Alternative polyadenylation and RNA-binding proteins

Ayse Elif Ersan-Bensan
Department of Biological Sciences, Orta Dogu Teknik Universitesi (ODTU) (METU), Universiteler Mahallesi, Cankaya, Ankara, Turkey

Abstract

Our understanding of the extent of microRNA-based gene regulation has expanded in an impressive pace over the past decade. Now, we are beginning to better appreciate the role of 3′-UTR (untranslated region) cis-elements which harbor not only microRNA but also RNA-binding protein (RBP) binding sites that have significant effect on the stability and translational rate of mRNAs. To add further complexity, alternative polyadenylation (APA) emerges as a widespread mechanism to regulate gene expression by producing shorter or longer mRNA isoforms that differ in the length of their 3′-UTRs or even coding sequences. Resulting shorter mRNA isoforms generally lack cis-elements where trans-acting factors bind, and hence are differentially regulated compared with the longer isoforms. This review focuses on the RBPs involved in APA regulation and their action mechanisms on APA-generated isoforms. A better understanding of the complex interactions between APA and RBPs is promising for mechanistic and clinical implications including biomarker discovery and new therapeutic approaches.

mRNA polyadenylation

All eukaryotic mRNAs, except histones, undergo cleavage and are polyadenylated by a multiprotein machinery consisting of subunits of cleavage and polyadenylation stimulatory factor (CPSF), cleavage stimulatory factor (CSTF), and cleavage factor IIm complexes. CPSF, CSTF, and CFIm are the core protein complexes that recruit other factors including CFIIm, the scaffolding protein symplekin and poly(A) polymerases (PAPs). CPSF recognizes poly(A) signals, which are found ~15–30 nucleotides upstream of the poly(A) site (the cleavage site). The canonical poly(A) signal recognized by CPSF is the AAUAAA sequence, which is strongly enriched and conserved among mammals. There are also other weaker poly(A) signal variants that are utilized with varying frequencies throughout the genome (Derti et al. 2012). mRNA cleavage at the poly(A) site is mediated by the endonuclease subunit of the CSTF complex that binds to the downstream U-/GU-rich element (DSE). The CFIm complex binds to U-rich/UGUA upstream elements (USEs) and also mediates the cleavage reaction. Finally, the recruited CFIIm complex aids termination of the RNA polymerase II-mediated transcription and PAPs catalyze the addition of untemplated adenosines (Proudfoot 2011).

APA

Approximately 70% of human genes have multiple poly(A) sites (Derti et al. 2012), suggesting that alternative poly(A) signals may be activated due to APA. Indeed, accumulating evidence show tightly regulated genome-wide APA in a tissue and developmental stage-specific manner. In addition, proliferation and/or activation
signals in physiological states induce genome-wide APA events. For example, widespread APA is seen during T lymphocyte activation, where usage of the proximal poly(A) site is associated with the induced proliferative state (Sandberg et al. 2008). On the contrary, genome-wide 3′-UTR lengthening is detected in mouse embryonic development, where differentiation/morphogenesis genes are upregulated and proliferation-related genes are downregulated (Ji et al. 2009).

Given the link between APA, proliferation, and differentiation, genome-wide deregulated APA events have been reported for various diseases including cancer and endocrine disorders (reviewed in Rehfeld et al. 2013). Therefore, to understand the consequences of deregulated APA events, the position of poly(A) signals is an important factor to define the composition of the resulting mRNA isoforms. If the alternate proximal poly(A) signal is at the 3′-UTR of the gene, the resulting mRNA isoforms differ only in the length of their 3′-UTRs. These isoforms with shorter 3′-UTRs generally lack the cis-elements where microRNAs and/or RBPs bind. Hence, 3′-UTR shortening generally has been linked to increased protein abundance. Numerous cases have been reported to support this correlation. For example, 3′-UTR shortening of CCND1 (Cyclin D1) in lymphomas leads to an increase in CCND1 protein levels by preventing the microRNA-mediated repression (Rosenwald et al. 2003). Another cell cycle regulator, CDC6 (cell division cycle 6), an important player in DNA replication, is upregulated, and its 3′-UTR is shortened in response to estrogen in breast cancer cells, leading to increased CDC6 protein levels and higher S-phase entry (Akman et al. 2012). HGRG14 (high-glucose-regulated gene) is differentially expressed in hyperglycemia, a complication of diabetes mellitus. HGRG14 goes through APA under high glucose conditions and produces a longer isoform that harbors adenylate-uridylate-rich elements (AU-rich elements; AREs), which eventually leads to lower protein production (Abdel Wahab et al. 1998). Similarly, shortening of the 3′-UTR of IGF2BP1 (insulin-like growth factor 2 mRNA-binding protein 1) transcript results in a more significant oncogenic transformation compared with the longer 3′-UTR isoform (Mayr & Bartel 2009).

