Molecular Omics

Interrogating the Transcriptome with Metabolically Incorporated Ribonucleosides

Journal:	*Molecular Omics*
Manuscript ID	MO-REV-08-2021-000334.R1
Article Type:	Review Article
Date Submitted by the Author:	05-Oct-2021
Complete List of Authors:	Kleiner, Ralph; Princeton University,

Interrogating the Transcriptome with Metabolically Incorporated Ribonucleosides

Ralph E. Kleiner

Department of Chemistry, Princeton University, Princeton, NJ, USA 08544

*rkleiner@princeton.edu

Abstract

RNA is a central player in biological processes, but there remain major gaps in our understanding of transcriptomic processes and the underlying biochemical mechanisms regulating RNA in cells. A powerful strategy to facilitate molecular analysis of cellular RNA is the metabolic incorporation of chemical probes. In this review, we discuss current approaches for RNA metabolic labeling with modified ribonucleosides and their integration with Next-Generation Sequencing, mass spectrometry-based proteomics, and fluorescence microscopy in order to interrogate RNA behavior in its native context.

1. Introduction

RNA plays a central role in biological processes. Gene expression programs are shaped by the dosage and spatial distribution of individual messenger RNAs as well as through the action of non-coding RNA species that mediate protein production. Achieving appropriate gene expression control over space and time involves multiple regulatory modalities that affect RNA behavior during its lifecycle -- transcription, maturation, export, translation and turnover[1-8]. Therefore, characterizing the mechanisms regulating RNA behavior are of fundamental importance to understanding cellular and organismal physiology and can reveal insight into disease-related processes.

The human genome encodes a vast complexity of RNA sequences spanning coding and non-coding RNAs – the transcriptome. While early studies in molecular biology focused on understanding the function and regulation of individual RNAs, technological advances in the form of DNA microarrays[9-11] and next-generation sequencing (NGS)[12] have ushered in the era of whole transcriptome analysis enabling a systems level “omics” understanding of RNA biology. The gold standard for transcriptome analysis is RNA sequencing or “RNA-seq”[13], which allows for the quantification of RNA transcript abundance across different biological conditions. Advances in NGS technology have provided increased throughput and lowered the cost of RNA-seq enabling its widespread application in biological studies with unprecedented depth of coverage. While canonical RNA-seq analysis has revolutionized our ability to probe RNA transcript abundance and variation across different conditions, it is limited in that it provides only a static “snapshot” of the transcriptome and does not report directly on RNA dynamics. Further, the study of additional features of RNA transcripts including their post-transcriptional modification state and interactions with associated RNA-binding proteins, which both have important roles in RNA transcript behavior and lifecycle, is typically inaccessible.

The incorporation of artificial functionality into biopolymers by exploiting the promiscuity of cellular metabolism and biosynthesis is a powerful approach to interrogate biological processes in their native context. This approach is known as metabolic labeling and has been applied to a variety of biomolecules in order to probe synthesis and turnover, subcellular localization, interaction partners, and other properties. RNA metabolic labeling with chemically modified ribonucleosides is a burgeoning field and provides unique opportunities for studying the transcriptome when combined with RNA sequencing, proteomics, and cellular imaging. In this review, we describe the methodological considerations for RNA metabolic labeling and highlight the available chemical probes and their uses. We end by discussing future prospects for the application of RNA metabolic labeling to transcriptomic analyses.

2. Considerations for metabolic labeling

The incorporation of artificial modifications into cellular RNA has generally relied upon a limited number of strategies (Figure 1). In order to broadly label RNA in a transcriptome wide manner, the modification must be presented in the form of a modified nucleoside triphosphate (NTP) that is a substrate for RNA polymerases. The most direct route for labeling is therefore to supply the cell with the modified NTP of interest; however, due to the limited cellular permeability of NTPs they cannot be directly taken up by cells and must be introduced using an artificial carrier or transporter method. These include transfection reagents[14, 15], which can be toxic, microinjection, which is limited to individual cells, or through the heterologous expression of nucleotide transporter proteins[16]. The latter approach has only been demonstrated in *E. coli* and it is unknown whether the strategy can be applied generally in different organisms. Further, the chemical synthesis of NTPs is cumbersome and generally low yielding.

Since the cellular uptake of NTPs is a major challenge, researchers have turned to supplying cells with simpler building blocks that can be metabolically activated to NTPs. This process relies on the nucleotide salvage pathway, a multi-enzyme cascade that is responsible for converting nucleotide precursors such as nucleobases, nucleosides, and nucleotide mono/di-phosphates into NTPs (Figure 1). Most often, the nucleotide precursor of choice is the nucleoside. Nucleosides are synthetically accessible and modifications can be incorporated on the nucleobase or on the ribose sugar. Selective incorporation into RNA or DNA can be biased by the usage of ribonucleosides or deoxyribonucleosides, respectively. In addition, while nucleosides are polar compounds and therefore not predicted to be particularly cell permeable, their uptake is aided by nucleoside transporters[17] that possess broad substrate tolerance for purine and pyrimidine nucleosides. Once in the cell, nucleosides must be phosphorylated three times by a combination of nucleoside and nucleotide kinases before they can be utilized by RNA polymerases. A related approach, designed to obviate the need for phosphorylation by nucleoside or nucleotide kinases, involves the delivery of protected nucleotide analogues in the form of phosphoramidate[18] or phosphate ester prodrugs[19] (Figure 1). Since the negative charge on the phosphate is masked, these compounds can enter cells through passive diffusion where they must then be converted into the native nucleotide by cellular esterases. One of the most common prodrug approaches is the “ProTide” strategy developed by McGuigan and co-workers[20] for delivering nucleoside monophosphates (NMPs) to cells (thereby bypassing the first cellular phosphorylation step). While ProTide compounds have been developed as antiviral drugs, the efficiency of the unmasking process is highly dependent on the expression of certain esterase enzymes and therefore the general application of this approach to RNA metabolic labeling in model cell lines has not been widely pursued.

