Mass spectrometric evidence for the underrepresented RNA-protein crosslinks

Jong Woo Bae[1]

Correspondence: jakenizer@gmail.com

1 No affiliation

Abstract

In protein-RNA complexes, UVC crosslinks proteins predominantly to uracil. Concrete evidence for other UVC-crosslinked RNA bases is lacking. Here, I elucidate mass spectrometric signatures of UVC-crosslinked guanine, cytidine, and methyladenine. This can help us capture the native protein-RNA interactions in more detail.

Main Text

RNA-binding sites (RBSs) are amino acids in RNA-binding proteins (RBPs) that make direct contacts with RNA. Through bottom-up mass spectrometric analysis, UV-crosslinked and hydrofluoride-treated RNA moieties can pinpoint the exact RBSs on peptides[1,2]. Previous studies showed that UVC irradiation dominantly crosslinks uracil to proteins[1,3-9], but finding solid evidence for other low-abundance RNA adducts has

remained challenging. Here I sought to identify the underrepresented UVC-crosslinked RNA moieties using a previous RBS-mapping dataset[1,10] via two different approaches.

First, I exploited previous observations that neutral loss of ribose occurs from UV-crosslinked RNA nucleosides[1,2]. Fragpipe open search with diagnostic ion mining[11-16] followed by inspection of fragment ions (b, y) for ribose loss unraveled two additional UVC-crosslinked RNA moieties (**Table 1**). One RNA adduct mass corresponded to guanosine with ammonia loss, while the other could be either cytosine with ammonia loss or uridine with water loss.

In addition, I looked into RNA-binding domains (RBDs) with well-defined RNA sequence specificity. I examined the precursor mass shifts from peptides located within PABPC1 RRM1-2 (poly(A)-binding)[17-20], RBFOX2 RRM (G-crosslinking)[21,22], and PTBP1 RRM1 (polypyrimidine-binding)[23-25]. Most unexpectedly, methyladenine/adenosine crosslinks[26], but not adenine/adenosine crosslinks, were detected on H144 of PAPBC1 RRM2 domain (**Figure 1a**). Meanwhile, guanosine crosslinks were identified on F154/V155 of RBFOX2 RRM (**Figure 1b**), and cytosine crosslinks on H133 of PTBP1 RRM1 (**Figure 1c**). All identified RBSs (except V155, which may have been catalogued owing to in-solution dynamic fluctuation, mislocalization during open search, or electron hopping during UVC-crosslinking[8]) were proximal to adenines, guanine, or cytosine in previously reported ribonucleoprotein structures[27-29], respectively.

RBSs harboring the five newly identified RNA adducts were predominantly identified in annotated RBPs (**Table 2**). These adduct types may be incorporated in closed search parameters on UVC-crosslinked RBS datasets, but at the cost of search space inflation[1].

Figures and Tables

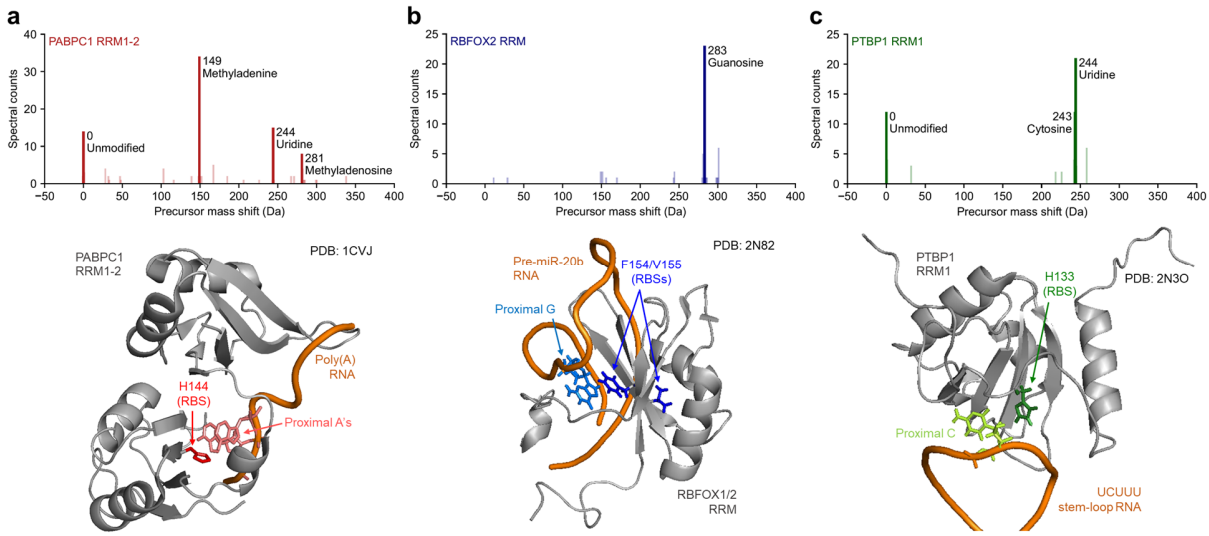

Figure 1. UVC-crosslinks on RBDs with well-defined sequence specificity.

