SURVEY AND SUMMARY

Molecular mechanisms of eukaryotic pre-mRNA 3’ end processing regulation

Stefania Millevoi1,2,3,* and Stéphan Vagner1,2,3

1Institut National de la Santé et de la Recherche Médicale (INSERM) U563, Toulouse, F-31000, 2Institut Claudius Regaud, Toulouse, F-31052 and 3Université Toulouse III Paul Sabatier, Toulouse, F-31000, France

Received October 2, 2009; Revised November 27, 2009; Accepted December 1, 2009

ABSTRACT

Messenger RNA (mRNA) 3’ end formation is a nuclear process through which all eukaryotic primary transcripts are endonucleolytically cleaved and most of them acquire a poly(A) tail. This process, which consists in the recognition of defined poly(A) signals of the pre-mRNAs by a large cleavage/polyadenylation machinery, plays a critical role in gene expression. Indeed, the poly(A) tail of a mature mRNA is essential for its functions, including stability, translocation to the cytoplasm and translation. In addition, this process serves as a bridge in the network connecting the different transcription, capping, splicing and export machineries. Here, we review the eukaryotic 3’ end processing machineries as well as the comprehensive set of regulatory factors and discuss the different molecular mechanisms of 3’ end processing regulation by proposing several overlapping models of regulation.

INTRODUCTION

In eukaryotes, 3’ end cleavage of transcripts generated by RNA polymerase II (pol II) is a universal step of gene expression that proceeds through the recognition of cis-acting elements of the pre-messenger RNA (mRNA) [defined as the poly(A) signal] by a complex machinery. After cleavage, most pre-mRNAs, with the exception of histone replication-dependent transcripts, acquire a polyadenylated tail. 3’ end processing is a nuclear co-transcriptional process that promotes transport of mRNAs from the nucleus to the cytoplasm and affects the stability and the translation of mRNAs.

Although cleavage and polyadenylation can be studied as isolated processes in vitro, mRNA 3’ end formation in vivo is an integral component of the coupled network in which the different machines carrying out separate steps of the gene expression pathway are tethered to each other to form a gene expression factory. In this network, 3’ end processing cross-talks with the transcription and splicing steps to optimize the efficiency and specificity of each enzymatic reaction (Figure 1) (1). The physical interconnections between the splicing/transcription and 3’ end processing machineries create a strong functional interdependence. Indeed, 3’ end polyadenylation factors (or pA factors, including factors involved in both cleavage and polyadenylation) and sequence elements of the poly(A) signal modulate transcription termination (2–5) and, in turn, transcription factors/activators affect processing at the poly(A) signal (6–9). The phosphorylated carboxyl-terminal domain (CTD) of pol II also plays a major role in this coupling network by serving as a gathering/delivering platform of pA factors and is an integral component of the 3’ end processing complex (10,11). The functional interdependence between splicing and 3’ end processing is mediated by the molecular link between splicing factors bound at the last intron 3’ splice site and pA factors associated to the poly(A) signal in the terminal exon [(12–16) and references inside] and contributes to define the last exon of a pre-mRNA (17).

In addition to playing an essential role in the extensive network that coordinates the activities of the different gene expression machineries, 3’ end processing...
also participates in quantitative and qualitative regulatory aspects of gene expression. In transcripts carrying a single poly(A) signal, the function of the regulatory factors is to define whether to process the transcript. The regulation of the efficiency of poly(A) signal recognition will determine the level of protein expression. Indeed, transcripts that are not processed at the 3' end will be degraded or not transported efficiently to the cytoplasm. In transcripts containing more than one poly(A) signal, that is the majority of the transcription units (18–20), the role of the regulatory factors is to define where to process the transcript. Alternative 3' end processing proceeds through the choice of alternative pA signals located in the same exon or in different exons (Figure 1). The consequence of this regulation is either to change the coding sequences, resulting in different protein isoforms, or the sequences included in the 3' untranslated region (UTR) region, resulting in transcripts which may differ in their stability, localization, transport and translation properties (21–24).

