Mechanisms of mRNA polyadenylation

Hızlan Hıncal AĞUŞ, Ayşe Elif ERSON BENSAN*
Department of Biology, Arts and Sciences, Middle East Technical University, Ankara, Turkey

Received: 26.05.2015 • Accepted/Published Online: 21.08.2015 • Final Version: 18.05.2016

Abstract: mRNA 3'-end processing involves the addition of a poly(A) tail based on the recognition of the poly(A) signal and subsequent cleavage of the mRNA at the poly(A) site. Alternative polyadenylation (APA) is emerging as a novel mechanism of gene expression regulation in normal and in disease states. APA results from the recognition of less canonical proximal or distal poly(A) signals leading to changes in the 3' untranslated region (UTR) lengths and even in some cases changes in the coding sequence of the distal part of the transcript. Consequently, RNA-binding proteins and/or microRNAs may differentially bind to shorter or longer isoforms. These changes may eventually alter the stability, localization, and/or translational efficiency of the mRNAs. Overall, the 3' UTRs are gaining more attention as they possess a significant posttranscriptional regulation potential guided by APA, microRNAs, and RNA-binding proteins. Here we provide an overview of the recent developments in the APA field in connection with cancer as a potential oncogene activator and/or tumor suppressor silencing mechanism. A better understanding of the extent and significance of APA deregulation will pave the way to possible new developments to utilize the APA machinery and its downstream effects in cancer cells for diagnostic and therapeutic applications.

Key words: Alternative polyadenylation, mRNA, untranslated region, cancer, cleavage stimulatory factor, cleavage and polyadenylation stimulatory factor

1. Introduction

Polyadenylation, a cotranscriptional process, was first identified in the nuclear extracts of calf thymus as early as the 1960s (Edmonds and Abrams, 1960). All eukaryotic pre-mRNAs undergo polyadenylation, i.e. cleavage and polyadenylation at the 3' untranslated region (UTR), except replication-dependent histone transcripts that are processed by small nuclear ribonucleoproteins (Marzluff et al., 2008).

In essence, a poly(A) tail is added to the 3' UTR of newly synthesized pre-mRNAs by the poly(A) polymerase, followed by the recognition of the poly(A) signal and endonucleolytic cleavage of the pre-mRNA at the poly(A) site. Conserved upstream elements (U-rich elements, UGUA elements) found at ~10–35 nucleotides upstream of the poly(A) site aid the poly(A) signal (AAUAAA, AUUAAA, and other variants) selection. Downstream elements are the U-rich and GU-rich elements (Tian and Graber, 2012). Polyadenylation prolongs mRNA lifespan by protecting the 3’ downstream sequences against exosome nucleases and plays roles in export to cytosol, localization, stability, and translation (reviewed by Guhaniyogi and Brewer, 2001; Lutz and Moreira, 2011). Given these functional roles of polyadenylation, variations in 3’ UTR processing mechanisms such as alternative polyadenylation (APA) are to be expected to affect the length/reading frame, stability, localization, and translational efficiency (Lutz and Moreira, 2011; Elkon et al., 2013). As polyadenylation is a cotranscriptional process, it is also plausible to consider mRNA processing steps to affect and/or regulate polyadenylation efficiency and specificity (Proudfoot et al., 2002).

While the extent of APA is better understood today with improvements in high-throughput sequencing techniques, isolated cases of APA events were reported much earlier, in the 1980s, for the IgM (immunoglobulin M) and DHFR (dihydrofolate reductase) genes. Intronic APA-generated isoforms of IgM were found to encode two distinct proteins, secreted if the proximal APA site is used or membrane-bound if the distal site is used (Alt et al., 1980; Early et al., 1980; Rogers et al., 1980). DHFR, on the other hand, has four isoforms with identical reading frames but different 3’ UTR lengths (Setzer et al., 1980). Since then, an improved understanding of the extent of APA events in the genome suggests an impressive 70% of known human genes to harbor multiple poly(A) sites in their 3’ UTRs and high conservation of APA sites among different species (Derti et al., 2012).
While the role of APA-based isoform generation has been linked to normal processes, findings suggest APA to be a novel and important deregulated mechanism in diseases, especially cancer. In this review, we aim to provide an overview of current data on APA and known consequences of altered 3’ UTR lengths during the complex events of tumorigenesis.