Overall, while 3′-UTR shortening generally has been linked to increased protein levels of proliferative-related genes, interestingly, recently, upstream regions of APA sites were reported to be enriched for conserved miRNA-binding sites for pro-differentiation/antiproliferative genes, conferring stronger inhibitory activity. This finding may suggest 3′-UTR shortening to provide an additional repression mechanism for antiproliferative genes (Hoffman et al. 2016).

A second group of APA events occurs due to activation of proximal intronic poly(A) sites. Experimental evidence showed the existence of RTK (receptor tyrosine kinase) mRNA isoforms that are predicted to encode dominant-negative and secreted variants (Vorlová et al. 2011). Further evidence for intronic poly(A) site usage in receptor genes was reported in adrenocorticotrophic hormone (ACTH) producing nonpituitary tumors. To investigate the mechanism behind glucocorticoid resistance, a C-terminally truncated isoform of glucocorticoid receptor that lacks the steroid-binding domain due to APA was reported in ACTH-producing small cell lung cancer (Parks et al. 1998). Another intronic APA event was reported for the TCF7L2 (transcription factor 7-like 2) transcript that modulates insulin secretion and is implicated in Type II diabetes. Activation of an intronic poly(A) site due to APA produces a truncated TCF7L2 mRNA transcript which may be linked to a predisposition to type II diabetes (Locke et al. 2011). While the genome-wide extent of intronic APA will have to be further investigated, functional consequences of such truncated protein isoforms may have substantial importance.

Overall, based on current findings, APA is a newly appreciated genome-wide regulator of mRNA isoform diversity in normal and disease states. It is also becoming clear that we are facing a very complicated and multi-layered network of interactions between cis-elements on APA isoforms and trans-acting factors such as RBPs. The action mechanisms of RBPs in relation to APA can be investigated from two perspectives: (1) role of RBPs in APA decisions and (2) activity of RBPs on APA-generated isoforms (Fig. 1A and B).

**Role of RBPs in APA decisions**

The core RBPs in the polyadenylation machinery are the main regulators of poly(A) signal selection, cleavage, and polyadenylation. Hence, not surprisingly, changes in expression levels of these proteins cause differential selection of poly(A) signals. For example, upregulation of CSTF subunits during stem cell induction from somatic cells results in a genome-wide shift toward proximal poly(A) site activation. Specifically, CSTF2, a subunit of the CSTF complex, is a strong regulator of APA. Depletion of CSTF2 results in increased usage of distal poly(A) sites. Altered expressions of CPSF and CFIm complex members have also been implicated in the selection of alternate poly(A) signals in normal physiological settings including
self-renewal of embryonic stem cells and in disease states (reviewed in Zheng & Tian 2014). For example, decreased CFIm25 expression correlated with 3′-UTR shortening events in glioblastoma patients (Masamha et al. 2014).

Furthermore, a combinatorial approach of computation and experimentation revealed a pattern of 3′-UTR length changes in breast cancer patients/cells and a potential link between APA, CSTF2, and EGF (epidermal growth factor) signaling (Akman et al. 2015).

Apart from the core proteins, numerous RBPs have been implicated in regulating APA events by either competing with or enhancing the binding of the polyadenylation machinery proteins to their target sites (Fig. 1A). For example, HNRNP H (heterogeneous nuclear ribonucleoprotein H) binding to proximal poly(A) sites recruits polyadenylation machinery proteins, resulting in the production of shorter 3′-UTR isoforms (Katz et al. 2010). Similarly, the cytoplasmic polyadenylation element-binding protein 1 (CPEB1) binds upstream of the weaker proximal poly(A) signal and recruits the CPSF complex to promote 3′-UTR shortening events (Bava et al. 2013). By contrast, Drosophila ELAV (embryonic lethal abnormal vision) mediates neural-specific 3′-UTR lengthening by inhibiting proximal APA, which leads to transcriptional read-through and formation of a longer isoform (Hilgers et al. 2012).

