
The importance of being modified: an unrealized code to RNA structure and function

PAUL F. AGRIS

The RNA Institute, State University of New York, Albany, New York 12222, USA

A, U, C, and G are boring. Truly, four nucleosides sometimes repeated in seemingly endless sequence has a numbing effect even in a computational and structural context. However, there is nothing boring about the chemistries and structures of post-transcriptional modifications of RNA as indicated by the rising interest and rapidly increasing numbers of publications. A plethora of papers demonstrate that modified nucleosides of RNA are functionally indispensable for RNA processing and translation. Over the last 20 years, efficiency, accuracy, enhanced activity and responsiveness are terms that accompany studies of modifications. Modifications are found in all types of RNAs, in every structural motif and in unstructured regions. The large numbers of modifications in tRNAs are quite familiar. The presence, chemistries and RNA sequence locations of modified nucleosides are only now beginning to be appreciated as both temporal—in response to environmental (internal and external) signals—and spatial, in their effect on RNA conformation, dynamics, and functions. Modifications can be recognition determinants for proteins as well as incredible determiners of RNA structure and function. Unfortunately, we tend to neglect the modified nucleosides. Using amplification and transcription *in vitro* that disregard the native modified nucleosides, we study the interesting low abundance, non-coding RNAs for which one molecule per cell is sometimes enough to affect gene expression. On the other hand, we lack technologies sensitive enough to explore modifications and their functions in these same low abundance RNAs. Our lack of understanding of RNA modification regrettably limits innovation in the use of modified nucleosides for synthetic biology and medicine. However, the past two decades have led to some of the most interesting of findings.

Modifications change everything

The presence of modified nucleosides in RNA has been recognized for more than half a century. Fifty years ago, inosine was identified by Holley's lab at the first position of the anticodon

of yeast tRNA^{Ala} and quickly became the hallmark of Crick's Wobble Hypothesis. Wondrous as that seemed at the time, we now know that RNA editing, deaminations of A to I and C to U, occur in other RNAs, particularly mRNAs. The deamination of A to I at tRNAs' anticodon's Wobble position allows tRNA to read three synonymous codons, NNU, NNC, and NNA. The non-canonical I•A base pair begs the question of how a purine•purine interaction is accommodated at the decoding site of the ribosome. With ribosome crystal structures, Ramakrishnan's and our group showed that the decoding site accommodates the distorted anticodon-codon double helix of I•A and extensively modified anticodon•codon base pairs. These biochemical and biophysical results certainly seem to be a simple enough explanation as to how the wobble position inosine and modified pyrimidines allow a number of tRNAs to read cognate and synonymous codons. However, nature did not leave well enough alone even when it comes to the inosine adaptation of tRNAs. Our lab demonstrated that 2-thiocytidine occurring in the anticodon loop of the isoaccepting tRNA^{Arg1} species negates inosine's ability to read adenosine while allowing the recognition of U and C. Thus, sometimes multiple modifications that occur close in three dimensional space work together to affect RNA conformation, dynamics and function. Wobble at the third position of the anticodon together with modification chemistry and structure enable some 40 cytoplasmic tRNAs to translate the 61 Universal Genetic Codes, and sometimes the altered use of the three termination codes. In the human mitochondria, only 22 tRNAs with distinctive modified anticodons are used for decoding.

Modifications reflect amino acids

RNA's modified nucleosides are more than simple modulators of structure and function. The genetic memory, material requirements and energy used in producing modified nucleosides by all organisms has evolved into a substantial, evolutionarily conserved and ubiquitous investment. If the chemistries of RNA modifications and their effects on structure seem familiar, they should. The over 100 post-

Corresponding author: pagris@albany.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.050575.115>. Freely available online through the RNA Open Access option.

© 2015 Agris This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

transcriptional modifications of RNA have chemistries akin to those of the amino acid side chains: hydrophilic, polar and charged, and hydrophobic, aliphatic and aromatic. In 1996, we attributed 10 physicochemical properties of modification chemistry, including spatial effects, to their functions. Though some modifications are chemically complex and physically large, with biosyntheses requiring sequential and multiple enzymatic activities, other modifications represent simple but still very important changes. Deamination of A to I in mRNA editing and tRNA anticodons is not the only relatively modest alteration that results in a large effect on structure and function. Steitz's lab showed that pseudouridine in U2 snRNA is important for the processing of the mRNA at the intron branch point A. In this journal 15 years ago, we demonstrated that a single atom change from 2-keto oxygen to 2-sulfur of uridine at tRNA's wobble position restricted codon recognition through a change in conformation and dynamics. Other modifications contribute considerable chemical and physical changes. These include amino acid derivations of nucleobases. Yokoyama and co-workers illustrated that the post-transcriptional addition of lysine to the wobble position C in the CAU anticodon of a tRNA^{Ile} isoacceptor promotes isoleucine aminoacylation, inhibits methionine misacylation and negates the tRNA^{Ile} recognition of the methionine codon AUG in favor of the isoleucine codon AUA. Björk and others demonstrated that chemically complex and physically large modifications at both tRNA's wobble position and 3'-adjacent to the anticodon are important to maintaining the translational reading frame. The amino acid threonine modification of A is conserved across the three domains of life. It is located 3'-adjacent to anticodons in those tRNAs that read codons beginning with A. We showed this modification stabilizes the first anticodon-codon base pair, U•A, by straddling the base pair as a hydrophobic platform within 3 Å.