Once a suitable modified nucleotide and labeling approach have been chosen, several requirements must be met. First, the modification needs to be compatible with RNA

polymerases. Typically, this precludes the use of modifications that grossly perturb Watson-Crick base pairing or that interfere with polymerization, for instance by modification of the 3' hydroxyl. Interestingly, W-C pairing is not a formal requirement for nucleotide polymerization as hydrophobic base pairs can be tolerated by DNA and RNA polymerases[21-23], although the general shape of the nucleobase and base pair must still be accommodated by the polymerase active site. There is considerably more flexibility on the non-Watson-Crick face of the nucleobase, and RNA polymerases have been shown to tolerate large modifications at the C5 position of pyrimidines[24]. Next, when feeding a modified nucleoside or other nucleotide precursor, the modified structure must be compatible with the nucleotide salvage pathway. Phosphorylation by nucleoside kinases to the NMP, in particular, has been shown to be particularly stringent, and is often the limiting step for the activation of modified nucleosides[25]. Another important consideration is the toxicity of the metabolic label. Toxicity can arise from several factors. For one, incorporation into RNA is often associated with cytotoxicity, presumably through interference with various steps of RNA processing, modification, translation or metabolism[26, 27]. Additionally, when performing metabolic labeling with a modified nucleoside, the nucleoside itself, or any of its metabolic nucleotide products can also interfere with cellular processes, typically through inhibition of enzymes involved in nucleotide metabolism[28]. While some amount of toxicity when using an artificial nucleotide is usually unavoidable, the degree to which this can be tolerated is determined by the particular experimental setup. Finally, the type of label must be chosen carefully based upon the goals of the experiment. In the following sections we will discuss the available modified nucleotides that can be used for RNA metabolic labeling experiments and the applications made possible by these structures.

3. RNA synthesis and turnover

Perhaps the most widespread application of RNA metabolic labeling is for the study of RNA synthesis and turnover dynamics (Figure 2). Since standard RNA-seq analysis provides only a

snapshot of the transcriptome frozen in time, there is a clear utility in probes that allow for the direct investigation of RNA metabolism, as these probes can provide temporal insight into transcriptional programs and enable the quantitation of the relative stability of RNA transcripts. Two types of experiments are typically performed to study RNA dynamics –"pulse" experiments can label newly synthesized transcripts at any particular time, whereas "pulse-chase" experiments enable the measurement of RNA half-life during the "chase". As with other biopolymers, the earliest implementations of RNA metabolic labeling utilized radioisotopes. This approach, often performed with ^{3}H-uridine (Figure 2)[29], benefits from efficient labeling stoichiometry and minimal perturbation to the native nucleotide structure. However, working with radioisotopes poses a number of technical and operational challenges and researchers have moved away from these reagents when possible. Moreover, radioisotopes are not compatible with RNA-seq workflows and are typically applied to study either the bulk population or individual transcripts.

A major advance in RNA metabolic labeling came in the use of 5-bromouridine (BrU) (Figure 2). This C5-modified nucleoside is readily available, recognized by the nucleotide salvage pathway, and minimally toxic. Detection or affinity isolation of BrU-containing RNAs can be accomplished using a BrU-specific monoclonal antibody[30]. With the development of Next-Generation Sequencing (NGS) technology, RNA labeling with BrU has been applied to measure RNA stability transcriptome-wide – a method that has been named bromouridine immunoprecipitation chase-deep sequencing analysis (BRIC-seq)[31].

While BrU-labeling has been widely used, the requirement for antibody detection makes this workflow reagent and labor intensive and also limits its immunostaining application to specimens that are easily antibody permeated. With innovations in biorthogonal labeling and nucleotide chemistry, further development in RNA metabolic labeling has focused on modified

nucleosides that are amenable to functionalization via bioorthogonal chemistry as well as those that allow for measurement of RNA dynamics without the need for affinity isolation. The first bioorthogonal ribonucleoside used for transcriptomic analysis, 5-ethynyluridine (EU) (Figure 2), was developed by Jao and Salic in 2008[32]. After incorporation, EU-labeled RNA can be detected in fixed cells by fluorescence microscopy using Cu-catalyzed azide-alkyne cycloaddition (CuACC) with an azide-conjugated fluorophore. Alternatively, CuACC labeling with azide-conjugated biotin can be performed on purified RNA in order to affinity isolate the EU-labeled RNA for RNA-seq analysis. Due to the lability of RNA, modified CuACC conditions must be employed in order to minimize RNA degradation[33]. The versatility of EU, simplicity of use, and its ability to broadly label the transcriptome has made it a popular metabolic label for RNA synthesis and turnover analysis. In addition to EU, several alkyne-containing ribonucleosides have been shown to be suitable for RNA metabolic labeling including 2-ethynyladenosine[34], N^6-propargyladenosine[35], and 5-ethynylcytidine[36] (Figure 2). The nucleotide metabolites of the alkyne-containing adenosine analogs are incorporated both transcriptionally by RNA polymerases and post-transcriptionally by poly(A)-polymerase enzymes enabling their use in the study of polyadenylation dynamics. 5-ethynylcytidine has similar labeling properties to 5-ethynyluridine but is metabolized more rapidly.

The rate enhancement afforded by Cu(I)-catalysis makes CuAAC chemistry well-suited for bioconjugation reactions in fixed cells as it is rapid and has low background. However, the use of CuAAC on purified RNA is challenging due to Cu(I)-mediated RNA degradation, and CuAAC is toxic to living cells. Therefore, it can be preferable to incorporate an azido group into RNA since it can be further derivatized using the Cu-free strain-promoted azide-alkyne cycyloaddition (SPAAC) reaction[37] which is significantly milder than CuAAC chemistry. In contrast to alkyne-modified ribonucleosides, which have been typically prepared by modification of the nucleobase, azido modifications have been primarily incorporated in place of the ribose 2'-

hydroxyl (Figure 2). Spitale and co-workers demonstrated the efficient metabolic incorporation of 2'-azidoadenosine into cellular RNA[38]. Similar to the alkyne-containing adenosine analogs, 2'-azidoadenosine is incorporated transcriptionally by RNA polymerase enzymes and post-transcriptionally by poly(A)polymerases. Additional azido-modified nucleosides have been used for RNA metabolic labeling through the manipulation of enzymes in the nucleotide salvage pathway. Our group showed that 5-azidomethyluridine can be incorporated into cellular RNA by expression of a mutated version of the nucleoside kinase UCK2 containing an expanded active site[25], and can be used for live cell imaging of azido-modified RNA after in-cell functionalization with a BODIPY-BCN fluorophore. Analogously, overexpression of wild-type UCK2 or the deoxynucleoside kinase dCK enables RNA labeling with 2'-azidouridine[39] or 2'-azidocytidine[40], respectively. These examples illustrate the critical importance of the nucleotide salvage pathway enzymes for RNA metabolic labeling, and also provide opportunities for tissue-specific metabolic labeling through targeted expression and engineering of nucleoside kinase enzymes.

