Published in final edited form as:

Nat Struct Mol Biol. 2014 November ; 21(11): 945–946. doi:10.1038/nsmb.2911.

An expanding universe of mRNA modifications

Samie R. Jaffrey

Department of Pharmacology, Weill Cornell Medical College, Cornell University, New York, NY 10065, USA. srj2003@med.cornell.edu

Abstract

The fate of mRNA can be regulated by internal base modifications, with the currently known modified bases being *N*⁶-methyladenosine, 5-methylcytosine, and inosine. Three new studies show that yeast and human mRNA also contain pseudouridine residues and that pseudouridylation is induced in various stress states, hinting at a new pathway for post-transcriptional control of mRNA.

When we think of pathways that regulate mRNA, we usually think of mRNA-binding proteins or microRNAs that get recruited to the mRNA in response to cellular signaling events. However, an emerging model is that mRNA can be regulated by base modifications, analogous to amino acid modifications in proteins. The most well-characterized mRNA base modification pathway is RNA editing, which involves the conversion of adenosine to inosine¹. Adenosine residues can also be methylated to *N*⁶-methyladenosine (m6A)^{2,3}. Definitive demonstration of the presence of m6A came with the recent development of transcriptome-wide m6A mapping technologies that showed that m6A residues are present in at least 8,000 transcripts, often near stop codons or in the 5'UTR of transcripts^{4,5}. This, along with the recent mapping of 5-methylcytosine sites in mammalian mRNA^{6,7}, points to the existence of a diverse landscape of “epitranscriptomic” modifications that influence mRNA fate and function in cells.

The prevalence of these mRNA base modifications raises the question of whether there are other base modifications with signaling functions. Pseudouridine is a particularly compelling candidate. Pseudouridine is found in tRNA and rRNA, as well as small nuclear RNAs (snRNAs) and other noncoding RNAs (ncRNAs), which makes it the most abundant modified base in cells⁸. Pseudouridine is formed by a remarkable mechanism involving uracil base detachment, flipping, and reattachment to the ribose⁹ (Figure 1). This is catalyzed by either dyskerin (or its yeast homolog Cbf5) pseudouridine synthase or by a family of pseudouridine synthase (Pus) enzymes⁸. Pus family enzymes pseudouridylate uridines in specific sequence contexts. On the other hand, Cbf5/dyskerin pseudouridylates in concert with one of many different small nucleolar RNAs (snoRNAs), which guide the Cbf5/dyskerin complex to target RNAs, and direct pseudouridylation to a central uridine residue (Figure 1). The idea that mRNA might also be pseudouridylated was originally suggested by the finding that some snoRNAs lack complementarity to ncRNA targets but match mRNAs^{10,11}. Additionally, 23 putative pseudouridine synthase genes are predicted in human, with unclear function¹². Thus, the presence of putative pseudouridine synthase

enzyme activities without known targets raises the tantalizing possibility that their endogenous substrates might be mRNA⁸.

Perhaps more intriguing is that pseudouridylation can be regulated. Two uridine residues in the U2 snRNA are individually pseudouridylated by various stress stimuli in yeast, such as heat shock and nutrient deprivation¹³, and rRNA pseudouridylation at two sites is regulated by mTOR in mammalian cells¹⁴. The idea that pseudouridylation is controlled by signaling pathways makes pseudouridine appealing as a potential mRNA modification⁸.

Three groups now report the first transcriptome-wide maps of pseudouridine, revealing the pseudouridine landscape in yeast and the human transcriptome^{15–17}. These groups each used a primer-extension based assay to detect pseudouridine. This assay uses CMC, a chemical that modifies numerous bases on nitrogen atoms. CMC adducts are readily reversed with alkali, except CMC-pseudouridine adducts, which are stable and abort cDNA synthesis by reverse transcriptases¹⁸. The position of pseudouridine in various RNAs can be determined by mapping the termination sites of these cDNAs¹⁸.

