Trends in
Biochemical Sciences

CellPress

Scientific Life

Defining the commonalities between post-transcriptional and post-translational modification communities

Zachary T. Baumer,[1,12,13] Luke Erber,[2,12,13] Elizabeth Jolley,[3,12,13] Sheldon Lawrence,[4,12,13] Chuwei Lin,[5,12,13] Shino Murakami,[6,12,13] Veronica Perez,[7,12,13] Wil Prall,[8,12,13] Cassandra Schaening-Burgos,[9,12,13] Megan Sylvia,[10,12,13] Sixue Chen,[11,*] and Brian D. Gregory[8,*]

Post-transcriptional modifications of RNA (PRMs) and post-translational modifications of proteins (PTMs) are important regulatory mechanisms in biological processes and have many commonalities. However, the integration of these research areas is lacking. A recent discussion identified the priorities, areas of emphasis, and necessary technologies to advance and integrate these areas of study.

Ribonucleotides within the various RNA molecules in prokaryotes and eukaryotes can be marked by more than 160 distinct covalent chemical modifications [1,2]. Relatedly, a set of more than 300 covalent chemical amino acid modifications can occur in the cellular proteome [3,4]. Recent findings demonstrated that PRMs and PTMs impact the structure, functionality, stability, and degradation of the modified molecules [1–3,5,6]. These covalent additions are dynamically added and removed through writer and eraser complexes, respectively, providing a layer of post-transcriptional and post-translational regulation that affects the quality and quantity of RNA and protein [1–3,5,6]. Thus, it is not surprising that PRMs and PTMs can have widespread effects on the development, behavior, and disease susceptibility of organisms and their responses to abiotic and biotic stresses [1–3,7–9].

The importance of PRMs and PTMs to cellular regulatory mechanisms has become unquestioned [1–3,7–9]. However, significant gaps remain in the integrated understanding of PRMs and PTMs and their effects on overall cellular metabolism. Thus, an open discussion on the conceptual, theoretical, technical, and analytical resource needs in PRM and PTM studies was organized to bring together researchers from both areas. Overall, this discussion identified the priorities, methodologies, and technologies needed to advance and integrate future research in these areas.

Areas of future synergy

During the discussion, numerous areas for research synergy were noted. First, it was recognized that PRM and PTM research should be carried out across a broad and distantly related set of species, providing opportunities to bring together large communities of researchers to address the questions of covalent additions to RNAs and proteins across the tree of life. Such studies are likely to help define sets of common rules or characteristics driving the addition, removal, and functionality of PRMs and PTMs (Figure 1).

It was deemed crucial for proper validation of covalent modifications to characterize the true functionality. Currently, a comprehensive functional validation of RNA and protein modification sites is lacking. However, such studies are needed to truly define the biological processes and/or stresses affected by specific modifications.

Additionally, the reversibility of PRMs and PTMs was noted as an open area of research. For at least some PTMs, the effects of adding and removing them on protein structure and function have been studied. However, this is not the case for others, and it is still under debate for even the best studied PRM, m^6A. Therefore, we need to better characterize the potential dynamic reversibility of PRM and PTM systems to learn the rules and functions associated with these regulatory systems.

Another area of synergy involves the interplay and crosstalk between PRMs and PTMs. For instance, the pervasiveness and effects of PTMs on the machinery involved in writing, erasing, and interacting with PRMs have not been well studied. However, there have been suggestions that some PRM writers can be modified post-translationally, but the biological effects of these events are currently unknown. Furthermore, the pervasiveness and effects of PRMs on the transcripts encoding the proteins involved in adding, removing, and sensing PTMs is understudied.

The final area identified was the overall effects of PRMs and PTMs on the cellular metabolome. It was noted that these effects are likely to impact the overall cellular and organismal phenotype through the alteration of an organism's metabolome, a hypothesis that has been supported by previous findings [10]. Thus, metabolomic studies in systems with altered PRM and PTM systems should be carried out to better understand the connection of PRMs and PTMs with phenotypes.