2. Polyadenylation mechanisms
To better understand the complexity of polyadenylation and APA, proteins known to be involved in these processes will be described with potential roles in cancer development when their expressions are altered. Pre-mRNA 3’ UTRs have long been known to be important for the nuclear export, stability, and translational efficiency of mRNAs. With recent developments, 3’ UTR-based posttranscriptional processing mechanisms are gaining attention as important modulators of gene expression. APA is an example of how much complexity exists for posttranscriptional regulation, specifically for 3’-end processing. So far, we know that the 3’-end processing machinery is quite complex, comprising approximately 85 proteins, some of which are known to be core elements whereas others may mediate crosstalk with other processes including splicing (Shi et al., 2009). Not surprisingly, the polyadenylation complex is quite large and has a comparable size similar to that of bacterial large ribosomal subunits (1.5 MDa) (Radermacher et al., 1987).

The polyadenylation complex consists of cleavage and polyadenylation stimulatory factor (CPSF), cleavage stimulatory factor (CSTF), cleavage factor IIm (CFIm), and cleavage factor IIm (CFIm), along with the scaffold protein symplekin and poly(A) polymerase (PAP). Known key APA machinery proteins are summarized in Figure 1.

The most conserved motif among eukaryotic mRNA polyadenylation signals is AAUAAA, specifically recognized by the multisubunit CPSF complex and therefore providing sequence specificity in the pre-mRNA cleavage site and polyadenylation. CPSF consists of hFip1, WDR33 (WD repeat domain 33), CPSF30, CPSF73, CPSF100, and CPSF160 subunits (Millevoi and Wagner, 2010). CPSF subunits CPSF30 and WDR33 directly contact the AAUAAA region (Chan et al., 2014; Schönenmann et al., 2014) and the CPSF73 subunit is the endonuclease positioned on the cleavage site (Mandel et al., 2006). PAP associates with cleavage and polyadenylation specificity factors CPSF160 and CPSF73 subunits and also with the targeted pre-mRNA (Laishram and Anderson, 2010). Evidence suggests the CPSF complex to also crosstalk with transcriptional initiation and elongation steps through interactions with transcription factor IID (TFIID), RNA polymerase II (Dantonel et al., 1997; Nag et al., 2007; Glover-Cutter et al., 2008), and splicing machinery through spliceosomal factor TRAP150 (Kwon et al., 2014).

Another complex, CSTF, recognizes U- and GU-rich sequences and is composed of 3 subunits: CSTF1 (50 kDa), CSTF2 (64 kDa), and CSTF3 (77 kDa) (Shi et al., 2009; Millevoi and Wagner, 2010). CSTF2 (also known as CSTF64) has 557 amino acid residues, including an RNA-binding domain and a pentapeptide repeat region that appears to be necessary for cleavage and polyadenylation (Takagaki et al., 1992; Takagaki and Manley, 1997). While CSTF2 interacts directly with the GU-rich region downstream of the cleavage site (Takagaki and Manley, 2000; Yao et al., 2012), CSTF1 and CSTF3 subunits interact with the C-terminal domain of RNA polymerase II, possibly to facilitate activation of 3’ UTR processing proteins (Cevher et al., 2010). Among CSTF subunits, CSTF64 has gained special attention due to its potential role in the selection of alternate (proximal or distal) poly(A) signals. A transcriptome study reported the depletion of CSTF64 (and its parologue, CSTF64r) to result in increased usage of distal poly(A) sites in HeLa cells, indicating that, when present, these factors increase the selection of proximal sites, leading to shorter isoform generation (Yao et al., 2012). Interestingly, CSTF64 was also proposed to play a key role in modulating the cell cycle in normal and in embryonic stem cells while also simultaneously controlling histone mRNA 3’-end processing (Romeo et al., 2014).