Interestingly, for some RBPs, the position of the binding site determines whether RBP enhances or represses APA. For example, MBNL (muscleblind-like) differentially affects the binding of polyadenylation proteins (e.g. CSTF and CFIm) to their target regions. MBNL enhances APA by recruiting the core machinery proteins if the MBNL-binding site is upstream of the poly(A) signal. However, when the MBNL-binding site overlaps with a poly(A) site, APA is repressed (Batra et al. 2014). Likewise, NOVA2 (neuro-endocortical ventral antigen 2) can inhibit polyadenylation only when it is bound close to a poly(A) signal. On the other hand, hnRNP I inhibits polyadenylation if bound to DSEs but enhances polyadenylation when bound to USEs (Hall-Pogar et al. 2007). Another RBP with differential function is FUS (fused in sarcoma). When an APA site is upstream of an FUS-binding site, FUS enhances polyadenylation at that site by recruiting CPSF. However, when an APA site is found downstream from an FUS-binding site, polyadenylation is not activated (Masuda et al. 2015). Adding further complexity into regulation mediated by RBP- and APA-based gene expression, multiple RBPs can bind the same mRNA as was exemplified by the PTGS2 (prostaglandin-endoperoxide synthase 2, synonym COX2) USE, where several RBPs including PSF (polypyrimidine tract-binding protein-associated splicing factor), SRSF11 (serine/arginine-rich splicing factor 11, p54), hnRNP I (synonyms PTBP1 (polypyrimidine tract-binding protein 1, PTB), and U1A proteins bind to regulate APA (Hall-Pogar et al. 2007).

RBPs often have multifunctional roles in every step of the RNA lifecycle; hence, RBPs implicated in polyadenylation generally also have roles in splicing. CPSF2 and symplekin function as cofactors of the well-known splicing regulator RBFOX2 (RNA-binding protein, fox-1 homolog 2) (Misra & Green 2016). NOVA2 can bind to both introns and 3′-UTRs of target genes regulating both splicing and polyadenylation (Licatalosi et al. 2008). Similarly, U1 snRNP (small nuclear ribonucleoprotein), an essential component of the spliceosome, suppresses
Activity of RBPs in APA-generated isoforms

Considering that 3′-UTRs have major roles in determining the fate of mRNAs, APA-generated 3′-UTR isoforms are of interest to better understand the complexity of gene expression regulation. Indeed, recent developments highlight the availability of numerous cis-elements including RBP binding sites on 3′-UTRs in addition to microRNA-binding sites. AREs are the most common RBP-binding sites that have defined roles in mRNA stability and half-life regulation. Molecular mechanisms by which RBPs enhance ARE-mediated mRNA stability are not fully known, whereas destabilizing RBPs can recruit the exosome to AREs, thus promoting rapid mRNA decay (Chen et al. 2001).

To date, various ARE-binding proteins have been described. Among these, TTP (tristetraprolin) is one of the well-known proteins that bind to AREs on target mRNAs and promote their degradation. Known targets of TTP are: TNF-α (tumor necrosis factor), PTGS2, VEGF (vascular endothelial growth factor), and IL10 (interleukin 10) (reviewed in Matoulovka et al. 2012). KSRP (K homology splicing regulatory protein) is another protein involved in mRNA decay. By contrast, HuR (mammalian homolog of Drosophila ELAV) has been implicated in stabilization of various cell proliferation and cell-cycle-regulated target mRNAs including EGF (epidermal growth factor), FOS (FBj murine osteosarcoma viral oncogene homolog), BCL2 (B-cell CLL/lymphoma 2), VEGF (vascular endothelial growth factor), CDKN1A (cyclin-dependent kinase inhibitor 1A, synonym p21), CCNA1 (Cyclin A), and CCNB1 (Cyclin B1) (Wang et al. 2013). HuR is localized in the nucleus but translocates to the cytoplasm in response to stress conditions such as UV radiation and oxidative stress. HuR can also repress the translation of target mRNAs as exemplified by the cell adhesion molecule WNT5A (wingless-type MMTV integration site family, member 5A), contributing to invasion and metastasis (Leandersson et al. 2006).