All areas of future synergy will be aided by the development of new methods,

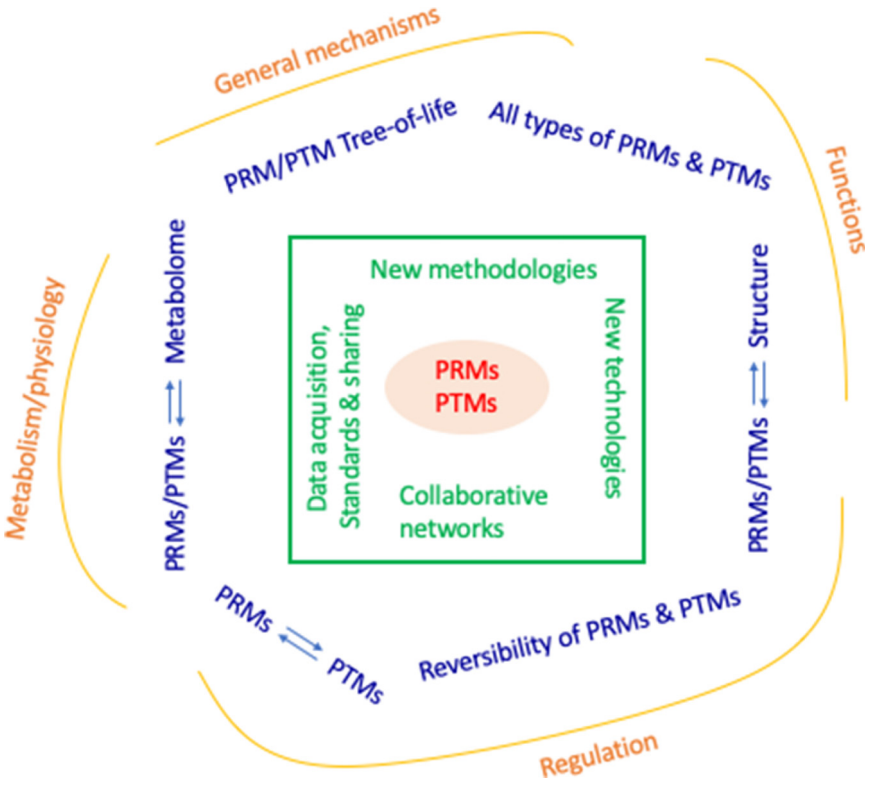

Figure 1. Diagram summarizing the future areas of synergistic research (in blue) and potential outcomes (in orange) for post-transcriptional modifications of RNA (PRMs) and post-translational modifications of proteins (PTMs) identified by the workshop. The central box depicts what it takes to achieve the outcomes (i.e., the workshop recommendations, in green). What is currently composed of various unique and conceptually distinct research endeavors can be united by the similarities of the underlying rules and features that govern these cellular pathways. These ideas can synergize PRM and PTM research, which has the potential to reveal the fundamental rules of life specific to these chemical moieties in biological systems.

technologies, and community resources, as well as data standards (Figure 1).

Data acquisition, standards, and storage needed

There was a consensus that data quality standards need to be developed. Specifically, very high-quality validated datasets defining the positions, metabolome effects, and functions of *bona fide* PRMs and PTMs are necessary to serve as training sets for developing the necessary machine learning (ML) and artificial intelligence (AI) analytical tools that will significantly aid in driving discoveries forward and will provide the resources for the development of large-scale PRM and PTM databases and browsers. Additionally, standardization of the methodologies and analytical approaches will ensure the ability to comprehensively compare all available datasets. The democratization of available datasets and resources can increase engagement in PRM and PTM research, driving future hypothesis-driven experiments focused on uncovering the rules governing these modifications.

