14:22906–22932. [PubMed: 24264039]
10. Oberstrass FC, et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science. 2005; 309:2054–2057. [PubMed: 16179478]
11. Han A, et al. De novo prediction of PTBP1 binding and splicing targets reveals unexpected features of its RNA recognition and function. PLoS Comput Biol. 2014; 10:e1003442. [PubMed: 24499931]
12. Kafasla P, et al. Polypyrimidine tract binding protein stabilizes the encephalomyocarditis virus IRES structure via binding multiple sites in a unique orientation. Mol Cell. 2009; 34:556–568. [PubMed: 19524536]
13. Jang SK, Wimmer E. Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. Genes Dev. 1990; 4:1560–1572. [PubMed: 2174810]
14. Rinner O, et al. Identification of cross-linked peptides from large sequence databases. Nat Methods. 2008; 5:315–318. [PubMed: 18327264]
15. Lamichhane R, et al. RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. Proc Natl Acad Sci U S A. 2010; 107:4105–4110. [PubMed: 20160105]
16. Gygi SP, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999; 17:994–999. [PubMed: 10504701]
17. Ong SE, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002; 1:376–386. [PubMed: 12118079]
18. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci U S A. 1999; 96:6591–6596. [PubMed: 10359756]
19. Muller DR, et al. Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. Anal Chem. 2001; 73:1927–1934. [PubMed: 11354472]
20. Duss O, Diarra Dit Konte N, Allain FH. Cut and paste RNA for nuclear magnetic resonance, paramagnetic resonance enhancement, and electron paramagnetic resonance structural studies. Methods Enzymol. 2015; 565:537–562. [PubMed: 26577744]
21. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004; 76:4193–4201. [PubMed: 15253663]
22. Daubner GM, Clery A, Jayne S, Stevenin J, Allain FH. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. EMBO J. 2012; 31:162–174. [PubMed: 22002536]

23. Will CL, Luhrmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011; 3
24. Kondo Y, Oubridge C, van Roon AM, Nagai K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. Elife. 2015; 4
25. Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature. 2009; 458:475–480. [PubMed: 19325628]
26. Vizcaino JA, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2013; 41:D1063–1069. [PubMed: 23203882]
27. Tugarinov V, Kanelis V, Kay LE. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. Nat Protoc. 2006; 1:749–754. [PubMed: 17406304]
28. Sambrook, J., Russell, DW. Molecular cloning a laboratory manual. Edn. 3rd. Cold Spring Harbor Laboratory Press, Cold Spring Harbor; New York: 2001.
29. Gasteiger, E., et al. The Proteomics Protocols Handbook. Walker, JM., editor. Humana Press; Totowa, NJ: 2005. p. 571-607.
30. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res. 1987; 15:8783–8798. [PubMed: 3684574]
31. Kaminski A, Jackson RJ. The polypyrimidine tract binding protein (PTB) requirement for internal initiation of translation of cardiovirus RNAs is conditional rather than absolute. RNA. 1998; 4:626–638. [PubMed: 9622122]
32. Duke GM, Hoffman MA, Palmenberg AC. Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. J Virol. 1992; 66:1602–1609. [PubMed: 1310768]
33. Duss O, Maris C, von Schroetter C, Allain FH. A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. Nucleic Acids Res. 2010; 38:e188. [PubMed: 20798173]
34. Maniatis T, Jeffrey A, van deSande H. Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. Biochemistry. 1975; 14:3787–3794. [PubMed: 1174504]
35. Sharma K, et al. Analysis of protein-RNA interactions in CRISPR proteins and effector complexes by UV-induced cross-linking and mass spectrometry. Methods. 2015; 89:138–148. [PubMed: 26071038]
36. Leitner A, et al. Probing the phosphoproteome of HeLa cells using nanocast metal oxide microspheres for phosphopeptide enrichment. Anal Chem. 2010; 82:2726–2733. [PubMed: 20201521]
37. Chambers MC, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012; 30:918–920. [PubMed: 23051804]
38. Walzthoeni T, et al. xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. Nat Methods. 2015; 12:1185–1190. [PubMed: 26501516]
39. Dominguez C, Schubert M, Duss O, Ravindranathan S, Allain FH. Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. Progress in nuclear magnetic resonance spectroscopy. 2011; 58:1–61. [PubMed: 21241883]
40. Lee W, Revington MJ, Arrowsmith C, Kay LE. A pulsed field gradient isotope-filtered 3D 13C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. FEBS Lett. 1994; 350:87–90. [PubMed: 8062930]
41. Peterson RD, Theimer CA, Wu H, Feigon J. New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. J Biomol NMR. 2004; 28:59–67. [PubMed: 14739639]
42. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR. 2002; 24:171–189. [PubMed: 12522306]

43. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol. 2002; 319:209–227. [PubMed: 12051947]

44. Guntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol. 2004; 278:353–378. [PubMed: 15318003]

45. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature. 2008; 452:51–55. [PubMed: 18322526]

46. Case, DA., D, TA., Cheatham, TE., III, Simmerling, CL., Wang, J., Duke, RE., R, et al. AMBER 12. University of California; San Francisco: 2012.

Europe PMC Funders Author Manuscripts

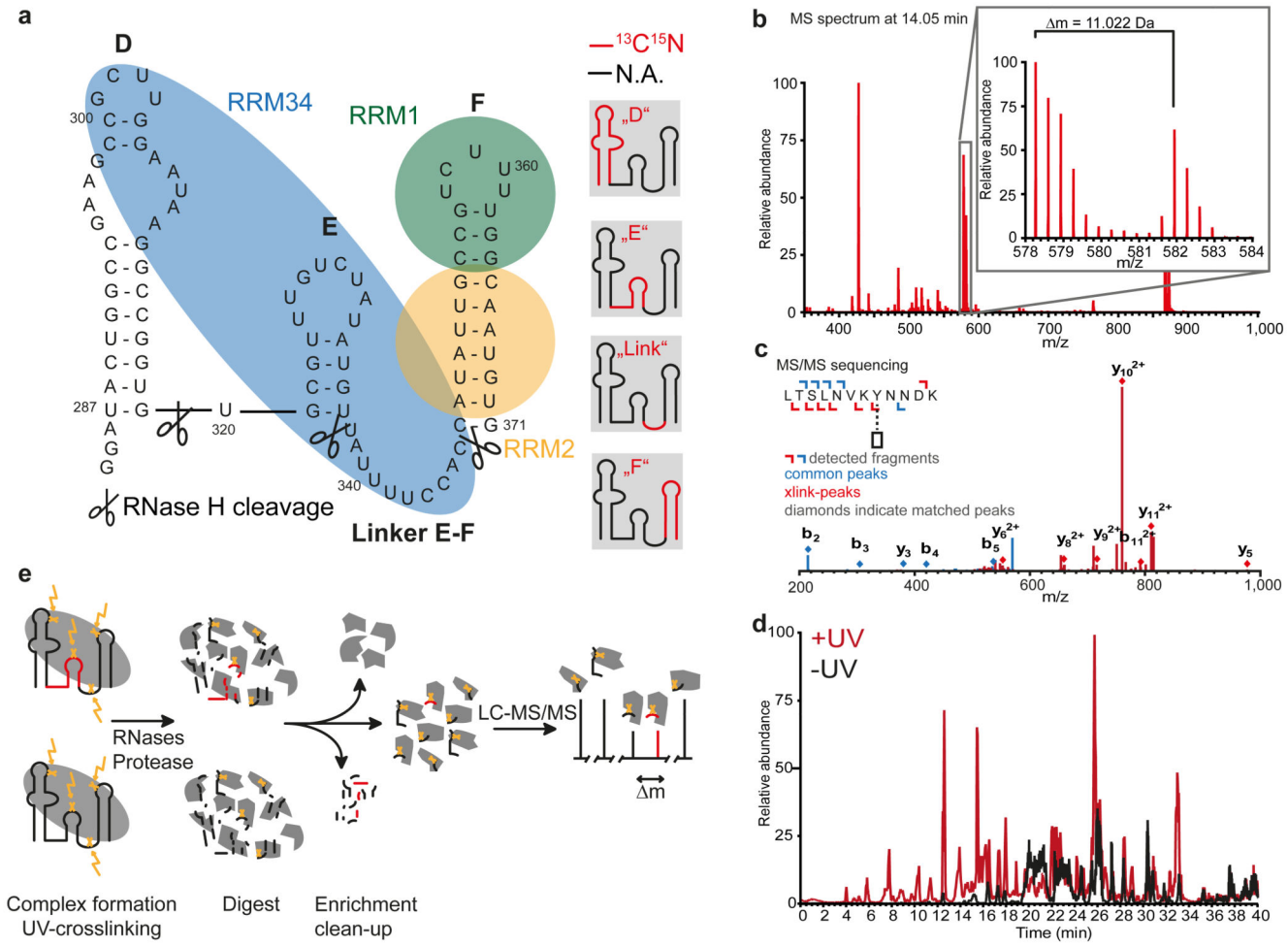

Fig. 1. CLIR-MS/MS analysis of crosslinked PTBP1-peptides.

a, Sequence and predicted secondary structure of the EMCV^{DElinkF} RNA, including the localization of PTBP1-RRMs as suggested previously12, N.A. = natural abundance. RNase H cleavage sites for segmental isotope labeling and the four segmentally isotope labeled constructs used for UV-crosslinking are represented schematically. **b**, Mass spectrum at 14.05 min and the typical isotope pattern due to the attachment of 50% isotope labeled uracil (actual labeling rate is slightly less because of incomplete isotopic enrichment in the “heavy” RNA.). **c**, MS/MS spectrum of the peptide LTSLNVKYNNDK from RRM2 (residues 260-271) with a uracil modification at Tyr267 or Asn268. **d**, Total ion chromatograms of PTBP1 in complex with 50% uniformly labeled EMCV^{DElinkF} with (red) and without (black) prior UV-treatment. **e**, schematic representation of crosslinking, enrichment and LC-MS/MS analysis. Specific isotope splitting (Δm) in the precursor ion spectrum facilitates the crude localization of peptides to the specific differentially labeled RNA-segment, reads of crosslinked di- and trinucleotides allows single nucleotide resolution.