(a) UVC-crosslinked methyladenine/adenosine on PABPC1 RRM1-2. Top: Nominal precursor mass shifts of peptide-spectrum matches (PSMs) within PABPC1 RRM1-2 (amino acid positions: 11-175)[16]. Bottom: Structure of PABPC1 RRM1-2 in complex with poly(A) RNA (PDB: 1CVJ)[27]. (b) UVC-crosslinked guanosine on RBFOX2 RRM. Top: Nominal precursor mass shifts of PSMs within RBFOX2 RRM (amino acid positions: 111-187)[16]. Bottom: Structure of RBFOX2 RRM in complex with pre-miR-20b RNA (PDB: 2N82)[28]. (c) UVC-crosslinked cytosine on PTBP1 RRM1. Top: Nominal precursor mass shifts of PSMs within PTBP1 RRM1 (amino acid positions: 59-143)[16]. Bottom: Structure of PTBP1 RRM1 in complex with UCUUU stem-loop RNA (PDB: 2N3O)[29].

Table 1. UVC-crosslinked RNA moieties with ribose fragmentation.

MS1 shift (Da)	Chemical formula	Nucleoside-(neutral loss)
226.0594	$C_9H_{10}N_2O_5$	Cytosine-NH_3
		Uridine-H_2O
266.065	$C_{10}H_{10}N_4O_5$	Guanosine-NH_3

Table 2. RBSs/proteins identified with the underrepresented RNA adducts.

Evidence type	Ribose fragmentation		RBDs with sequence specificity		
RNA adduct	Cytosine-H_2O/ Uridine-NH_3	Guanosine-NH_3	Methyladenine/ adenosine	Guanosine	Cytosine
Nominal mass	226	266	149/281	283	243
#Protein	141	40	62	55	210
#RBS	411	89	120	112	614
%RBS in RBP	97.8	100.0	98.3	99.1	99.0

Data & code availability

The mass spectrometry data and the structural data have been retrieved from the PRIDE repository (PXD016254) or the RCSB PDB database (1CVJ, 2N82, 2N3O), respectively. Fragpipe workflows, analysis results, protein sequence/annotation resources, and the custom Python codes used in this study have been deposited to the GitHub repository (https://github.com/jwbaebio/RBS-ID-mining).

Acknowledgments

I sincerely thank V. Narry Kim, Jong-Seo Kim, and Alexey I. Nesvizhskii's labs for the valuable MS datasets and software.

Ethics declarations

The author declares no competing financial interests.

Methods

Four .raw files (UVC, RBS-ID) in PXD016254 dataset were searched against the Swiss-prot human proteome database (2022.10.11.) using Fragpipe (v18.0) with the parameters specified in the Open_RBSID.workflow file. The global.diagmine.tsv file was looked up for signatures of ribose loss from b/y-ions. The four psm.tsv files were merged and looked up for precursor modifications on peptides from PABPC1 RRM1-2, RBFOX2 RRM, or PTBP1 RRM1 domains. For nominal modification masses 149/281, 283, or 243, PSMs with unique localizations were taken to identify RBSs in PABPC1 RRM1-2, RBFOX2 RRM, or PTBP1 RRM1, respectively. The RNA-protein complex structures were visualized from .pdb files (1CVJ, 2N82, 2N3O) using PyMOL (v.1.7.2.1.). Also, the number of identified RBSs and proteins with respective modifications proteome-wide were counted, with RBP annotations integrated (2022.10.28.). All custom Python codes are available in the RBS-ID-mining.ipynb file.

References

1. Bae, J. W., Kwon, S. C., Na, Y., Kim, V. N. & Kim, J. S. Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution. *Nat Struct Mol Biol* **27**, 678-682, doi:10.1038/s41594-020-0436-2 (2020).

2. Bae, J. W., Kim, S., Kim, V. N. & Kim, J. S. Photoactivatable ribonucleosides mark base-specific RNA-binding sites. *Nat Commun* **12**, 6026, doi:10.1038/s41467-021-26317-5 (2021).

3. 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, doi:10.1038/nmeth.3092 (2014).

4. Dorn, G. *et al.* Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat Methods* **14**, 487-490, doi:10.1038/nmeth.4235 (2017).

5. Trendel, J. *et al.* The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391-403 e319, doi:10.1016/j.cell.2018.11.004 (2019).