While click chemistry and BrU-antibody detection provide powerful platforms to image RNA synthesis in cells, affinity enrichment of nascent RNA for RNA-seq analysis does not require derivatization of labeled RNA in the complex cellular environment. Instead, total cellular RNA can be first extracted and functionalization of labeled RNA with biotin or other reagent can be performed on purified RNA *in vitro*. This greatly expands the scope of accessible bioconjugation chemistry since native RNA contains few strong nucleophilic or electrophilic functional groups. One of the most versatile metabolic labels that provides an orthogonal handle for reactivity on purified RNA is the modified nucleoside 4-thiouridine (4-SU) (Figure 2). 4-thiouridine has been widely used in a number of transcriptomic studies to measure RNA synthesis and degradation across a large number of transcripts[41-50]. 4-thiouridine is readily accepted by the nucleotide salvage pathway and its triphosphate is installed in place of native uridine residues. In cells that express uracil phosphoribosyltransferase (UPRT) enzyme, labeling can also be performed using

the 4-thiouracil nucleobase[51, 52]. 4-SU-labeled RNA can be subjected to multiple modes of thiol-specific reactivity including alkylation, disulfide formation, and oxidative-nucleophilic aromatic substitution. These reactions can be used to distinguish newly synthesized 4-SU-containing RNA from unmodified RNA using primarily two approaches: 1) affinity isolation after appending biotin or other suitable affinity handle; and 2) mutational analysis of chemically recoded 4-SU residues. Due to the variability inherent in pulldown-based methods, approaches in the second category such as SLAM-seq[42], TimeLapse-seq[41], and TUC-seq[53], which do not require biochemical enrichment, offer considerable advantages. Affinity-based enrichment approaches offer higher sensitivity, which is important when metabolic labeling is performed for short periods of time. Chemical recoding of a modified base has also been reported for metabolically incorporated N^6-allyladenosine (Figure 2)[54], which can be cyclized in an iodine-dependent reaction to result in mutations upon reverse transcription. Metabolic incorporation of N^6-allyladenosine is much lower than that of pyrimidine nucleosides like 4-SU, which has precluded the widespread use of this derivative.

Instead of chemical recoding or affinity enrichment, a recent strategy for measuring RNA turnover relies upon direct RNA sequencing using nanopores (commercialized by Oxford Nanopore Technologies) to detect metabolic labeling events. In this approach, known as Nano-ID[55], the modified nucleoside is detected directly in RNA through perturbations in the nanopore current. Appropriate selection of nucleoside is critical since compatibility with nanopore sequencing and detection efficiency is unknown for most modified nucleosides. Maier *et al.*[55] demonstrated that 5-EU can be reliably detected, whereas other structures failed to produce distinctive signatures or resulted in truncated nanopore reads. This approach has several advantages in that long-read nanopore sequencing can discriminate between RNA isoforms and direct label detection does not require chemical treatment or biochemical enrichment of RNA prior to sequencing. However, currently nanopore sequencing lacks the throughput provided by

short-read NGS platforms limiting the number of RNA transcripts that can be studied in a single experiment.

4. RNA binding proteins and modifying enzymes

In addition to studying transcriptome dynamics through RNA sequencing, another major application of RNA metabolic labeling is to investigate biochemical interactions between RNA and associated RNA-binding proteins that occur within the cell (Figure 3). RNA-protein interactions can be mapped transcriptome-wide using crosslinking and immunoprecipitation sequencing combined with high-throughput sequencing (CLIP-seq or HITS-CLIP)[56, 57]. These studies are critical for characterizing the RNA clients of an individual RNA binding protein, which typically reveal insight into its biological role. While RNA-protein interactions can be covalently crosslinked in situ by exposing cells to short wavelength UV (~254 nm), the efficiency of crosslinking is low and short wavelength UV irradiation can be damaging to biomolecules. An improvement over conventional UV crosslinking with native nucleobases can be achieved by the incorporation of photoreactive nucleosides into RNA. A number of photoreactive nucleotide have been studied in vitro, but the most widely used nucleotide for in-cell crosslinking is 4-thiouridine (4-SU) (Figure 3). This is due to its high photoreactivity and similarity in size to native uridine. Tuschl and co-workers explored RNA labeling and protein-RNA crosslinking with different modified nucleosides including 4-SU, 6-thioguanosine (6-SG), and 5-iodouridine (5-IU) in the development of the PAR-CLIP[58] approach that relies upon metabolic labeling prior to crosslinking, immunoprecipitation and sequencing; they ultimately settled on 4-SU as a superior photoreactive nucleoside. In addition to affording higher yield of RNA-protein crosslinking, bonafide 4-SU-mediated crosslinks can be more confidently distinguished from non-specific background contamination and mapped at near nucleotide resolution by relying upon characteristic T to C mutations occurring at the site of RNA-protein binding. Another application

of 4-SU-induced protein-RNA crosslinking is to characterize the complement of RNA binding proteins in a particular sample. This method is termed “RNA interactome capture”[59, 60] and involves RNA-protein enrichment through denaturing antisense oligo-dT pulldown and mass spectrometry-based proteomics after metabolic labeling and in-cell crosslinking with 4-SU. Related approaches for studying RNA-binding proteins relying upon 4-SU-mediated protein-RNA crosslinking have also been described including RBR-ID[61], TRAPP[62], CARIC[63], XRNAX[64], and OOPS[65].