Modifications control RNA conformational dynamics at the site of modification, spatially, and at the time of modification, temporally. Modification directs the RNA into the conformation most needed for the task at hand, whether that be snRNAs, mRNAs or perhaps, miRNAs or lncRNAs. The last 20 years has shown that the simplest of modification chemistries are the most ubiquitous elements of RNA. Very early, the ribothymidine and dihydrouridine arms of tRNA's structure were named for the presence and locations of these modifications in tRNA. However, the sequences of the ribosome's peptidyl transferase site differ among organisms, but modifications are found conserved within the functional structure. The positions of pseudouridines and 2'-O-methyls in many types of RNA, rRNAs, snRNAs, from different organisms also are conserved. These most ubiquitous of modifications are often found to facilitate RNA folding, structure, and conformational dynamics. The temporal aspect of these and other modifications are only now becoming recognized for their importance to cell biology.

The modified wobble hypothesis we presented in 1991 invoked the chemical and structural contributions of modified

nucleosides to account for tRNA being either constrained in its codon recognition, or relaxed in its abilities to read multiple codons. Modified nucleoside prestructure RNA for function. The prestructuring lowers energy barriers to RNA/RNA and RNA/protein interactions. The fact that many modification enzymes are at the ready when RNA is transcribed makes one ponder whether the modification of any RNA can be an immediate temporal response to internal and external signals in a time frame shorter than that of gene expression. While RNA is inherently transient, we now know that some modifications are even more so with the discovery of demethylations. Thus, modifications are chemistries that allow cells to respond to signals without the need for additional transcription and then are lost when the RNA is degraded only to re-emerge! Modifications are temporally more responsive than the transcription apparatus, and result in conformational diversity, and population heterogeneity. The modified nucleosides are disposable; they are not recognized by native RNA polymerases *in vivo*. Thus, post-transcriptional modification can be responsive to signals, as well as be important to the most conserved of "housekeeping" RNA species.

Questions remaining: Technological innovation generates sophisticated fundamental and biomedical inquiries

There will be inspiring but difficult evolutionary questions to answer. In the RNA World were RNA polymerases capable of incorporating modified nucleosides in contrast to the protein-based polymerases of today? Was post-transcriptional modification of RNA first achieved by ribozymes? There are questions important to today's biomedical applications that are more readily approached. Do the least abundant of RNAs, the ncRNAs regulating gene expression, contain modified nucleosides and if so which modifications and for what purposes? Now we are capable of detecting and identifying 10–100 attomol of a modification obtained from fmole of RNA in tens of thousands of cells, rather than from millions. Thus, single cell analysis of modified nucleosides is feasible. It will provide insights into modified nucleoside function in the least abundant RNAs. Other attainable goals would be the technological advancement of a high throughput sequencing of low abundant RNAs that includes sequence locations and identification of modified nucleosides. Do changes in modification over time relate to cell signaling in normal biology and in diseased cells? The tracking of modifications in time, as a consequence of cells receiving signals that affect differentiation and development, stress response, infection, etc., is also an important goal. Since modification amounts and chemistries respond to signals and environmental insults, epitranscriptomics or RNA epigenetics, our ability to identify and quantify native and unnatural modified nucleoside changes may reveal a novel and distinctive set of biomarkers for health and disease. In fact, it is not too much of a stretch to propose that non-native modifications resulting from

chronic use of legal and illegal drugs could cause intolerance in the former and addiction in the latter. The more we understand about RNA modifications, the more we will be able to use them in the design and implementation of new RNAs, in synthetic biology and in biologic therapeutics. Already modified nucleosides have been employed for delivery and stability of siRNAs, miRNAs, and mRNAs important for abating the body's immune response to therapeutics. More recently, we find that modified nucleoside character, location, and proximity to each other can be identity determinants for protein recognition. Exploration of exciting modified nucleoside chemistries as a Third Code is inspiring and motivating. It holds the promise of an in-depth understanding of fundamental RNA biology, answers to important present and future biomedical questions, and to the use of modified nucleosides in synthetic biology.

Final thought: Modifications represent an additional, operational code, a Third Code

The Universal Genetic Code provided insight into the decoding of genes, but as demonstrated by Söll and others is not frozen in time. Nucleosides proximal to tRNA's 3'-aminoacyl-terminus, including the discriminator nucleotide (position 73), are recognition determinants for aminoacylation and as such are a second code discovered by Schimmel. RNA modifications as determiners of functional structure, indicators of cellular function and health and a distinctive set of biomarkers for health and disease, are a third code. We are gaining insights into this code with improvements in analytical sensitivity and high throughput methods that relate modified nucleoside chemistry and structure to cellular function and human health issues.