In addition to RBPs having opposing functions in regulating mRNA stability, different RBPs can bind to the same target 3′-UTR. In such cases, both cooperation and antagonism have been reported between RBPs as exemplified by HuR and decay-promoting AUF1 (AU-binding factor 1, also known as hnRNP D) collectively destabilizing p16INK4A, whereas there is a competition between HuR and AUF1 for target region binding for cyclin D1 (Chang et al. 2010). Similarly, AUF1 competes for binding to BCL2 with another RBP, nucleolin, which protects BCL2 mRNA from exosomal decay (Ishimaru et al. 2010). In addition, similar cooperation or antagonism is valid for RBPs and micro-RNAs which may share common and/or overlapping binding sites. For example, HuR relieves the translational repression of ERBB2 (erb-b2 receptor tyrosine kinase 2) by blocking microRNA-binding sites, while for MYC (v-myc avian myelocytomatisis viral oncogene homolog), let-7a binding can be enhanced by increasing accessibility of the 3′-UTR (reviewed in Connerty et al. 2015).

Furthermore, binding of RBPs to the 3′-UTRs can provide a platform to facilitate protein–protein interactions. CD47, a transmembrane protein, has HuR-binding sites on its 3′-UTR. HuR binding to longer 3′-UTR isoform results in the recruitment of SET (SET nuclear oncogene) to its site of translation on the rER (rough endoplasmic reticulum), which allows SET to bind to the newly translated CD47 protein and translocate it to the plasma membrane via activated RAC1 (ras-related C3 botulinum toxin substrate 1). However, CD47 translated from the short isoform, which does not have any binding sites for HuR, is retained in the rER (Berkovits & Mayr 2015).

Conclusions

Overall, while mechanisms controlling mRNA turnover are being increasingly recognized as critical regulators of gene expression, APA-generated isoform diversity adds another layer of complexity. Moreover, the intricate relationship (summarized in Fig. 1B) between APA and trans-acting factors is possibly physiological state-specific. Hence, deregulation of the components of this crosstalk is likely to have functional consequences in disease pathologies as was shown in cancer (reviewed in Erson-Bensan & Can 2016) and in several endocrine diseases including type I and II diabetes, pre-eclampsia, ectopic Cushing
syndrome, and fragile X-associated premature ovarian insufficiency (reviewed in Rehfeld et al. 2013). Given the evidence linking proliferation and differentiation to APA, endocrine pathways are very likely to exert their intricate effects on gene expression at least partially by APA. Our current understanding of APA in endocrine diseases is somewhat limited to individual cases of mutational inactivation or activation of poly(A) sites. Therefore, a global and mechanistic understanding of how endocrine system components alter APA patterns may pave the way to a better understanding of endocrine response in diverse tissues in normal and disease states. In addition, unraveling the functional consequences of deregulated APA isoforms is promising for clinical applications such as biomarker discovery and development of novel therapies. In conclusion, considering the diversity of isoforms as well as the trans-factors acting on these isoforms, it is clear that the relationship between DNA, mRNA, and protein is not linear, and we have to tailor our experimental designs and understanding of gene expression accordingly.

Declaration of interest
The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding
Our APA work is funded by TUBITAK 112S478 and 114Z884.

Acknowledgments
The author thanks Drs S Banerjee and M Muyan for critical reading of the manuscript and apologize to many authors whose works on APA and RBPs have not been cited here because of length constraints.

References
Locke JM, Da Silva Xavier G, Rutter GA & Harries LW 2011 An alternative polyadenylation signal in TCFL2A generates isoforms that inhibit T cell factor/lymphoid-enhancer factor

http://jme.endocrinology-journals.org DOI: 10.1530/ME-16-0070 © 2016 Society for Endocrinology Printed in Great Britain


Received in final form 27 April 2016
Accepted 20 May 2016
Accepted Preprint published online 20 May 2016