While UV-crosslinking provides a general method to stabilize diverse RBP-RNA interactions, more specialized metabolic labeling approaches have been developed to study RNA modifying enzymes and identify their associated RNA modification sites at single-nucleotide resolution. In particular, while UV crosslinking can capture binding events between RNA modifying enzymes and RNA transcripts, it cannot distinguish between catalytic and non-catalytic RNA-substrate interactions. Further, UV crosslinking can happen at a site distal to the modification site, which limits the resolution of modification mapping. In order to address these limitations, a mechanism-based nucleoside probe that is specific to the particular RNA modifying enzyme family of interest can be used as a metabolic label. When incorporation of the modified nucleoside overlaps with the substrate site of an RNA modifying enzyme, and assuming that the modified nucleoside does not perturb enzyme-substrate recognition, a covalent crosslink may be formed. This enables the profiling of RNA modification sites transcriptome-wide through CLIP-seq based methodology. While such an approach is appealing in concept, it is limited to RNA modifying enzymes for which suitable mechanism-based nucleoside probes have been identified. One such example is RNA 5-methylcytidine (m^5C) methyltransferases[66]. In mammals these enzymes fall into the NSUN/DNMT family of proteins, and act upon diverse RNA substrates. All known RNA m^5C MTases rely upon a similar thymidylate-synthase-like catalytic mechanism involving the formation of a covalent enzyme-substrate intermediate through an

essential cysteine residue in the protein. Typically, a second cysteine residue catalyzes beta-elimination and enzyme release after methylation by SAM or other suitable methyl donor. This mechanism lends itself to covalent inhibition since modification of the cytosine C5 position in a manner that prevents beta-elimination can result in a stable RNA-protein adduct. Cairns and Khoddami applied this chemistry in their "Aza-IP"[67] approach to profile the substrates of the m^5C methyltransferases NSUN2 and DNMT2 through metabolic labeling of RNA with 5-azacytidine (5-AzaC) feeding (Figure 3). More recently, 5-fluorouracil (Figure 3) was utilized to profile the substrates of the 5-methyluridine (m^5U) methyltransferase TRMT2A[68], which forms an analogous enzyme-RNA substrate intermediate to that of the m^5C RNA methyltransferases. These strategies lay the groundwork for applying analogous mechanism-based crosslinking approaches to other classes of RNA modifying enzymes, however relatively few RNA modifying enzymes are known to form covalent intermediates with their substrates, therefore more general, reactivity-based crosslinking chemistry may be required for these efforts. Our group has recently developed a reactivity-based method to profile RNA modifying enzyme activity in native samples that we named RNA-mediated activity-based protein profiling (RNABPP)[69]. RNABPP combines metabolic labeling with modified nucleoside probes, RNA-protein enrichment and quantitative mass spectrometry-based proteomics in order to identify RNA modifying enzymes that form covalent adducts with a candidate modified nucleoside. This strategy not only reports on enzyme activity and catalytic mechanism in cells, but also provides nucleoside probes for characterizing RNA modifying enzyme substrates using catalytic CLIP methods. In our first demonstration of this approach, we profiled RNA modifying enzymes that react with 5-fluoropyrimidines (Figure 3) showing that the uncharacterized human dihydrouridine synthase enzyme DUS3L crosslinks with 5-fluorouridine-containing RNA. We further employed catalytic crosslinking and sequencing to profile the substrates of DUS3L for the first time. Adaptation of our approach to other modified ribonucleosides compatible with metabolic labeling will allow further characterization of the RNA modifying proteome.

While lying outside the scope of this review, which is focused on metabolic labeling with ribonucleosides, it is worth noting that studies of RNA modifications (in particular, m^6A) and RNA modifying enzymes have also relied upon modified SAM analogs (generated metabolically by feeding homocysteine derivatives) to install modifications in cellular RNA in an enzyme-dependent fashion[70, 71].

5. Discussion

The incorporation of chemically modified ribonucleosides into cellular RNA has opened up a number of new avenues for transcriptomic study. As described in this review, when combined with Next Generation Sequencing, mass-spectrometry based proteomics, or fluorescence imaging, these nucleoside probes enable the analysis of RNA transcript dynamics, modification status, and associated proteins in cells and in whole organisms. Given the availability of these platforms and of many modified nucleosides, we anticipate that the adoption of RNA metabolic labeling approaches will become increasingly more commonplace among the RNA biology community.

Moving forward, there is a clear need to expand the rather limited repertoire of modified ribonucleosides that can be used for metabolic labeling in order to further our ability to probe RNA behavior in biological systems. Here, we propose several possible directions for further development. First, methods for live cell imaging of RNA are primarily limited to the investigation of individual transcripts using artificial reporter tags. This can be accomplished by expressing a fusion of the MS2 hairpin[72] or a fluorescent RNA aptamer (e.g. Spinach, Mango, etc.)[73] to the RNA transcript of interest in cells, or synthesizing a fluorescently labeled RNA *in vitro* followed by cellular transfection[74]. In order to image bulk RNA synthesis or turnover, SPAAC labeling of a metabolically incorporated azido-nucleosides has been used[25, 38], however, the reaction kinetics

and high background leave room for optimization, and extensive “wash-out” procedures must be used to eliminate background originating from the free, unconjugated fluorophore. Therefore, the development of novel bioorthogonal reactions with faster kinetics and “turn-on” fluorescence together with the appropriate modified nucleosides have the potential to advance research in this area. One such example is the tetrazine ligation[75], and vinylpyrimidine derivatives have been shown to be compatible with metabolic labeling[76], however live cell labeling and RNA imaging using this strategy has yet to be reported. A complementary approach to live cell RNA imaging would involve the application of fluorescent nucleosides that can be directly incorporated during cellular RNA synthesis. While a number of fluorescent nucleoside-based compounds have been reported and applied *in vitro*[77], currently it is unknown whether any of these structures would be compatible with the nucleotide salvage pathway and suitable for metabolic labeling. Further many of the nucleoside-based fluorophores have spectral properties that are not ideally suited for cellular imaging. Nevertheless, this type of approach would obviate the need for post-transcriptional fluorophore conjugation, which is often a major source of background fluorescence. In addition, the incorporation of fluorescent nucleosides directly into the transcript would also open opportunities beyond studying RNA localization, such as quantifying RNA structure and microenvironment.