Processing at a single or multiple poly(A) signals not only can be influenced by physiological conditions (including cell growth, cell cycle position, differentiation and development) but also can be altered in pathological situations (including cancer, immunity and inflammation and viral infection). The crucial role of 3' end processing in gene expression is highlighted by the increasing number of different disease entities caused by defects in the formation of proper mRNA ends (25). Indeed, disruption of this process can profoundly perturb cell viability, growth and development.

The purpose of this review is to summarize the current knowledge on the molecular mechanisms of eukaryotic 3' end processing regulation with particular emphasis on the different models of regulation supported by a comprehensive account of examples known at present. Indeed, a growing number of proteins have been identified as regulators of the 3' end processing reaction (Table 1).

These can be factors involved in other steps of gene expression (capping, splicing, transcription, stability/translation and export) or proteins of the basic cleavage/polyadenylation machinery. Their function in 3' end processing may depend on their ability to bind auxiliary/essential poly(A) signal sequences or to redistribute pA factors in alternative complexes or in different cellular compartments. Post-translational modification plays also a critical role in the regulation of the assembly of this machinery. More extensive reviews concerning various aspects of mRNA 3' end formation and its implication for health and disease can be found elsewhere (23,25–29).

**THE EUKARYOTIC PRE-MRNA 3' END PROCESSING MACHINERIES**

The 3' end processing reaction requires multiple protein factors that are generally conserved in eukaryotes and assemble onto defined sequence elements within the 3' end region of the pre-mRNA. Although the cis-elements differ in sequence and location among mammalian, yeast and plant pre-mRNAs, there appears to exist a common tripartite arrangement in which the cleavage site is associated with one A-rich element and one or more U-rich regions (Figure 2).

**Metazoan 3' end processing machinery**

The machinery leading to the formation of metazoan polyadenylated mRNAs contains several sub-complexes [Figure 2A, for a recent review see ref. 26], including cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CFIm), cleavage factor II (CFIIm), poly(A) polymerase (PAP), symplekin and the pol II. All these factors contribute to the cleavage reaction. The addition of the poly(A) tail, when tested in *in vitro* reconstituted assays, only requires CPSF, PAP and the poly(A) binding protein (PABP). PABP stimulates PAP to catalyze the addition of adenosine residues and controls poly(A) tail length by regulating the interaction between CPSF and PAP (30).

The poly(A) signal is defined by two primary sequence elements: the AAUAAA hexamer (or the more frequent variant AUUAAA) found 10–30 nt upstream of the cleavage site that binds CPSF and the U/GU-rich region located 30 nt downstream of the cleavage site (downstream sequence element or DSE) that associates with CstF. Recognition of the poly(A) signal in the absence of the canonical A(A/U)UAAA element primarily depends on the presence of an upstream UGUA sequence motif which functions in association with CFIm (31,32).

The optimal cleavage site is generally a CA dinucleotide and cleavage is catalyzed by the 73-kDa subunit of CPSF at the 3' side of the adenosine residue (33). Recently, the purification and subsequent proteomic identification and structural characterization of the human 3' end processing complex revealed a complex architecture containing ~85 proteins, including new essential factors and over 50 proteins that mediate the interplay with other processes. Among these factors, two proteins, the CPSF-associated factor, WDR33 (WD repeat domain 33),...
and the serine/threonine protein phosphatase 1 (PP1), have been shown to be required for the 3' end processing reactions (34).

The efficiency of the 3' end processing reaction is modulated by additional sequence elements located upstream (upstream sequence element or USE) or downstream (auxiliary downstream sequence element or AuxDSE) of the cleavage site. USEs are generally U-rich and serve as an additional anchor for the 3' end processing machinery by recruiting auxiliary or essential 3' end processing factors (25,35–44). AuxDSEs are generally G-rich and function by binding regulatory factors resulting in enhanced mRNA 3' end formation (45–51).