The UGUA upstream element provides binding sites for the CFI complex (CFIm25, CFIm59, and CFIm68), which facilitates the binding of CPSF (Fip1 subunit) to the UTR (Coseno et al., 2008; Yang et al., 2010). The tetrameric complex consists of the two 25-kDa subunits partnering with either the 59- or 68-kDa subunits (Martin et al., 2010). Structural studies revealed that the two CFI complexes bind to two UGUA elements in an antiparallel manner and provide the looping of RNA (Li et al., 2011; Yang et al., 2011). It appears that the CFI complex is important for poly(A) site selection as, for several genes, the sites shifted to distal ones when CFIm25 (also known as NUDT21) levels were reduced by siRNA in HeLa cells (Kubo et al., 2006; Fukumitsu et al., 2012; Gruber et al., 2012; Martin et al., 2012). In addition, loss of function of CFIm68 and CFIm25 but not of CFIm59 leads to a transcriptome-wide increase in the use of proximal polyadenylation sites in HEK293 cells (Gruber et al., 2012; Martin et al., 2012). CFI proteins are also likely to function in splicing but have different roles as CFIm59 but not CFIm68 was found to interact with splicing factor U2AF65 (Millevoi et al., 2006), whereas CFIm68 interacts with SR proteins (Dettwiller et al., 2004). CFIm68 can also shuttle between the nucleus and cytoplasm during the cell cycle (Cardinale et al., 2007) and possibly participates in mRNA export (Ruepp et al., 2000).

While the detailed mechanisms needs to be further studied, the CFIm complex seems to be important for
the efficient degradation of the 3’ product of poly(A) site cleavage and transcriptional termination (West and Proudfoot, 2008). Symplekin, as an essential polyadenylation factor, was shown to bind to CPSF73 and CSTF64 at the early steps of the polyadenylation process, possibly to function as a scaffold protein (Barnard et al., 2004; Ruepp et al., 2010).

3. APA and cancer
Following the pivotal work of Sandberg et al. (2008), describing the increase in expression of shortened mRNA 3’ UTR isoforms after activation of primary murine CD4+ T lymphocytes, various other studies demonstrated APA changes in normal and/or cancer cells by delineating functional APA sites using EST databases, microarray data, and modified RNA-seq techniques. These studies provided a link between proliferation/APA and cancer, as well as between development and tissue-specific polyadenylation patterns. For example, transcripts in the nervous system and brain and during mouse embryonic development are very specifically characterized by distal poly(A) site usage, whereas induced pluripotent stem cells (iPSCs) from differentiated cells are accompanied by global 3’ UTR shortening (Ji et al., 2009).
Given the specific APA patterns in normal cells, defective global alternative polyadenylation was reported for numerous cancers including colorectal carcinomas, esophageal cancers, breast cancers, and small intestinal neuroendocrine tumors, suggesting APA as a new target for both diagnosis and treatment options (Mayr and Bartel, 2009; Fu et al., 2011; Morris et al., 2012; Rehfeld et al., 2014; Sun et al., 2014). One of the recent findings on APA and tumorigenesis comes from glioblastomas. Approximately 1500 genes with shortened 3’ UTRs were discovered after CFIm25 knockdown followed by a marked increase in the expression of several known oncogenes, including CYCLIN D1. Furthermore, a subset of APA-deregulated genes was discovered in the low-CFIm25-expressing glioblastomas (Masamha et al., 2014). A functional link between APA and glioblastoma was also reported based on downregulation of CFIm25 expression in glioblastoma cells and enhanced tumorigenic properties. In addition to the high-throughput approaches demonstrating a global bias towards generation of short or long isoforms, individual mRNA-APA studies have also been imperative to improve the understanding of the significance of APA in cancers. For example, ECE-1 (endothelin-converting enzyme-1) is responsible for producing active endothelin-1, an overexpressed mitogenic peptide in cancers especially associated with poor prognosis and aiding the acquisition of androgen independence in prostate cancers. APA results in the production of ECE-1 transcripts with shorter 3’ UTRs, which promote elevated protein expression as a means to activate oncogenes (Whyteside et al., 2014). In line with earlier findings, our group reported the role of hormones in APA, and specifically estradiol, in breast cancers to cause 3’ UTR shortening of CDC6, possibly through E2Fs, transcription factors involved in cell cycle regulation. CDC6 itself is also a major regulator of the S phase and 3’ UTR shortening of CDC6 contributes to the consequent proliferation of ER(+) breast cancer cells in response to estradiol (Akman et al., 2012). In connection, earlier, a role for E2F was proposed for the transcriptional regulation of some APA proteins, providing a mechanistic link between APA and proliferation (Elkon et al., 2012). Another example important for cellular transformation processes is VEGFR2 (vascular endothelial growth factor receptor 2), controlled by intronic APA. The VEGFR2 gene can transcribe a functional receptor using the distal poly(A) signal or a soluble decoy protein, if the proximal intronic poly(A) signal is preferred in connection with the splicing machinery (Vorlová et al., 2011). Furthermore, as VEGFR2 mediates important angiogenic signaling, inhibitory antisense oligonucleotides to the nearest upstream splicing site of the intronic poly(A) signal results in increased expression of the soluble form of VEGFR2 protein and blockade of angiogenic signals, which may indicate the potential therapeutic implication of APA (Vorlová et al., 2011).