Another emerging strategy in RNA biology is the crosslinking of RNA modifying enzymes and their substrates with mechanism-based probes for identifying RNA modification sites using Next Generation Sequencing. These chemical crosslinking strategies have relied upon metabolic incorporated nucleosides and nucleobases, but currently few such examples have been reported. Given the broad interest in RNA modifications and their regulatory role, as well as the large array of post-transcriptional RNA modifications and corresponding writer enzymes, identification of novel activity-based probes for RNA modification writer enzymes should be

prioritized. These may also serve as starting points for inhibitors and biophysical probes of these enzymes.

While bulk RNA incorporation is often the goal of metabolic labeling experiments, targeted labeling can also be of great utility. Enzymes involved in nucleoside metabolism can be tethered to certain subcellular compartments to bias the generation of modified NTPs in the area of interest[78]. In addition, some modified nucleosides show specificity for certain RNA polymerases or template-independent polymerase enzymes[40], although the basis of this behavior is not well understood. Further exploration of these ideas will provide more robust approaches for transcript and polymerase-specific metabolic labeling.

Finally, the viability of RNA metabolic labeling strategies and expansion of these approaches depends on robust and general methods to introduce probes into cellular RNA. Currently, our understanding of delivery and incorporation strategies for modified nucleotide probes is largely anecdotal. Moving forward, a more systematic analysis of each individual approach will be beneficial to the field and will hopefully lead to a unified toolbox for metabolic labeling integrating nucleoside chemistry, protein engineering, and nucleotide pro-drug strategies.

Acknowledgements

Research in the Kleiner lab is supported by the NIH (R01 GM132189) and the NSF (MCB-1942565).

Conflict of interest

No competing financial interests have been declared.

Figure Legends

Figure 1. Methods for the metabolic incorporation of modified nucleotides.

Figure 2. Modified ribonucleosides for studying RNA synthesis and degradation.

Figure 3. Modified ribonucleosides for studying RNA-associated binding proteins and enzymes.

TOC legend. This review summarizes recent developments in metabolic labeling of RNA to study RNA synthesis and turnover, RNA binding proteins, and RNA modifications and modifying enzymes.