6. Panhale, A. *et al.* CAPRI enables comparison of evolutionarily conserved RNA interacting regions. *Nat Commun* **10**, 2682, doi:10.1038/s41467-019-10585-3 (2019).

7. Gotze, M. *et al.* Single Nucleotide Resolution RNA-Protein Cross-Linking Mass Spectrometry: A Simple Extension of the CLIR-MS Workflow. *Anal Chem* **93**, 14626-14634, doi:10.1021/acs.analchem.1c02384 (2021).

8. Knorlein, A. *et al.* Nucleotide-amino acid pi-stacking interactions initiate photo cross-linking in RNA-protein complexes. *Nat Commun* **13**, 2719, doi:10.1038/s41467-022-30284-w (2022).

9. Sarnowski, C. P. *et al*. Sensitive detection and structural characterisation of UV-induced cross-links in protein-RNA complexes using CLIR-MS. *bioRxiv*, doi:10.1101/2022.03.31.486537 (2022).

10. Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543-D552, doi:10.1093/nar/gkab1038 (2022).

11. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* **14**, 513-520, doi:10.1038/nmeth.4256 (2017).

12. Geiszler, D. J., Polasky, D. A., Yu, F. & Nesvizhskii, A. I. Detecting diagnostic features in MS/MS spectra of post-translationally modified peptides. *Nat Commun* **14**, 4132, doi:10.1038/s41467-023-39828-0 (2023).

13. Chang, H. Y. *et al.* Crystal-C: A Computational Tool for Refinement of Open Search Results. *J Proteome Res* **19**, 2511-2515, doi:10.1021/acs.jproteome.0c00119 (2020).

14. da Veiga Leprevost, F. *et al.* Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat Methods* **17**, 869-870, doi:10.1038/s41592-020-0912-y (2020).

15. Geiszler, D. J. *et al.* PTM-Shepherd: Analysis and Summarization of Post-Translational and Chemical Modifications From Open Search Results. *Mol Cell Proteomics* **20**, 100018, doi:10.1074/mcp.TIR120.002216 (2021).

16. UniProt, C. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523-D531, doi:10.1093/nar/gkac1052 (2023).

17. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K. & Dreyfuss, G. mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol Cell Biol* **6**, 2932-2943, doi:10.1128/mcb.6.8.2932-2943.1986 (1986).

18. Blobel, G. A protein of molecular weight 78,000 bound to the polyadenylate region of eukaryotic messenger RNAs. *Proc Natl Acad Sci U S A* **70**, 924-928, doi:10.1073/pnas.70.3.924 (1973).

19. Nietfeld, W., Mentzel, H. & Pieler, T. The Xenopus laevis poly(A) binding protein is composed of multiple functionally independent RNA binding domains. *EMBO J* **9**, 3699-3705, doi:10.1002/j.1460-2075.1990.tb07582.x (1990).

20. Burd, C. G., Matunis, E. L. & Dreyfuss, G. The multiple RNA-binding domains of the mRNA poly(A)-binding protein have different RNA-binding activities. *Mol Cell Biol* **11**, 3419-3424, doi:10.1128/mcb.11.7.3419-3424.1991 (1991).

21. Weyn-Vanhentenryck, S. M. *et al.* HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep* **6**, 1139-1152, doi:10.1016/j.celrep.2014.02.005 (2014).

22. Feng, H. *et al.* Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. *Mol Cell* **74**, 1189-1204 e1186, doi:10.1016/j.molcel.2019.02.002 (2019).

23. Oberstrass, F. C. *et al.* Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054-2057, doi:10.1126/science.1114066 (2005).

24. Auweter, S. D., Oberstrass, F. C. & Allain, F. H. Solving the structure of PTB in complex with pyrimidine tracts: an NMR study of protein-RNA complexes of weak affinities. *J Mol Biol* **367**, 174-186, doi:10.1016/j.jmb.2006.12.053 (2007).

25. Clerte, C. & Hall, K. B. The domains of polypyrimidine tract binding protein have distinct RNA structural preferences. *Biochemistry* **48**, 2063-2074, doi:10.1021/bi8016872 (2009).

26. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767-772, doi:10.1038/nmeth.3453 (2015).

27. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835-845, doi:10.1016/s0092-8674(00)81517-2 (1999).

28. Chen, Y. *et al.* Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. *Nucleic Acids Res* **44**, 4381-4395, doi:10.1093/nar/gkw177 (2016).

29. Maris, C. *et al.* A transient alpha-helix in the N-terminal RNA recognition motif of polypyrimidine tract binding protein senses RNA secondary structure. *Nucleic Acids Res* **48**, 4521-4537, doi:10.1093/nar/gkaa155 (2020).