Additionally, standardized data storage and the development of specific large-scale databases are necessary. For instance, central repositories for large-scale RNA sequencing (RNA-seq) and proteomics data have been established, such as the Gene Expression Omnibus (GEO), PeptideAtlas, ProteomeXchange, and the AlphaFold Protein Structure Database [11–13]. However, broad PRM and PTM data repositories are yet to be established. These data also need to be incorporated into visualization browsers so researchers can access these data in a user-friendly manner to aid hypothesis-driven experimental design. There should be a concerted effort to combine metabolomic data into these central repositories and data browsers, including the metabolic states of cells with and without intact systems for adding and removing PRMs and PTMs.

Finally, information regarding the affinity of the writers and erasers needs to be obtained and stored in these repositories. Overall, data acquisition, standardization, storage, and visualization are important parameters for driving synergistic PRM and PTM research.

Methodologies needed for future studies

Another theme noted was the need for new methodologies to answer the questions central to the synergy of PRM and PTM research due to the current limitations in all available technologies. Several methodologies were identified as truly necessary for driving these studies forward. For instance, approaches to simultaneously isolate and analyze the modified RNAs or proteins in conjunction with the pools of cellular metabolites from the desired samples are needed. These advances are necessary to study the relationships of various PRMs and PTMs with the pools of available cellular metabolites, allowing one to determine the effect of inhibiting the removal of a specific PRM or PTM on the cellular pool of that metabolite. It was also noted that it is necessary to develop chemical probes and related methodologies for determining the specific nucleotides and amino acids that are modified. Developing these resources will

enhance our capability to identify *bona fide* modification sites.

Current methodologies usually provide information on single modifications linked to a specific nucleotide or amino acid. However, a single RNA or protein molecule can have several modifications simultaneously. Thus, there is a need for tools that can define all the modifications that occur on a single RNA or protein. 'Top down' proteomic approaches can provide this for proteins but are largely limited to modifications with a lower molecular weight [3]. Additionally, the development of ML and AI approaches for identifying and characterizing modifications was determined to be a major future focus [14,15].

Altogether, the development of novel approaches will be necessary for synergizing and driving future PRM and PTM research, which hopefully can uncover the rules and true importance of these regulatory signals for RNA and protein molecules (Figure 1), which will be driven by the necessary development of technologies with a higher throughput and higher resolution for identifying and characterizing PRMs and PTMs (Box 1).

Recommendations for future collaboration and synergy

One of the most important ideas identified was the need to create a collaborative network of scientists focusing on PRMs and PTMs (Figure 1). This network would obtain funding to run yearly meetings and workshops to continually reunite these groups to discuss recent progress and to continue setting goals and milestones to keep the synergy between these scientists. Relatedly, conferences that are jointly focused on PRM and PTM research should also be a yearly focus. These events would allow these scientists to continue discussions on the topics described earlier. Such discussions would help ensure that future studies on PRMs and PTMs are aimed to address common goals, while also ensuring that all datasets can be utilized and reanalyzed by all. These gatherings would provide a continual forum for determining future research goals in this research area.

Additionally, funding mechanisms to provide research support to drive synergistic projects forward were seen as necessary to coalesce PRM and PTM research. For instance, one idea was for funding agencies to promote funding of large-scale multi-principal investigator (PI) and multi-institution proposals aimed at addressing fundamental questions that unite these areas of inquiry. These funding opportunities were suggested to include interactions with industry partners, especially where the development of new technologies would be driven by such partnerships. The establishment of industry partnerships is likely to accelerate the development of novel technologies that will allow significant growth and development of this research. Future funding should also be focused on rewarding researchers who make their analytical tools, research methodologies, novel technologies, and large-scale databases readily accessible and user-friendly. Notably, cross-agency funding initiatives would be an excellent mechanism to push the synergistic research goals outlined earlier. In general, funding mechanisms that focus on developing synergies between PRM and PTM researchers are needed for future synergism in these research areas.