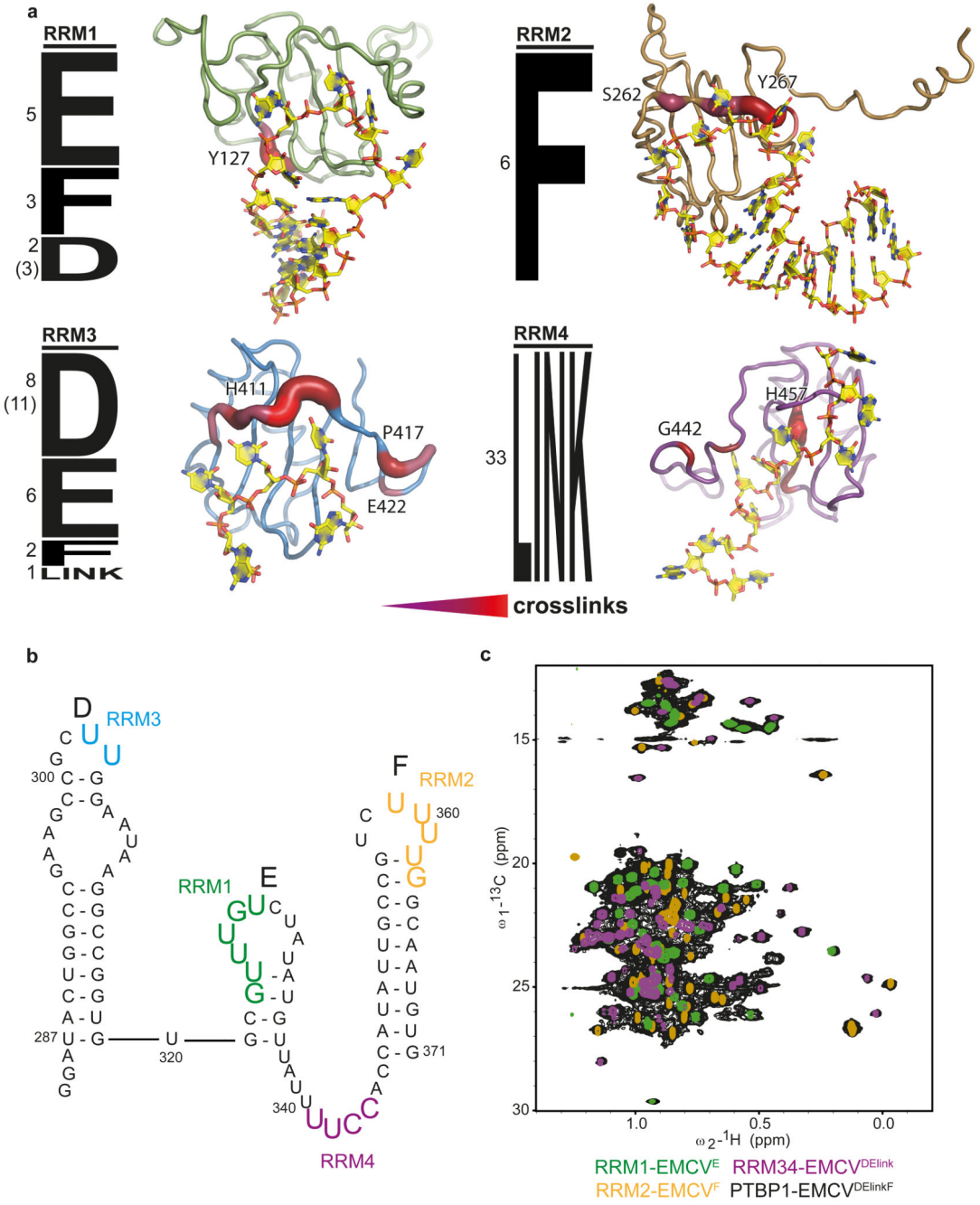

Fig. 2. CLIR-MS/MS mapping of PTBP1 on EMCV-IRES domains D-F.

a, Mapping of PTBP1 binding on the EMCV-IRES RNA. RRM2 and RRM4 exclusively crosslinked to “F” and “Link”, respectively, RRM1 and RRM3 appear to crosslink predominantly to “E” and “D”, respectively (spectral counts as indicated, numbers in brackets correspond to the spectral counts detected in an independent replicate of sample “D”). Crosslinking sites are highlighted on the CLIR-MS/MS derived models according to their relative crosslinking reactivity **b,** RRM-binding sites on the RNA as derived from the analysis of crosslinked di- and trinucleotides and NMR titration experiments (for details see

text). **c,** Overlay of $^{1}H^{13}C$-methyl-TROSY spectra of ILV-methyl groups of PTBP1 in complex with $EMCV^{DElinkF}$ (black), RRM1, RRM2 and RRM34 in complex with $EMCV^{E}$ (green), $EMCV^{F}$ (yellow) and $EMCV^{DElink}$ (magenta). The identical peak position in the full-length complex and in the subcomplexes confirms binding of RRM1, RRM2, RRM3 and RRM4 to SLE, SLF, SLD and the linker E-F, respectively.

