RNA modification: the Golden Period 1995–2015

HENRI GROSJEAN

Institute for Intégrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, F-91198 Gif-sur-Yvette cedex, France

The study of RNA modification started in 1957 when the first modified nucleoside was discovered in bulk yeast RNA, pseudouridine: a fifth nucleotide, was born. After a long period of arduous and somewhat frustrating works, remarkable results have been obtained in the last two decades, made possible by new technology developed in many laboratories, and have engendered a renaissance of interest in the field. Research on RNA modification has always been and remains a real chest of treasures. I am happy to add my personal comments to this anniversary edition of *RNA*.

In 1995, 93 naturally occurring modified nucleosides were known. They were identified in molecules that were abundant and stable in the cell, namely tRNAs, rRNAs, and snRNAs. Today the count is 114. The number increases to 141 if we include residues that are doubly modified, like those containing a modified base and a methylation of the ribose (see: http://modomics.genesilico.pl/ and http://mods.rna.albany.edu). Initially, however, just a few modification enzymes were identified and characterized. Their investigation was limited by the presence of these enzymes in small cellular amounts, making them difficult to detect, let alone purify in sufficient amounts for enzymatic characterization. Additionally, the functional relevance of RNA modified was unclear because the genes coding for modification enzymes could be genetically deleted with only modest or no effect on cell growth. Fortunately, interest in this research was sustained by the demonstration of a functional importance of several modified nucleotides in the anticodon loop of tRNA, namely the wobble nucleotide (position 34) and the nucleotide (position 37) that is next to the third anticodon nucleotide. Modifications at these sites correlate with clear-cut increases in the efficacy and accuracy of translation, a process that is termed translational fine-tuning.

A dramatic burst into systematic analyses of RNA modifications and their functional consequences followed in the footsteps of the genomics revolution. This stage opened with the genome sequence of *Heamophilus influenza* in 1995. Today, more than six thousands of completed genome sequences are known, and they span the gamut of phylogenetic diversity (see: https://gold.jgi-psf.org). The availability of techniques for facile cloning and purification of recombinant genes, the over-expression of cloned genes providing copious amounts of their protein enzymes, and the ability to produce large amounts of RNA transcripts that can be used as substrates for the enzymes, have launched rapid development of our knowledge of RNA modifications. Additionally, bioinformatics tools have identified a large collection of genes coding for orthologous RNA modification enzymes. To date, more than 160 modification enzymes (not counting transient "helpers" protein and components of multi-protein complexes) have been identified. The activities and specificities of many of these enzymes have been determined both in vitro and in vivo (see: http://modomics.genesilico.pl). The year 1996 saw the first X-ray crystal structure of an RNA modification enzyme. The protein was an archaeal tRNA transglycosylase (aTgt enzyme) which catalyses the formation of archaeosine (G^+). A co-crystal structure of aTgt complexed with its tRNA substrate was published later in 2003. To date, the crystal structures of 70 core-enzymes and a few multi-protein complexes have been solved at high resolution. Many of them are described in complexes with their cofactors and/or their respective cognate RNA transcripts.

Bioinformatics analyses identified orthologs and paralogs existing in genes as in the translated proteins showing characteristic domains, motifs and conserve residues within modification enzyme sequences that are important for enzyme activity and specificity. Evolutionary links between several RNA editing enzymes and methylases acting on either RNA or DNA are also evident, suggesting common evolutionary origins.

Great diversity exists in modification enzyme structure and function. Some enzymes act while the target RNA is undergoing transcription, while others act at later stages, usually at the level of the full-length RNA. Some modification enzymes act alone on their RNA target, whereas others are part of multi-protein complexes. The latter complexes can exist either transiently or as stable entities. One remarkable such entity involves enzyme polypeptides bearing no intrinsic ability to recognize an RNA substrate alone. Together with other proteins of the multi-component complex, they

---

cooperate with a small non-coding RNAs (named snoRNAs) to guide to function *in trans* for substrate RNA recognition, via Watson–Crick base-pairing interactions. These RNA-guided enzymes, discovered in 1996–97, catalyse the methylation of the 2′-hydroxyl ribose moiety, or the isomerisation of uridine into pseudouridine in rRNAs and snRNAs in eukaryotes; they also can act on tRNAs and rRNAs in Archaea. However, this "RNA-assisted" modification strategy has thus far not been observed in bacteria or organelles.