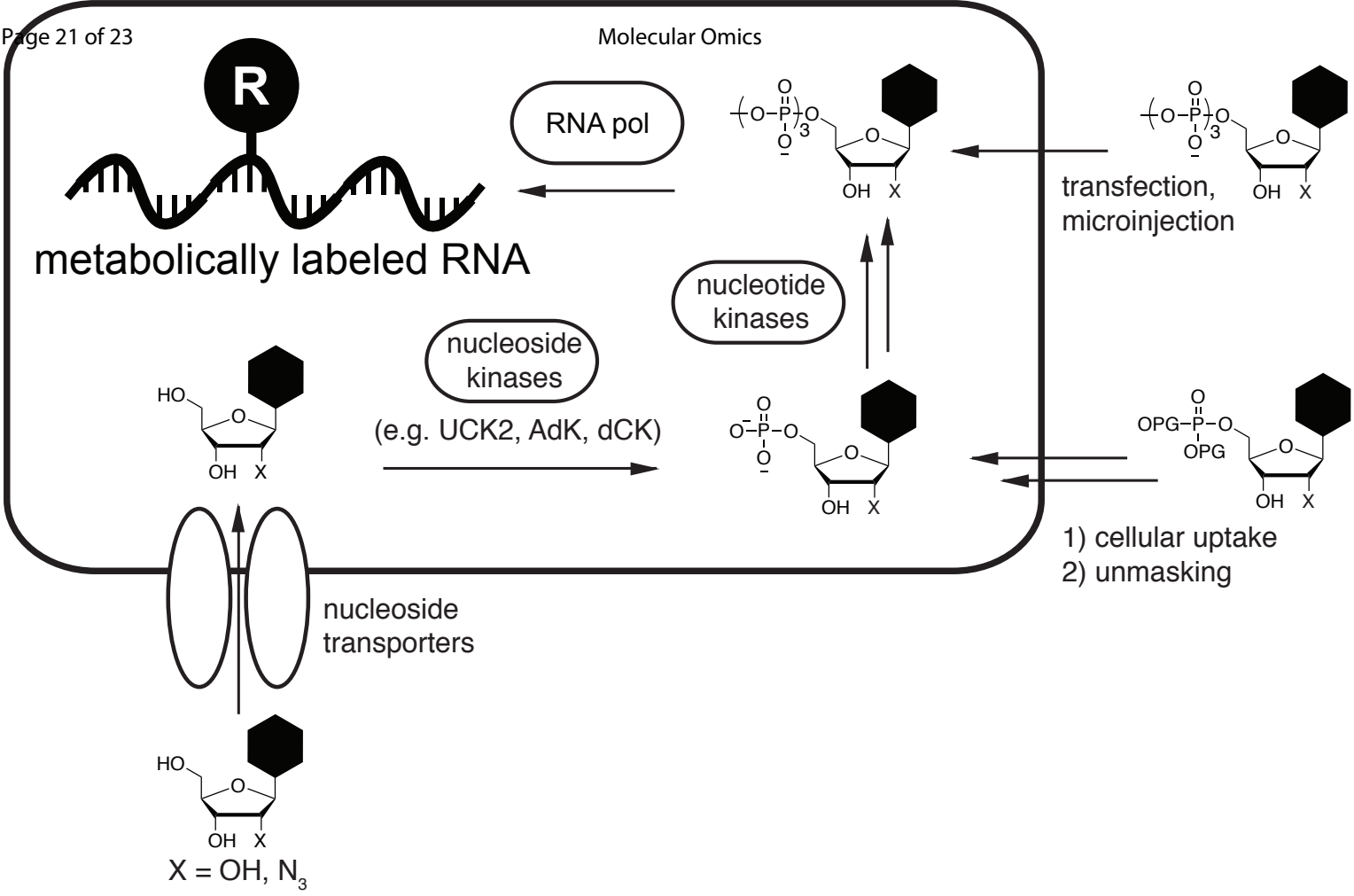
References

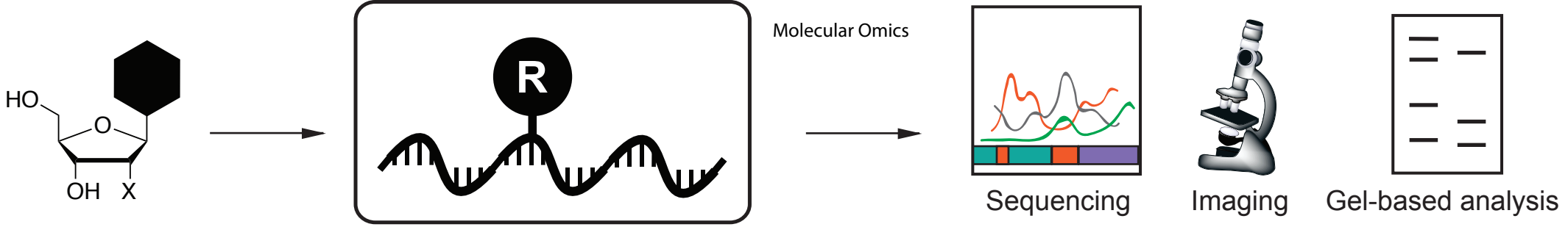
1. Licatalosi, D.D. & Darnell, R.B. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* **11**, 75-87 (2010).
2. Sonenberg, N. & Hinnebusch, A.G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745 (2009).
3. Sharp, P.A. The centrality of RNA. *Cell* **136**, 577-580 (2009).
4. Martin, K.C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719-730 (2009).
5. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763-776 (2009).
6. Rodriguez, A.J., Czaplinski, K., Condeelis, J.S. & Singer, R.H. Mechanisms and cellular roles of local protein synthesis in mammalian cells. *Curr Opin Cell Biol* **20**, 144-149 (2008).
7. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499-506 (2002).
8. Beelman, C.A. & Parker, R. Degradation of mRNA in eukaryotes. *Cell* **81**, 179-183 (1995).
9. Hughes, T.R. et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**, 342-347 (2001).
10. Shalon, D., Smith, S.J. & Brown, P.O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6**, 639-645 (1996).
11. Pease, A.C. et al. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci U S A* **91**, 5022-5026 (1994).
12. Goodwin, S., McPherson, J.D. & McCombie, W.R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333-351 (2016).
13. Chu, Y. & Corey, D.R. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther* **22**, 271-274 (2012).
14. Haukenes, G., Szilvay, A.M., Brokstad, K.A., Kanestrom, A. & Kalland, K.H. Labeling of RNA transcripts of eukaryotic cells in culture with BrUTP using a liposome transfection reagent (DOTAP). *Biotechniques* **22**, 308-312 (1997).

15. Zawada, Z., Tatar, A., Mocilac, P., Budesinsky, M. & Kraus, T. Transport of Nucleoside Triphosphates into Cells by Artificial Molecular Transporters. *Angew Chem Int Ed Engl* **57**, 9891-9895 (2018).
16. Feldman, A.W. et al. A Tool for the Import of Natural and Unnatural Nucleoside Triphosphates into Bacteria. *Journal of the American Chemical Society* **140**, 1447-1454 (2018).
17. Pastor-Anglada, M. & Perez-Torras, S. Emerging Roles of Nucleoside Transporters. *Front Pharmacol* **9**, 606 (2018).
18. McGuigan, C., Pathirana, R.N., Mahmood, N., Devine, K.G. & Hay, A.J. Aryl phosphate derivatives of AZT retain activity against HIV1 in cell lines which are resistant to the action of AZT. *Antiviral Res* **17**, 311-321 (1992).
19. Gollnest, T., de Oliveira, T.D., Schols, D., Balzarini, J. & Meier, C. Lipophilic prodrugs of nucleoside triphosphates as biochemical probes and potential antivirals. *Nat Commun* **6**, 8716 (2015).
20. Mehellou, Y., Rattan, H.S. & Balzarini, J. The ProTide Prodrug Technology: From the Concept to the Clinic. *J Med Chem* **61**, 2211-2226 (2018).
21. Schweitzer, B.A. & Kool, E.T. Hydrophobic, Non-Hydrogen-Bonding Bases and Base Pairs in DNA. *J Am Chem Soc* **117**, 1863-1872 (1995).
22. Hirao, I. et al. An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nat Methods* **3**, 729-735 (2006).
23. Zhang, Y. et al. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551**, 644-647 (2017).
24. Tarasow, T.M., Tarasow, S.L. & Eaton, B.E. RNA-catalysed carbon-carbon bond formation. *Nature* **389**, 54-57 (1997).
25. Zhang, Y. & Kleiner, R.E. A Metabolic Engineering Approach to Incorporate Modified Pyrimidine Nucleosides into Cellular RNA. *J Am Chem Soc* **141**, 3347-3351 (2019).
26. Silverstein, R.A., Gonzalez de Valdivia, E. & Visa, N. The incorporation of 5-fluorouracil into RNA affects the ribonucleolytic activity of the exosome subunit Rrp6. *Molecular cancer research : MCR* **9**, 332-340 (2011).
27. Burger, K. et al. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol* **10**, 1623-1630 (2013).
28. Pinedo, H.M. & Peters, G.F. Fluorouracil: biochemistry and pharmacology. *J Clin Oncol* **6**, 1653-1664 (1988).
29. Harwood, R. & Itzhaki, S. The use of (5-3H)uridine for pulse-labelling ribonucleic acid of mouse Ehrlich ascites-tumour cells. *Biochem J* **128**, 91P-92P (1972).
30. Halicka, H.D., Bedner, E. & Darzynkiewicz, Z. Segregation of RNA and separate packaging of DNA and RNA in apoptotic bodies during apoptosis. *Exp Cell Res* **260**, 248-256 (2000).
31. Tani, H. et al. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res* **22**, 947-956 (2012).
32. Jao, C.Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc Natl Acad Sci U S A* **105**, 15779-15784 (2008).
33. Paredes, E. & Das, S.R. Click chemistry for rapid labeling and ligation of RNA. *Chembiochem* **12**, 125-131 (2011).
34. Curanovic, D. et al. Global profiling of stimulus-induced polyadenylation in cells using a poly(A) trap. *Nat Chem Biol* **9**, 671-673 (2013).
35. Grammel, M., Hang, H. & Conrad, N.K. Chemical reporters for monitoring RNA synthesis and poly(A) tail dynamics. *Chembiochem* **13**, 1112-1115 (2012).
36. Qu, D. et al. 5-Ethynylcytidine as a new agent for detecting RNA synthesis in live cells by "click" chemistry. *Anal Biochem* **434**, 128-135 (2013).