Europe PMC Funders Author Manuscripts

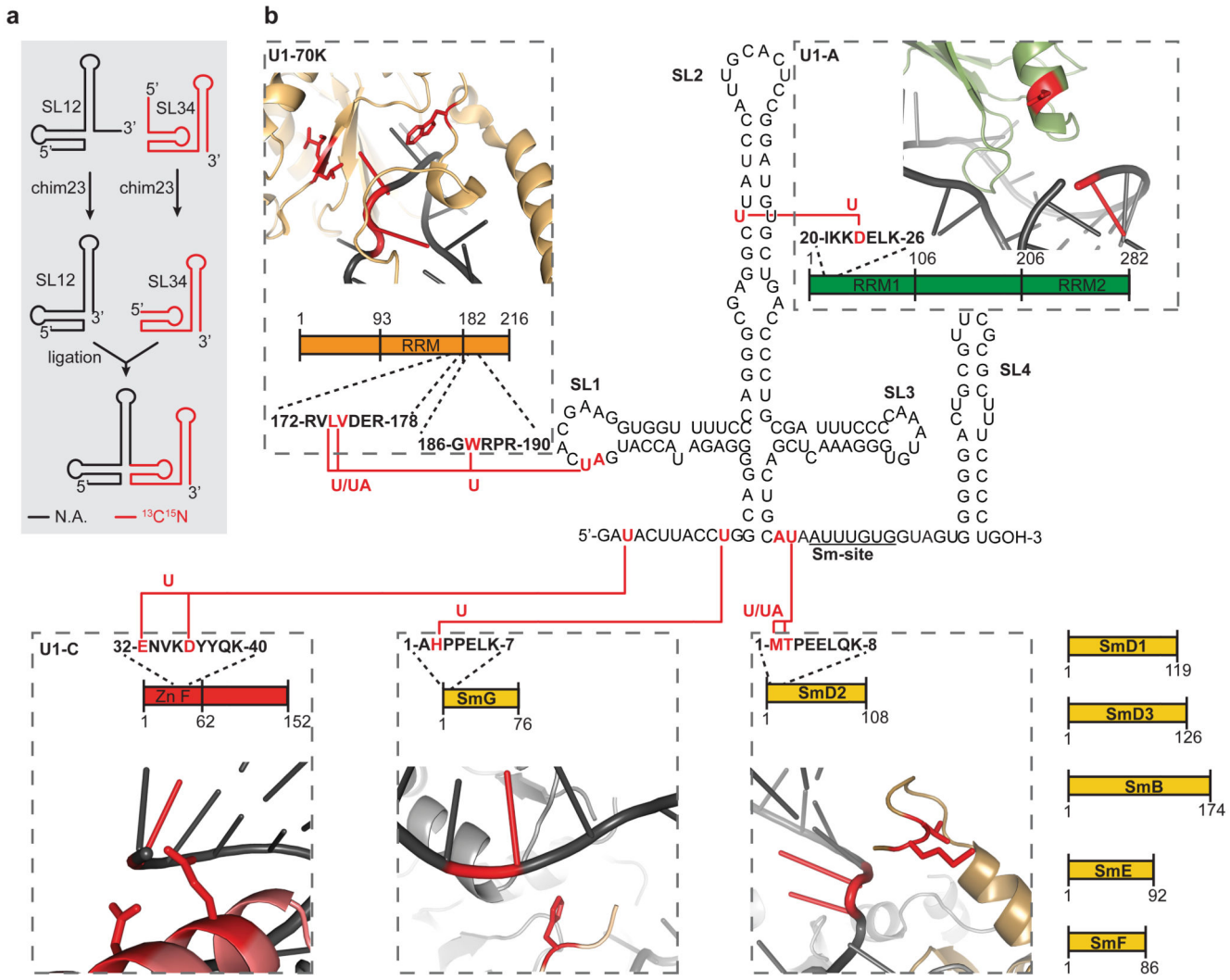

Fig. 3. CLIR-MS/MS applied to U1snRNP.

a, U1-SL12 and U1-SL34 precursors were cleaved by RNase H directed by chimera 23 (chim23) and purified. Isotope labeled SL34 and unlabeled SL12 (as indicated here) or isotope labeled SL12 and unlabeled SL34 were ligated, purified and used for U1snRNP reconstitution, N.A. = natural abundance. **b**, 5 of 10 U1snRNP proteins were identified and mapped onto the U1snRNA sequence (red letters). Crosslinks are illustrated using previously published structures24.

Europe PMC Funders Author Manuscripts