The three groups used CMC to map pseudouridine on a transcriptome-wide scale. This approach, called PSI-Seq, Pseudo-Seq, and Ψ -Seq, identified ~50–100 pseudouridine sites in approximately as many mRNAs in yeast grown in non-stress conditions, and ~100–400 pseudouridine sites in human cell lines. Pseudouridine was also detected in ncRNAs. The different number of pseudouridine sites called in the different studies relates to the read depth and different stringency criteria used to call a pseudouridine site. Notably, pseudouridines are not concentrated in specific regions of transcripts, but can be found in 5' untranslated regions, coding sequences, and 3' untranslated regions^{16,17}.

One the most interesting findings is that the levels of pseudouridine in mRNA increases substantially in response to stress states in yeast. Both Lovejoy *et al.* and Schwartz *et al.* used heat shock and found marked increases in the number of pseudouridine sites. Schwartz *et al.* found 265 new pseudouridine sites, most of which were formed in a Pus7p-dependent manner, similar to the Pus7p-dependent pseudouridylation of U2 in heat shock^{13,15}. Carlile *et al.* used nutrient stress, finding that the number of pseudouridine sites nearly doubled¹⁶. The majority of the sites were dependent on Pus1p and Pus7p, although other Pus family members also contributed to the pattern of pseudouridylation in these cells¹⁶. Although the overall number of pseudouridine sites is substantially smaller than what is seen with other modifications such as m6A, the targeted pseudouridylation of a cohort of transcripts could allow them to be coregulated in cellular stress responses.

In addition to Pus enzymes, Cbf5/dyskerin-mediated pseudouridylation contributed to the pseudouridine profile, but did not appear to mediate stress-mediated increases in pseudouridine¹⁵. Instead, Cbf5/dyskerin appears to mediate the smaller number of baseline pseudouridines seen in non-stress conditions. Importantly, not all Cbf5/dyskerin-dependent sites could be linked to a canonical snoRNA¹⁵, suggesting that other snoRNAs might contribute to mRNA pseudouridylation.

An exciting aspect of these studies is the potential disease relevance. X-linked dyskeratosis congenita and mitochondrial myopathy and sideroblastic anemia are each associated with

mutations in pseudouridine synthases^{19,20}. snoRNA42 was recently shown to act as an oncogene in lung cancer²¹. Profiling pseudouridine in these diseases might begin to provide a hint as to whether misregulation of mRNA pseudouridylation contributes to these diseases.

The main question is whether pseudouridine in mRNA is biologically meaningful. The apparent lack of a dedicated mRNA pseudouridylase raises the possibility that these pseudouridines could reflect nonspecific pseudouridylation. For example, a stress-induced upregulation in pseudouridylase activity that is meant to be directed towards rRNA, snRNA, and other ncRNAs might unavoidably modify mRNAs that coincidentally have the same short pseudouridylation-directing motifs.

To address this, Lovejoy *et al.* asked if the sequences surrounding the pseudouridine in *RPL11a* and *TEF1* show evolutionary conservation across diverse fungi¹⁷. They found nearly complete conservation, which matched the conservation level seen surrounding the pseudouridine site in the U2 snRNA. A transcriptome-wide analysis of the sequence conservation surrounding pseudouridine sites in mRNA could provide further support for functional relevance.

Additional support for a biological role would come from the demonstration of a function for pseudouridine. Since pseudouridine basepairs with adenosine, and pseudouridine-containing transcripts are translated into functional proteins in living cells²², pseudouridylation does not appear to change the encoded protein sequence. Earlier studies using artificial pseudouridylation showed that pseudouridine at stop codons leads to readthrough²³. However, pseudouridylation of a stop codon was only observed in one transcript¹⁵, indicating that this is not its main role. Alternatively, pseudouridine could recruit an as-of-yet unknown pseudouridine-binding protein, or influence RNA structure due to its altered base pairing properties⁸.

To explore a role for pseudouridine, Lovejoy *et al.* examined *RPL11a* and *TEF1*, which are pseudouridylated by Pus1p and Pus4p respectively¹⁷. They found no change in the encoded protein sequence and they failed to see changes in mRNA or protein abundance in the Pus deletion strains¹⁷. However, Schwartz *et al.* observed that mRNAs that contained heat shock-induced Pus7p-dependent pseudouridine sites were 25% more highly expressed compared to the same mRNAs in Pus7p-deficient yeast cells¹⁵. This raises the possibility that pseudouridine could stabilize mRNA.