Taken together, it is interesting to observe how APA may provide an additional level of gene expression control. However, our understanding of how and why a specific polyadenylation site is selected over another is far from complete. Transcriptional rate, levels of key poly(A) proteins, and/or alternate targeting of isoforms by either RNA-binding proteins (RBPs) and/or microRNAs are found to be contributing factors to APA. In essence, APA (and its regulation) can be affected by many internal and external factors such as proliferation signals, hormones, developmental stages, epigenetic modifications, and even UV-induced DNA damage (Kleiman and Manley, 2001). Potential internal and external regulatory signals for APA are shown in Figure. 2. Hence, we would like to give an overview of reported APA events that are linked to the above cellular events given their significance in tumorigenesis.

4. Transcriptional kinetics

Several lines of evidence have linked APA to transcriptional activity (initiation or elongation). In mice and humans, short 3’ UTR isoforms are more abundant for highly expressed genes, whereas long 3’ UTR isoforms are more abundant for lower expression genes, possibly due to RNA polymerase II activity and/or chromatin structure (Ji et al., 2011). Since short 3’ UTRs are generally more stable due to avoidance of negative regulatory elements in the 3’ UTRs (Mayr and Bartel, 2009), more production of short 3’ UTR isoforms would make the overall expression level of proteins higher, consistent with oncogene expression cases in cancer cells. Given that highly proliferating cancer cells require increased protein synthesis, transcriptional kinetics and APA connections become interesting for further examination.

In addition to transcription initiation, transcriptional elongation may also be another factor to determine poly(A) signal selection. Supporting evidence comes from Drosophila mutants with a 50% decrease in RNA polymerase II elongation rates. In mutants, proximal poly(A) site selection was higher compared to wild-type animals (Pinto et al., 2011). Another interesting example is from yeast. In normally growing yeast cells, the proximal poly(A) site is preferred for the RPB2 (RNA polymerase II subunit B2) gene. However, after UV damage, transcription of RPB2 is initially inhibited. As transcription recovers, the distal poly(A) site is preferentially used instead, producing more of a longer form of RPB2 mRNA (Yu and Volkert, 2013).
5. Splicing
U1 is a ubiquitously expressed small nuclear ribonucleoprotein and is a major part of the spliceosome complex. Interestingly, morpholino oligonucleotide-induced knockdown of U1 in HeLa cells results in generation of premature cleavage and polyadenylation in numerous pre-mRNAs at cryptic polyadenylation signals in introns (Kaida et al., 2010). In connection to its role in polyadenylation, U1 was later proposed as a determinant of mRNA length and a regulator of different isoform expression patterns (Berg et al., 2012). Given the polyadenylation process being highly dependent on transcriptional activity, it is not surprising to observe potential interactions between the splicing machinery and APA. It is clear that we will see more examples of how APA and splicing may coregulate generation of specific UTR isoforms.

6. Epigenetic marks
Transcriptional activity is also known to affect epigenetic features; therefore, regulation of APA by transcriptional activity also correlates with differences in nucleosome positioning and histone methylation patterns around alternative poly(A) sites. Indeed, preferred poly(A) sites are strongly depleted of nucleosomes, whereas neighboring regions can be enriched for nucleosomes (Spies et al., 2009). Given the high conservation of APA signals among species and epigenetic alterations in cancer cells, further studies on mammalian systems will be of interest to investigate the depth of the connections among transcriptional kinetics, associated epigenetic marks, and APA.