In closing, the study of PRMs and PTMs has the potential to become a highly collaborative and integrated area of research inquiry.

Box 1. Technologies needed for synergistic PRM and PTM studies

The development of novel technologies that will allow newer and finer-scale data to be obtained are needed to drive the synergism of PRM and PTM research forward. The technological advances identified as needed to drive this research are outlined here.

- Algorithms to determine all RNA modifications from nanopore-based direct RNA sequencing (RNA-seq) results.
- Increased signal and output from nanopore-based sequencing approaches to aid algorithmic development for detecting PRM.
- Improved mass spectrometry-based technologies to allow untargeted analysis of RNA and protein modifications, which would allow the identification of multiple or all modifications on the same RNA or protein molecule simultaneously.
- Mass spectrometry advances to allow analyses of PRMs and PTMs from single cells with a higher throughput and a deeper resolution.
- The production of small molecules that can inhibit or act as agonists to specific PRMs and PTMs to aid the validation of *bona fide* sites and functional characterization.
- Development of high-throughput CRISPR tools for RNA and protein modification site mutagenesis to streamline validation of *bona fide* sites.
- Identification and/or engineering of reverse transcriptase enzymes that will insert easily identifiable base misincorporation events into RNA-seq libraries for high-throughput identification of modified RNA bases.
- Dedicated curated online databases with a broad focus on PRMs and PTMs backed by community-driven standards ensuring high-quality data.
- Technologies allowing the visualization of 3D structures of modified RNA and protein molecules in their cellular environment, such as improved NMR spectroscopy-based technologies.

PRM/PTM Consortium information

Sara M. Assmann, Pennsylvania State University, PA, USA; Philip C. Bevilacqua, Pennsylvania State University, PA, USA; Kyle Biggar, Carleton University, ON, Canada; Amanda Bryant-Friedrich, Wayne State University, Detroit, MI, USA; Steven P. Briggs, University of California, San Diego, CA, USA; Yue Chen, University of Minnesota, Minneapolis, MN, USA; Xuemei Chen, University of California, Riverside, CA, USA; Vitaly Citovsky, State University of NY, Stony Brook, NY, USA; Ana Conesa, University of Florida,

Gainesville, FL, USA; Lydia M. Contreras, University of Texas, Austin, TX, USA; Matthew DeLisa, Cornell University, Ithaca, NY, USA; Xinnian Dong, Duke University, Durham, NC, USA; Cristina M. Furdui, Wake Forest School of Medicine, Winston-Salem, NC, USA; Benjamin Garcia, Washington University in St. Louis, ST. Louis, MO, USA; Wendy Gilbert, Yale School of Medicine, New, Haven, CT, USA; Chuan He, University of Chicago, Chicago, IL, USA; Ya-ming Hou, Thomas Jefferson University, Philadelphia, PA, USA; Samie R. Jaffrey, Cornell University, Ithaca, NY, USA; Ralph E. Kleiner, Princeton University, Princeton, NJ, USA; Kristin S. Koutmou, University of Michigan, Ann, Arbor, MI, USA; Brian K. Law, University of Florida, Gainesville, FL, USA; Yoosook Lee, University of Florida, Gainesville, FL, USA; Julius Lucks, Northwestern University, Evanston, IL, USA; Ping Ma, University of Georgia, Athens, GA, USA; Matthias Mann, Max Planck Institute of Biochemistry, Planegg, Germany; Julie Maupin-Furlow, University of Florida, Gainesville, FL, USA; Robert L. Moritz, Institute for Systems Biology, Seattle, WA, USA; Andrew Nelson, Boyce Thompson Institute, Ithaca, NY, USA; Robert H. Newman, North Carolina A&T University, Greensboro, NC, USA; Tao Pan, University of Chicago, Chicago, IL, USA; Laurie Parker, University of Minnesota, Minneapolis, MN, USA; Boone M. Prentice, University of Florida, Gainesville, FL, USA; Christopher M. Rose, Genentech Inc., San Franciso, CA, USA; Sara H. Rouhanifard, Northeastern University, Boston, MA, USA; Millicent O. Sullivan, University of Delaware, Newark, DE, USA; Michael R. Sussman, University of Wisconsin, Madison, WI, USA; Klaas van Wijk, Cornell University, Ithaca, NY, USA; Rongsheng E. Wang, Temple University, Philadelphia. PA, USA; Timothy A. Whitehead, University of Colorado, Boulder, CO, USA; Christina M. Woo, Harvard University, Boston, MA, USA; Aaron T. Wright, Pacific Northwest National Laboratory, Richland, WA, USA; John Yates, Scripps Research Institute, San Diego, CA, USA; Wei Zhang, University of Central Florida, Orlando, FL, USA.

