Published in final edited form as:

Trends Biochem Sci. 2014 June ; 39(6): 255–257. doi:10.1016/j.tibs.2014.04.002.

Sizing up the poly(A) tail: insights from deep sequencing

Dinghai Zheng and Bin Tian*

Department of Biochemistry and Molecular Biology, Rutgers New Jersey Medical School, Newark, NJ 07103

Abstract

Global investigation of poly(A) tails has been hindered by technical challenges. In a recent advance, two groups developed deep sequencing methods to globally interrogate poly(A) tail length and sequence with high precision, opening new avenues for investigation of poly(A) tail functions in mRNA metabolism. Initial applications of these methods reveal insights into the relationship between poly(A) tail length and translational efficiency, and identify widespread uridylation and guanylation at the 3' ends of transcripts.

Keywords

poly(A) tail; translation; mRNA stability; post-transcriptional control; polyadenylation

It is well established that the poly(A) tail, which decorates the 3' end of most mature mRNAs in eukaryotes, plays important roles in mRNA stability and translation [1]. Regulation of poly(A) tail length impacts gene expression in processes such as early development, inflammation, learning and memory [2]. Although the poly(A) tail lengths of individual genes can be accurately measured using biochemical methods, genome-wide analyses have in the past yielded results with limited resolution [3, 4]. The advent of deep sequencing technologies brings a bonanza to RNA research. However, the difficulty in sequencing homopolymers (strings of identical nucleotides) [5] has hampered precise measurement of poly(A) tail length by deep sequencing. To work around this problem, Subtelny *et al.* and Chang *et al.* have recently developed two methods, named poly(A) tail length profiling by sequencing (PAL-seq) [6] and TAIL-seq [7], respectively. These two methods, both designed for the Illumina® platform, involve largely similar strategies to prepare cDNA libraries. The major difference lies at the sequencing step (summarized in Figure 1A).

Using these methods, Chang *et al.* examined the poly(A) tail lengths in HeLa and NIH3T3 cells, while Subtelny *et al.* analyzed cells or tissues from a variety of species, including yeast, plant, fly, zebrafish, xenopus, mouse and human. In line with previous reports, both

© 2014 Elsevier Ltd. All rights reserved.

*Corresponding author btian@rutgers.edu FAX: (973) 972-5594, TEL: (973) 972-3615.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

studies found that different genes exhibit widely different poly(A) tail lengths, with genes in certain Gene Ontology (GO) groups tending to have shorter or longer poly(A) tails than genes in other groups. However, in contrast to prevailing perceptions [1], poly(A) tail lengths in the steady state appear to be quite short: ~30 nucleotides (nt) in yeast [6] and ~50 to 100 nt in most metazoan samples [6, 7]. Subtelny *et al.* additionally found a gradual lengthening of the poly(A) tail (from ~20 nt to ~50–60 nt) in early embryonic development of zebrafish and xenopus. It is notable, however, that the two studies have some discrepancies in reported poly(A) tail lengths for genes expressed in the same cell types, with the mean poly(A) lengths reported by Subtelny *et al.* being 20–40 nt longer than those by Chang *et al.* Some Gene Ontology (GO) analysis results also differ. For example, genes in the “ribosomal subunit” group were found to express transcripts with short poly(A) tails in NIH3T3 and HeLa cells by Subtelny *et al.*, but not so by Chang *et al.* It is not clear whether cell conditions or technical reasons are behind these differences. Both studies found that alternative polyadenylation isoforms can have different poly(A) tail lengths; therefore, differences in defining gene 3' ends could also cause some of the inconsistencies.

Although poly(A) tail shortening, also termed deadenylation, is the first step for the degradation of most mRNAs [8], neither study found connections between the poly(A) tail length and mRNA abundance. However, Chang *et al.* did observe a correlation between poly(A) tail length and mRNA half-life for relatively long mRNAs whose decay rates could be more reliably determined. Using TAIL-seq, the group also found widespread terminal uridylation (addition of uridine) and guanylation (addition of guanosine) of the poly(A) tail. Importantly, uridylation and guanylation frequencies have negative and positive correlations with mRNA half-lives, respectively, suggesting that these modifications are related to mRNA decay (Figure 1B). Although the fractions of mRNAs containing these modifications appear to be small (most genes have uridylation and guanylation in less than 10% and 5% of their transcripts, respectively), uridylation is more frequently associated with short poly(A) tails (< 25 nt) and guanylation is slightly biased to long poly(A) tails (>40 nt). Further studies are needed to determine whether it is the poly(A) tail length *per se* or terminal modification that determines or influences mRNA stability. Notably, poly(A) tail length was not found to correlate with mRNA half-life by Subtelny *et al.* Whether this discrepancy is related to uridylation or guanylation of transcripts remains to be examined.