Unlike polyadenylated mRNAs, histone pre-mRNAs 3' end processing is governed by a set of rigid constraints that allow a precise coordination between regulation of their expression and DNA replication signals (Figure 2A; for recent reviews see refs 28 and 52). The replication-dependent histone processing signal lies within 100 nt downstream of the stop codon and is composed of a conserved stem–loop sequence and a more variable purine-rich element (histone downstream element or HDE) that begins 15–20 nt downstream of the stem-loop. The SLBP protein bound to the stem–loop structure acts to stabilize the binding of the U7 snRNA incorporated in the U7snRNP to the HDE. This interaction is bridged by a 100-kDa zinc finger protein (ZPF100) and involves Lsm11, a component of the U7snRNP that together with Lsm10 and the propaptotic (ZPF100) and involves Lsm11, a component of the interaction is bridged by a 100-kDa zinc finger protein snRNA incorporated in the U7snRNP to the HDE. This loop structure acts to stabilize the binding of the U7 element or HDE (that begins 15–20 nt downstream of the stop codon) that begins 15–20 nt downstream of the stem-loop. The SLBP protein bound to the stem–loop structure acts to stabilize the binding of the U7 snRNA incorporated in the U7snRNP to the HDE. This interaction is bridged by a 100-kDa zinc finger protein (ZPF100) and involves Lsm11, a component of the U7snRNP that together with Lsm10 and the propaptotic protein FLASH, are required for histone 3' end processing (53,54). The replication-dependent histone 3' end processing complex and the machinery producing polyadenylated mRNAs share a common cleavage site, the CA dinucleotide, and a core cleavage factor containing symplekin, CPSF100 and CPSF73 (55), with the endonuclease CPSF73 being the factor that cleaves the pre-mRNA (56). CPSF100 is also important for the cleavage reaction but it lacks residues critical for catalysis (57). Symplekin is the temperature-sensitive component of the essential heat-labile factor which also includes CPSF and CstF subunits (58). While using common catalytic core machinery, the histone processing reaction diverges from the process originating polyadenylated mRNAs in that it is a one-step process strictly dependent on specific signal elements and is incompatible with splicing. Although the transcription complex does not stimulate histone 3' end processing (59), recent findings uncover a physical and functional link between transcription and 3' end processing factors playing a role in the choice of the correct cleavage site to achieve the stem–loop pathway (60).

Yeast 3' end processing machinery

The factors comprised in the 3' end processing apparatus in mammals and in yeast are generally conserved but the poly(A) signals in the two organisms are rather different in consensus sequence and organization (Figure 2B; for a recent review see ref. 26). The yeast machinery comprises the cleavage and polyadenylation factor (CPF), the cleavage factor IA (CFIA) and the cleavage factor IB (CFIB). CPF contains subunits that are homologous to those in mammalian CPSF but distributed in two different sub-complexes, CFII and PFI, and include many additional factors, some of which are required for the 3' end processing functions (including Pls2, Ssu72, Mpe1, Glc7 and Ref2). The yeast homolog of symplekin, Pta1, is included in CFII and serves as a scaffold protein that is required for both cleavage and polyadenylation (61). CFIA contains subunits that are homologous to those in mammalian CFIIm and CstF, except for the absence of CstF50 in yeast. Hrp1 is the only member of CFIB and does not have a homolog in mammals. In vitro cleavage requires only CFIA, CFIB and CFII, while in vitro polyadenylation requires CPF, CFIA, CFIB and Pap1. The pol II CTD is not essential for 3' end formation at yeast poly(A) signals but it does enhance the efficiencies of both cleavage and polyadenylation (62). The yeast poly(A) signal is composed of three sequence elements: the AU-rich efficiency element (EE), the A-rich positioning element (PE) and the U-rich elements located upstream (UCE) or downstream (DUE) of the cleavage site. The latter is defined by a pyrimidine followed by multiple adenosines Y(A)n. In spite of the sequence homology, the RNA-binding specificity, the positioning and the specific function of the yeast factors are rather different from the mammalian counterpart. Indeed, in contrast to mammalian CPSF160, which interacts with the AAUAAA element, the yeast homolog, Yhl1, does not binds to the PE, the yeast counterpart of the hexamer, but near the A-rich cleavage site (63). Similarly, unlike mammalian CstF64 which associates with the U/GU rich element downstream of the cleavage site, the yeast homolog Rna15, with the help of Rna14, recognizes the A-rich PE located in the upstream region (64). As an example of different function between the yeast and mammalian homologs, mammalian CstF is involved only in the cleavage reaction, while the yeast multi-protein complex CFI is necessary for both steps. Conversely, factors without a clear sequence homology can share overlapping functions. Hrp1, which resembles the mammalian splicing factor hnRNP A1 in structure, shares several features of the mammalian CFIIm, i.e. the function in both cleavage and polyadenylation, the positioning of the binding site upstream of the cleavage site, the post-translational modifications (arginine methylation) and the association with the transcription unit ([32] and references inside).