7. APA protein levels
Another potential contributor to APA usage is the changes in APA protein levels in cells. Initial evidence supporting the role of key proteins in APA comes from B-cell differentiation where upregulation of CSTF64 leads to a switch from distal to proximal poly(A) site selection for the IgM heavy chain, leading to the conversion of IgM heavy chain from membrane-bound to secreted form, which is required for B-cell activation (Takagaki and Manley, 1998). CSTF64 was also shown to be upregulated upon T-cell stimulation, leading to shorter isoform generation of the transcription factor, NFATC, in effector cells, whereas both short and long isoforms are synthesized in unstimulated cells (Chuvpilo et al., 1999). Lipopolysaccharide stimulation has also been reported to increase CSTF64 expression in macrophages to promote

Figure 2. Potential regulatory signals for APA. APA is potentially affected by many factors including proliferation and activation signals, transcriptional activity, RNA polymerase elongation rate, splicing, epigenetic marks, APA protein levels, RNA-binding proteins, microRNAs, and DNA damage. These APA-inducing mechanisms are likely to crosstalk with each other. Following the APA decision, the short or long isoform of a transcript is generated. These isoforms may have the same coding sequence as different 3' UTRs or different reading frames based on the position of the poly(A) site.
alternative polyadenylation of several mRNAs (Shell et al., 2005). In addition to proliferative/activatory signals, cell death decisions may also have effects on APA. For example, CPSF73 is translocated from the nucleus to cytosol by CSR1 (cell stress response 1). CSR1 is overexpressed under oxidative stress conditions and induces apoptosis. Consequently, CPSF73 translocation to the cytosol results in a significant inhibition of overall polyadenylation activity in prostate cancers (Zhu et al., 2008).

8. RNA-binding proteins and microRNAs

Perhaps also in connection with some of the above mechanisms, growing lines of evidence have shown that RBPs play roles in the regulation of APA by either directing poly(A) site selection or being part of the APA machinery. RBPs may also regulate translation by interacting exclusively with specific cis elements in the 3' UTRs. Some of these cis regulatory elements are A+U-rich elements, CU-rich elements, stem-loop destabilizing elements (reviewed by Knapinska et al., 2005), polyadenylation inhibition elements (Clerte and Hall, 2004), and GU-rich elements (Vlasova et al., 2008). While the roles of RBPs binding to these cis elements are just beginning to be understood in cancer cells, few of these proteins have been implicated in APA as well. For example, HuR is a member of the human embryonic lethal abnormal vision (ELAV) family of proteins, originally identified in *Drosophila*, and is ubiquitously expressed. HuR has been shown to regulate many cancer-related mRNAs such as tumor suppressors (TP53, p21, and p27), oncogenes (c-fos, c-Myc), cyclins (A, B1, D1), various growth factors (VEGF, TGFβ, and TNFa), and apoptosis-related factors (BCL-2, MCL-1). HuR has also been implicated in cancer-related phenomena such as hypoxia or inflammatory signals (reviewed by Wurth and Gebauer, 2015). Interestingly, HuR overexpression in HeLa cells blocks polyadenylation at poly(A) signals only if they contain U-rich sequences, possibly by blocking the physical interaction between CSTF64 and the mRNA (Zhu et al., 2007). Therefore, alterations in HuR levels may have profound secondary effects on multiple target mRNAs both in terms of its role in stabilizing mRNAs and binding to potential poly(A) sites in cancer cells. Indeed, cytoplasmic localization of HuR has shown to be a poor prognostic factor in several cancer types including breast, lung, colon, esophageal squamous cell carcinoma, and ovarian carcinomas (Wurth and Gebauer, 2015).

PABPN1 (nuclear poly(A) binding protein) appears to be contributing to the regulation of APA by inhibiting proximal polyadenylation. PABPN1 directly interacts with noncanonical and weak proximal poly(A) signals and possibly competes with the APA machinery as loss of PABPN1 leads to a genome-wide 3' UTR shortening pattern (Elkon et al., 2012). Lower expression of PABPN1 is also linked to poor prognosis in nonsmall-cell lung cancers (Ichinose et al., 2014). Another RBP, which regulates mRNA translation, is CPEB1 (cytoplasmic polyadenylation binding protein 1), which interacts with CPSF and splicing factors in the nucleus. CPEB1 mediates shortening of hundreds of mRNA 3' UTRs. In addition to 3' UTR shortening due to alternative polyadenylation, CPEB1 mRNA interaction has also been linked to alternative splicing (Bava et al., 2013).