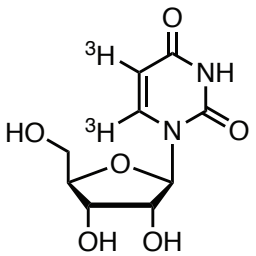
37. Baskin, J.M. et al. Copper-free click chemistry for dynamic in vivo imaging. *Proc Natl Acad Sci U S A* **104**, 16793-16797 (2007).
38. Nainar, S. et al. Metabolic Incorporation of Azide Functionality into Cellular RNA. *Chembiochem* **17**, 2149-2152 (2016).
39. Nainar, S. et al. An optimized chemical-genetic method for cell-specific metabolic labeling of RNA. *Nat Methods* **17**, 311-318 (2020).
40. Wang, D., Zhang, Y. & Kleiner, R.E. Cell- and Polymerase-Selective Metabolic Labeling of Cellular RNA with 2'-Azidocytidine. *J Am Chem Soc* **142**, 14417-14421 (2020).
41. Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C. & Simon, M.D. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods* **15**, 221-225 (2018).
42. Herzog, V.A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* **14**, 1198-1204 (2017).
43. Fuchs, G. et al. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol* **15**, R69 (2014).
44. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225-1228 (2016).
45. Dolken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959-1972 (2008).
46. Jurges, C., Dolken, L. & Erhard, F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* **34**, i218-i226 (2018).
47. Kenzelmann, M. et al. Microarray analysis of newly synthesized RNA in cells and animals. *Proc Natl Acad Sci U S A* **104**, 6164-6169 (2007).
48. Muhar, M. et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* **360**, 800-805 (2018).
49. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29**, 436-442 (2011).
50. Rabani, M. et al. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698-1710 (2014).
51. Miller, M.R., Robinson, K.J., Cleary, M.D. & Doe, C.Q. TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nat Methods* **6**, 439-441 (2009).
52. Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R. & Boothroyd, J.C. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol* **23**, 232-237 (2005).
53. Lusser, A. et al. Thiouridine-to-Cytidine Conversion Sequencing (TUC-Seq) to Measure mRNA Transcription and Degradation Rates. *Methods Mol Biol* **2062**, 191-211 (2020).
54. Shu, X. et al. N(6)-Allyladenosine: A New Small Molecule for RNA Labeling Identified by Mutation Assay. *J Am Chem Soc* **139**, 17213-17216 (2017).
55. Maier, K.C., Gressel, S., Cramer, P. & Schwalb, B. Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res* **30**, 1332-1344 (2020).
56. Lee, F.C.Y. & Ule, J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular cell* **69**, 354-369 (2018).
57. Nechay, M. & Kleiner, R.E. High-throughput approaches to profile RNA-protein interactions. *Curr Opin Chem Biol* **54**, 37-44 (2020).
58. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141 (2010).
59. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-1406 (2012).
60. Baltz, A.G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular cell* **46**, 674-690 (2012).

61. He, C. et al. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Molecular cell* **64**, 416-430 (2016).
62. Shchepachev, V. et al. Defining the RNA interactome by total RNA-associated protein purification. *Mol Syst Biol* **15**, e8689 (2019).
63. Huang, R., Han, M., Meng, L. & Chen, X. Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proc Natl Acad Sci U S A* **115**, E3879-E3887 (2018).
64. Trendel, J. et al. The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391-403 e319 (2019).
65. Queiroz, R.M.L. et al. Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat Biotechnol* **37**, 169-178 (2019).
66. Bohnsack, K.E., Hobartner, C. & Bohnsack, M.T. Eukaryotic 5-methylcytosine (m(5)C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease. *Genes (Basel)* **10** (2019).
67. Khoddami, V. & Cairns, B.R. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat Biotechnol* **31**, 458-464 (2013).
68. Carter, J.M. et al. FICC-Seq: a method for enzyme-specified profiling of methyl-5-uridine in cellular RNA. *Nucleic Acids Res* **47**, e113 (2019).
69. Dai, W. et al. Activity-based RNA-modifying enzyme probing reveals DUS3L-mediated dihydrouridylation. *Nat Chem Biol* (2021).
70. Shu, X. et al. A metabolic labeling method detects m(6)A transcriptome-wide at single base resolution. *Nat Chem Biol* **16**, 887-895 (2020).
71. Hartstock, K. et al. Enzymatic or In Vivo Installation of Propargyl Groups in Combination with Click Chemistry for the Enrichment and Detection of Methyltransferase Target Sites in RNA. *Angew Chem Int Ed Engl* **57**, 6342-6346 (2018).
72. Bertrand, E. et al. Localization of ASH1 mRNA particles in living yeast. *Molecular cell* **2**, 437-445 (1998).
73. Paige, J.S., Wu, K.Y. & Jaffrey, S.R. RNA mimics of green fluorescent protein. *Science* **333**, 642-646 (2011).
74. Baladi, T. et al. Stealth Fluorescence Labeling for Live Microscopy Imaging of mRNA Delivery. *J Am Chem Soc* **143**, 5413-5424 (2021).
75. Blackman, M.L., Royzen, M. & Fox, J.M. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J Am Chem Soc* **130**, 13518-13519 (2008).
76. Kubota, M. et al. Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels-Alder Chemistry. *ACS Chem Biol* **14**, 1698-1707 (2019).
77. Xu, W., Chan, K.M. & Kool, E.T. Fluorescent nucleobases as tools for studying DNA and RNA. *Nat Chem* **9**, 1043-1055 (2017).
78. Nguyen, K. et al. Spatially Restricting Bioorthogonal Nucleoside Biosynthesis Enables Selective Metabolic Labeling of the Mitochondrial Transcriptome. *ACS Chem Biol* **13**, 1474-1479 (2018).