To ultimately get to the bottom of a potential role for pseudouridine, the most straightforward approach will be to mutate pseudouridine sites. Many pseudouridylated mRNAs may have pseudouridine at a very low stoichiometry and mutagenesis might not reveal a role for this modification. Thus, identification of the transcripts with the highest pseudouridine stoichiometry, and mutagenesis of those uridines, will likely reveal functions of pseudouridine.

REFERENCES

1. Paul MS, Bass BL. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* 1998; 17:1120–1127. [PubMed: 9463389]

2. Perry RP, Kelley DE. Existence of methylated messenger RNA in mouse L cells. *Cell*. 1974; 1:37–42.
3. Desrosiers R, Friderici K, Rottman F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci U S A*. 1974; 71:3971–3975. [PubMed: 4372599]
4. Meyer KD, et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*. 2012; 149:1635–1646. [PubMed: 22608085]
5. Dominissini D, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. 2012; 485:201–206. [PubMed: 22575960]
6. Squires JE, et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res*. 2012; 40:5023–5033. [PubMed: 22344696]
7. Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat Biotechnol*. 2013; 31:458–464. [PubMed: 23604283]
8. Ge J, Yu YT. RNA pseudouridylation: new insights into an old modification. *Trends Biochem Sci*. 2013; 38:210–218. [PubMed: 23391857]
9. Hoang C, Ferre-D'Amare AR. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*. 2001; 107:929–939. [PubMed: 11779468]
10. Cavaille J, et al. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A*. 2000; 97:14311–14316. [PubMed: 11106375]
11. Huttenhofer A, Brosius J, Bachellerie JP. RNomics: identification and function of small, non-messenger RNAs. *Curr Opin Chem Biol*. 2002; 6:835–843. [PubMed: 12470739]
12. Hunter S, et al. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*. 2012; 40:D306–D312. [PubMed: 22096229]
13. Wu G, Xiao M, Yang C, Yu YT. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J*. 2011; 30:79–89. [PubMed: 21131909]
14. Courtes FC, et al. 28S rRNA is inducibly pseudouridylated by the mTOR pathway translational control in CHO cell cultures. *J Biotechnol*. 2014; 174:16–21. [PubMed: 24480570]
15. Schwartz S, et al. Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell*. 2014; 159:148–162. [PubMed: 25219674]
16. Carlile TM, et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*. 2014
17. Lovejoy AF, Riordan DP, Brown PO. Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in *S. cerevisiae*. *PLoS One*. 2014 **In press**.
18. Bakin A, Ofengand J. Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry*. 1993; 32:9754–9762. [PubMed: 8373778]
19. Heiss NS, et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet*. 1998; 19:32–38. [PubMed: 9590285]
20. Bykhovskaya Y, Casas K, Mengesha E, Inbal A, Fischel-Ghodsian N. Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *Am J Hum Genet*. 2004; 74:1303–1308. [PubMed: 15108122]
21. Mei YP, et al. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene*. 2012; 31:2794–2804. [PubMed: 21986946]
22. Kariko K, et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol Ther*. 2008; 16:1833–1840. [PubMed: 18797453]
23. Huang C, Wu G, Yu YT. Inducing nonsense suppression by targeted pseudouridylation. *Nat Protoc*. 2012; 7:789–800. [PubMed: 22461068]

Figure 1.

Pseudouridylation enzymes that introduce pseudouridine in mRNA. **(a)** Schematic showing the 180° flip of the uracil base that is common to the mechanism of all pseudouridine synthases. **(b)** RNA-independent and snoRNA-dependent pathways for pseudouridylation contribute to pseudouridine levels in mRNA. Diverse Pus enzymes introduce pseudouridine in mRNA, with Pus7p having a major role in heat-shock induced pseudouridylation. Shown are four Pus family members. The Pus1p and Pus2p lack a clear consensus site for pseudouridylation, while Pus4p and Pus7p recognize a short sequence motif that directs pseudouridylation. Pseudouridylation mediated by the Cbf5/dyskerin pathway requires a snoRNA that guides the Cbf5/dyskerin complex to a target RNA by base complementarity and directs methylation of a central uridine.