Interestingly, two cold-induced RBPs, CIRBP and RBM3, have been found to repress proximal poly(A) site usage under cold shock conditions, leading to longer 3' UTR isoforms (Liu et al., 2013). Further investigations revealed that many of these APA events had circadian oscillations depending on the ambient temperature. It will be of interest to investigate the mechanism of this connection.

In addition to RBPs, microRNAs are small, endogenous, noncoding RNAs that also target 3' UTRs. By binding to target mRNAs, microRNAs can degrade, destabilize, delay, or suppress translation of mRNAs (Akman and Erson-Bensan, 2014). While the role of deregulated microRNAs has been well established in cancers, an interesting metaanalysis of 3' UTR lengths and microRNA expression profiles suggested upregulated microRNA expression profiles to be highly correlated with a shift towards shorter isoforms in breast cancer cells, possibly due to increased degradation of the longer isoforms (Liaw et al., 2013). Indeed, when microRNA prediction algorithms are used, there is generally a clear preference for microRNA binding sites for the longer isoforms simply due to the availability of more binding sites. To support this observation, earlier we used several microRNA binding predictions for the short and long CDC6 isoforms and detected a common microRNA binding profile for the longer isoform. Indeed, western blot analysis and luciferase reporter assays confirmed increased CDC6 levels upon 3' UTR shortening and higher translation of the shorter isoform (Akman and Erson-Bensan, 2014). Cyclooxygenase 2 (COX-2), overexpressed in some cancers, has also been shown to be regulated by microRNAs (miR-146) in connection with APA (Cornett and Lutz, 2014). Cellular inflammatory responses, APA, and microRNA connection may be functionally interesting to examine in cancer cells as a potential way to activate oncogenes.

From the perspective of normal cellular events and cancer, it appears that APA is not regulated by a single mechanism and that different physiological states and/or signals can potentially be functionally important for APA-based UTR isoform generation. Considering the hallmarks of cancer, increased proliferation rates and activated signaling cascades, APA may emerge as an additional mechanism to explain certain oncogene activation and/
or tumor suppressor inactivation cases where no other explanation exists. Exciting developments clearly show the significance of UTR shortening and lengthening events in cancers and it is interesting to further evaluate microRNA deregulation in connection with APA. It should also be emphasized that one needs to be careful and take into consideration how RBPs may also bind to short or long isoforms in addition to microRNAs as RBPs may have stabilizing roles. Moreover, 3' UTR-isoform choice and influence on the stability and translational efficiency may not always correlate, as was shown in mouse fibroblasts (Spies et al., 2013).

On the other hand, several examples of mutated poly(A) sites were found to be associated with other diseases such as neonatal diabetes, type I and II diabetes, preeclampsia, fragile X-associated premature ovarian insufficiency, ectopic Cushing syndrome, and myotonic dystrophy type I (Rehfeld et al., 2013; Batra et al. 2014). For example, an A-to-G conversion was shown to drive the polyadenylation complex to a distal poly(A) site, which resulted in destabilization of FOX3P mRNA and decreased protein levels in IPEX (immune dysfunction, polyendocrinopathy, enteropathy, X-linked) patients (Di Giammartino et al., 2011). Aberrant spermatogenesis was also reported when testis-specific CSTF64 was disrupted (Dass et al., 2007).

In summary, the extent and functionality of APA deregulation in disease states including cancer must be established so as to pave the way to new developments to utilize APA machinery and its downstream effects for diagnostic and therapeutic applications.

Acknowledgment
Hızlan Hıncal Ağuş was funded by TÜBİTAK grants 112S478 and 114Z884.

References


Laishram RS, Anderson RA (2010). The poly A polymerase Star-PAP controls 3'-end cleavage by promoting CPSF interaction and specificity toward the pre-mRNA. EMBO J 29: 4132–4145.


AGUŞ and ERSON BENSAN / Turk J Biol