R
metabolically labeled RNA
RNA pol
transfection,
microinjection
nucleotide
kinases
nucleoside
kinases
(e.g. UCK2, AdK, dCK)
1) cellular uptake
2) unmasking
nucleoside
transporters
X = OH, N3

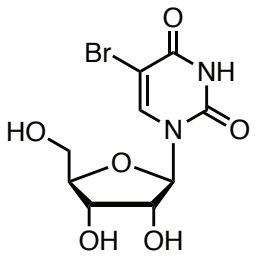


Nucleosides for autoradiography



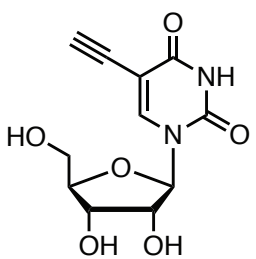
^{3}H-uridine
(^{3}H-U)

Nucleosides for antibody enrichment

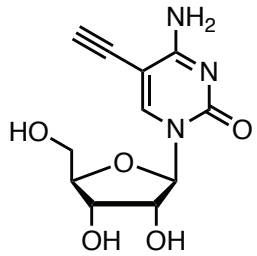


5-bromouridine
(BrU)

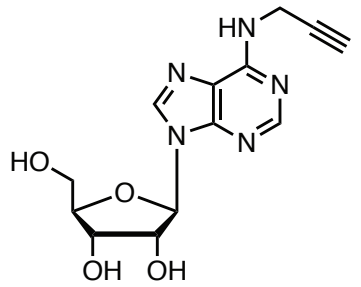
Alkyne-containing nucleosides (CuAAC labeling)



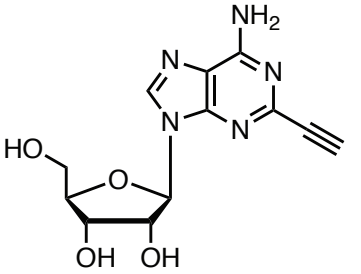
5-ethynyluridine
(EU)



5-ethynylcytidine
(EC)

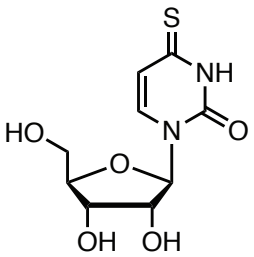


N^{6}-propargyladenosine
(N^{6}pA)

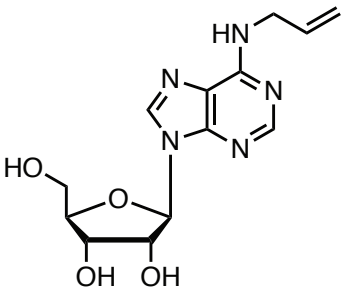


2-ethynyladenosine
(2-EA)

Nucleosides for chemical sequencing

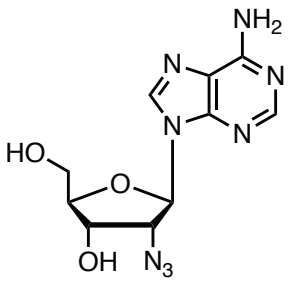


4-thiouridine
(4-SU)

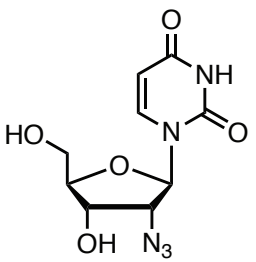


N^{6}-allyladenosine
(a^{6}A)

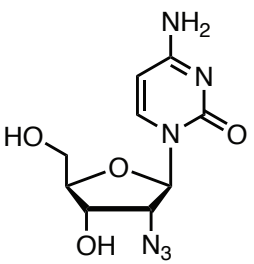
Azide-containing nucleosides (CuAAC and SPAAC labeling)



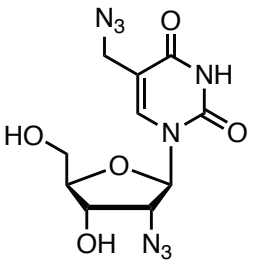
2'-azidoadenosine
(2'-AzAd)



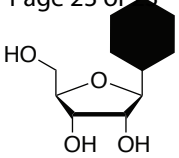
2'-azidouridine
(2'-AzUrd)



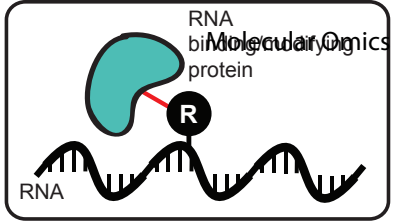
2'-azidocytidine
(2'-AzCyd)



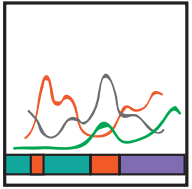
5-azidomethyluridine
(5-AmU)



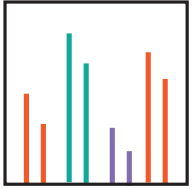
hυ
or
chemical
crosslinking



RNA-protein
enrichment

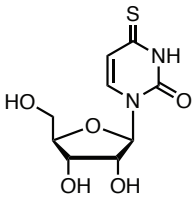


RNA-seq
(protein IP)

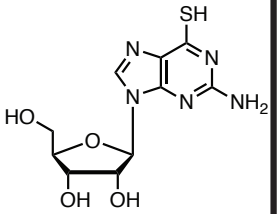


proteomics
(RNA pulldown)

photocrosslinking

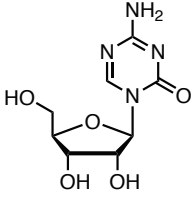


4-thiouridine
(4-SU)

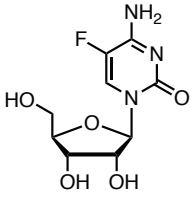


6-thioguanosine
(6-SG)

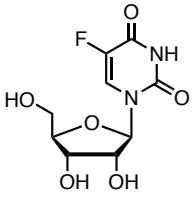
mechanism-based crosslinking



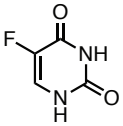
5-azacytidine
(5-AzaC)



5-fluorocytidine
(5-FCyd)



5-fluorouridine
(5-FUrd)



5-fluorouracil
(FU)