Aside from this RNA-dependent strategy, the molecular uses for substrate recognition by other RNA modifications remains largely unexplored territory. Most RNA modification enzymes are typically site-specific, while a few others operate on multi-sites of similar RNA substrates. Some enzymes are class-specific, while others demonstrate dual-specificity, acting on different types of RNA (e.g., tRNA and rRNA or snoRNA, tRNA and mRNA). In three cases, tRNA m^5U-methyltransferase (TrmA), tRNA transglycosylase (Tgt), and rRNA pseudouridine synthase (RluA), high-resolution X-ray crystallographic studies provide evidence for conformational changes in the protein and/or the substrate RNA upon catalysis. However, to unveil the dynamic aspect of RNA recognition, fast-kinetic techniques (e.g., stopped-flow or T-jump kinetics), and the more recently developed single-molecule technology, would be more valuable tools. Finally, while covalent addition of chemical moieties to RNA had long been considered as permanent modifications, the existence of RNA demethylases, analogous to those known for DNA, have been recently demonstrated (ecAlkB in 2003 and mammalian FTO/AlkBH5 in 2011). This remarkable discovery strongly supports the idea that particular nucleoside modifications can be under dynamic regulatory control.

The past two decades have also witnessed considerable progress in the identification and characterisation of new modified nucleosides in RNAs, especially in newly studied RNA of Archaea. Earlier methods mainly relied on analysis of nucleosides digests of purified RNA by thin-layer cellulose plates or liquid-column chromatography. The recent development of highly sensitive mass spectrometry based on triple quadripole detection (MS/MS), coupled with high performance liquid chromatography (LC), has considerably facilitated the detection of minute amounts of non-canonical nucleosides in RNA digests (2011). Also, combining LC-MS and LC-MS/MS with isotopomers (labelled with ^{13}C, ^{15}N or ^{18}O) allows detection of small differences between RNA samples (2009–2012). These new methodologies revealed novel biosynthetic pathways for RNA modifications and have led to the recent and exciting discovery of clear-cut signature changes in the spectrum and stoichiometry of RNA modification in response to different cell conditions, and in response to genetic changes in modification enzymes. Another useful method, yet more ancient (1977), is primer extension coupled with reverse transcription for the mapping

of modified nucleosides whose bases do not typically form stable base pair. If the RNA sequence is known, the stoichiometry of modification at a given position in the polymers is simply revealed by measuring the proportion of read-through product. Newly developed high-throughput sequencing methods and techniques allowing the isolation of RNAs containing any given modified nucleosides (2011–present) are revolutionizing our way of mapping certain modified bases and analyse their landscape profile in any cellular setting, not only in a limited numbers of model RNAs as before, but now in the whole cellular transcriptome. This new approach has recently allowed the detection in the eukaryotic transcriptome of inosine, m^5C, pseudouridine, and m^6A in RNA other than the already well-known t+r+snRNAs. The ubiquitous presence of modifications in many, if not all cellular RNA, including messenger (mRNA), within intronic RNA, and small or long intergenic non-coding RNA (lincRNA), is now evident. When applied to cells harboring a specific defect for a given modifying enzyme or cells subjected to stress or to any environmental condition, this transcriptome-wide analysis allows one to access information that was not accessible by earlier more traditional methods. One exiting breakthrough already obtained in this direction is the demonstration of a clear relationship between the dynamic character of the RNA modification and enzymatic de-modification systems (see above), and their active implication in transcriptional regulation of mRNA expression, notably reprogramming translation of selected and codon-biased mRNA sequences in response to changes in physiological or stress conditions. Applications of the latter methods have opened a new area of research with its own nickname: Epitranscriptomics.

In brief, biochemical processes leading to post-transcriptional RNA modifications are astonishingly diverse and conserved in all RNA classes and in all cell types. It is now clear that the combination of a multitude of tiny "decorations" or "outfits" in RNA, initially considered unimportant to cell metabolism, are now known to collectively play a strategic role in controlling the stability and quality of the cell-wide transcriptome, protecting individual RNA against undesirable molecular interactions (acting as molecular shields), improving the performance efficiency and specificity of the matured RNAs (fine-tuning), and serving as important biomarkers for epitranscriptomic regulations.

The next challenge will be to know how and why all these modification enzymes emerged during cellular evolution, where they are located in the crowded cell interior, how they are organized within that space, if they form multi-enzyme complexes (existence of modificosomes), how they are coordinate and finally how they regulate cell physiology. In relation to this last point, it is worth mentioning that a correlation exists between defects in RNA modifications and certain human pathologies. Modification enzymes may therefore constitute a new area of therapeutic intervention.