The biggest surprise from these two studies perhaps is the finding that poly(A) tail length in most samples is not coupled with translational efficiency, as indicated by the lack of such correlation across different genes [6, 7], as well as across transcripts from the same gene [6]. Notable exceptions are samples from early developmental stages prior to or around the maternal-to-zygotic transition in zebrafish and xenopus, for which a clear correlation between poly(A) tail length and translational efficiency exists [6]. Therefore, Subtelny *et al.* proposed that there are two regulatory regimes concerning the impact of poly(A) tail length on translation (Figure 1B). Explaining previous observations regarding microRNA (miRNA) regulation in the zebrafish embryo [9], Subtelny *et al.* found that miRNAs elicit deadenylation and translational inhibition at the developmental stages when poly(A) tail length and translational efficiency are coupled (2 and 4 hours post fertilization, hpf), but induce mRNA degradation at a later stage (6 hpf). Together with previous studies showing

that miRNAs in most cells function primarily by destabilizing their target mRNAs [10], this work indicates that the consequence of mRNA deadenylation elicited by miRNA depends on which regulatory regime is in play in a given cellular context (Figure 1B). Further studies are needed to elucidate the mechanistic details that differentiate the two regulatory regimes. How these mechanisms are exploited in other biological systems to mediate miRNA functions is also an open question.

In sum, despite some discrepancies that need to be reconciled by further experimentation, the studies by Subtelny *et al.* and Chang *et al.* present the first high resolution, global views of poly(A) tail length and its relevance to mRNA metabolism. Their methods provide powerful means to further unravel the functions of the poly(A) tail in gene regulation under physiological and pathological conditions.

ACKNOWLEDGMENTS

We thank Michael B. Mathews for helpful discussions. This work was supported by a grant from the National Institutes of Health (GM084089).

REFERENCES

1. Sachs A, Wahle E. Poly(A) tail metabolism and function in eucaryotes. *J Biol Chem.* 1993; 268:22955–22958. [PubMed: 8226806]
2. Weill L, et al. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat Struct Mol Biol.* 2012; 19:577–585. [PubMed: 22664985]
3. Meijer HA, et al. A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. *Nucleic acids research.* 2007; 35:e132. [PubMed: 17933768]
4. Beilharz TH, Preiss T. Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *Rna.* 2007; 13:982–997. [PubMed: 17586758]
5. Quail MA, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 2012; 13:341. [PubMed: 22827831]
6. Subtelny AO, et al. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature.* 2014
7. Chang H, et al. TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3' End Modifications. *Mol Cell.* 2014
8. Chen CY, Shyu AB. Mechanisms of deadenylation-dependent decay. *Wiley interdisciplinary reviews. RNA.* 2011; 2:167–183. [PubMed: 21957004]
9. Bazzini AA, et al. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science.* 2012; 336:233–237. [PubMed: 22422859]
10. Guo H, et al. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature.* 2010; 466:835–840. [PubMed: 20703300]

Figure 1.

(a) Flowchart showing key steps of PAL-seq and TAIL-seq at the sequencing stage. In PAL-seq (left), the poly(A) tail region of a cDNA library for 3' ends of transcripts is filled with a mixture of dTTP and biotinylated dUTP before regular sequencing of the region immediately upstream of the poly(A) tail (indicated by an arrow below the cDNA; 36 bases in the study). Poly(A) tail length is calculated based on the amount of incorporated biotinylated dUTP as measured by fluorescent streptavidin (which binds biotin). With spike-in RNAs containing defined poly(A) tail lengths as size markers, the poly(A) tail length for

each sequenced transcript is calculated. In TAIL-seq (right), the cDNA library is sequenced using a paired-end sequencing protocol. Read 1 (shown as an arrow above the cDNA; 51 bases in the study) is used for transcript identification, while Read 2 (shown as an arrow below the cDNA; 251 bases in the study) covers the poly(A) tail region. The boundary between the poly(A) tail and its upstream region, which is normally hard to define due to problems associated with homopolymer sequencing, is resolved by analysis of the fluorescence signals directly from the sequencing cycles using a machine learning method (Gaussian Mixture Hidden Markov Model was used in the study). Note that although these two methods involve largely similar strategies to prepare cDNA libraries, such as partial digestion of RNA with RNase T1 and 5' adapter ligation, there are some notable differences (not shown in the figure): TAIL-seq uses a single-stranded 3' adapter that ligates to any sequences; by contrast, PAL-seq uses splint ligation with a 3' adapter that hybridizes to terminal poly(A) sequences. As such, PAL-seq captures only the 3' ends with pure A-stretches whereas TAILseq can detect modified 3' ends as well. In addition, PAL-seq cDNA libraries are not amplified by PCR. Blue and red boxes are 5' and 3' adapter regions, respectively, and the grey box represents the region upstream of the poly(A) tail. AAA... AAA, poly(A) tail; green arrow, sequencing direction; "?", uncertain base; "V", A/C/G; "N", A/T/C/G. **(B)** The two gene regulatory regimes proposed by Subtelny *et al.* that govern the consequence of miRNA-mediated deadenylation of target mRNA. In early embryos where poly(A) tail length is coupled with translational efficiency, miRNAs elicit translational inhibition but not mRNA degradation; in most other cells, where poly(A) tail length is not coupled with translational efficiency, miRNAs induce mRNA decay without translational repression. AAA...AAA, original poly(A) tail; AAA, shortened poly(A) tail. Terminal modifications of the poly(A) tail observed by Chang *et al.* are also indicated in the figure. For some genes, uridylation and guanylation frequencies are associated with short and long half-lives, respectively.