Acknowledgments

The authors thank all participants and speakers that attended the National Science Foundation (NSF)-funded workshop, which resulted in the discussions documented herein. This workshop, discussion, and publication were supported by a US NSF grant (IOS-2135914) to S.C. and B.D.G. Z.T.B. is supported by a Graduate Research Fellowship from the US NSF (DGE-2040434).

Declaration of interests

The authors have no interests to declare.

[1]Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80305, USA
[2]Department of Medicinal Chemistry, University of Kansas, Lawrence, KS 66044, USA
[3]Department of Chemistry, Center for RNA Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA
[4]Department of Biology, Oxford College of Emory University, Oxford, GA 30054, USA
[5]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[6]Department of Pharmacology, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA
[7]Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, USA
[8]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA
[9]Remix Therapeutics, Watertown, MA 02472, USA
[10]Intercollege Graduate Degree Program in Plant Biology, Department of Biology, and Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA
[11]Department of Biology, University of Mississippi, University, MS 38677, USA
[12]These authors contributed equally to this work.
[13]The authors have been ordered alphabetically.

*Correspondence:
schen8@olemiss.edu (S. Chen) and
bdgregor@sas.upenn.edu (B.D. Gregory).

https://doi.org/10.1016/j.tibs.2023.09.011

© 2023 Elsevier Ltd. All rights reserved.

References

1. Roundtree, I.A. *et al.* (2017) Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200
2. Sharma, B. *et al.* (2023) The diversity and functions of plant RNA modifications: what we know and where we go from here. *Annu. Rev. Plant Biol.* 74, 53–85
3. Stillman, B. (2018) Histone modifications: insights into their influence on gene expression. *Cell* 175, 6–9
4. Bradley, D. (2022) The evolution of post-translational modifications. *Curr. Opin. Genet. Dev.* 76, 101956
5. Duan, G. and Walther, D. (2015) The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput. Biol.* 11, e1004049
6. Jumper, J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589
7. Kadumuri, R.V. and Janga, S.C. (2018) Epitranscriptomic code and its alterations in human disease. *Trends Mol. Med.* 24, 886–903
8. Tada, Y. *et al.* (2008) Plant immunity requires conformational charges of NPR1 via S-nitrosylation and thioredoxins. *Science* 321, 952–956
9. Viegas, I.J. *et al.* (2022) N6-methyladenosine in poly(A) tails stabilize VSG transcripts. *Nature* 604, 362–370
10. Yu, X. *et al.* (2021) Messenger RNA 5′ NAD+ capping is a dynamic regulatory epitranscriptome mark that is required for proper response to abscisic acid in *Arabidopsis*. *Dev. Cell* 56, 125–140
11. Deutsch, E.W. *et al.* (2019) The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* 48, D1145–D1152
12. Deutsch, E.W. *et al.* (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* 9, 429–434
13. Varadi, M. *et al.* (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444
14. Demichev, V. *et al.* (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* 17, 41–44
15. Gessulat, S. *et al.* (2019) Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* 16, 509–518