Plant 3' end processing machinery

Most human/yeast genes have homologs in plant Arabidopsis genome, except for the mammalian factor CFI (absent also in yeast) and the yeast factor Hrp1 (Figure 2C; for a recent review see ref. 27). The plant cleavage/polyadenylation machinery is still forthcoming but a working model starts to emerge. Arabidopsis CPSF complex (AtCPSF) includes AtCPSF30, AtCPSF73-I, AtCPSF73-II, AtCPSF100, AtCPSF160, AtFIPS5 and AtFY. AtCPSF100 serves as the core of the AtCPSF complex while AtCPSF30 is physically linked to AtCstF via its interaction with AtFip, and
mediates the interaction between CPSF and other factors, such as AtCLPS3, AtSYM5, and AtPCFS4 (65). The assembly of AtCPSF in plants is dynamic (65,66) and the interactions between AtCPSF30 and other CPSF subunits are different from those existing in other eukaryotes (66). Among the plant 3' end processing factors, CPSF73(II) and FY are plant-specific proteins and are involved in specific processes, such as plant female development and flowering. Unlike human/yeast 3' end processing factors that are encoded by single genes, some Arabidopsis factors are encoded by modest gene families (e.g. ref. 67). Another important difference is that instead of being essential, as in yeast and human, some Arabidopsis pA factors affect only specific biological functions, as is the case of AtCPSF30 (68,69). The plant poly(A) signal is composed of three poorly conserved sequence elements: a far-upstream U-rich element (FUE), a near-upstream A-rich element (NUE) and a U-rich element (CE) encompassing a pyrimidine-adenosine dinucleotide that functions as the cleavage site.

MECHANISTIC MODELS OF PRE-MRNA 3' END PROCESSING REGULATION

The large complexity of the 3' end processing complexes and the flexibility of the associated auxiliary/essential signal elements reflects the ability of this process to be subjected to extensive regulation. Based on a comprehensive set of examples of 3' end processing regulation in eukaryotic organisms, several overlapping mechanisms of action can be proposed as described below (Figure 3 and Supplementary Table S1).

Competition with pA factors for binding to a poly(A) signal

Since the assembly of the polyadenylation machinery primarily depends on the cooperative binding of CPSF and CstF to the core polyadenylation signal (70), an efficient way to regulate 3' end processing is to impede the formation of this heterotrimeric complex by competing with the basal pA factors for recognition of the core elements. In most cases, the regulatory factors block the association of CstF64 to the U/GU-rich DSE and this occurs by their direct binding to the DSE or by their association with elements in the close proximity of the DSE (Figure 3A).

A regulator which plays a negative function by competing directly for binding sites with CstF is the polypyrimidine tract binding protein (PTB), a major hnRNP protein that plays multiple roles in mRNA metabolism, including mRNA 3' end formation (71). The outcome of the competition between the two factors for binding to the DSE might be determined by the relative strength of the DSE and by the physiological variation in PTB levels (72). PTB has been shown to inhibit the α-, β-globin and complement C2 poly(A) signals (71) but, since the DSE is commonly present in metazoan poly(A) signals, this factor is expected to lead to a general downregulation in mRNA expression.

The second mechanism of action depends on the presence of auxiliary regulatory elements flanking or partially overlapping the DSE and, in most cases, is used to repress the choice of alternative poly(A) sites. The U1A

<table>
<thead>
<tr>
<th>Factor</th>
<th>Role</th>
<th>RNA</th>
<th>Sequence element</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metazoan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-3-3-Σε</td>
<td>−</td>
<td>Most cellular pre-mRNAs (tested on c-mos)</td>
<td>Upstream consensus</td>
<td>(126)</td>
</tr>
<tr>
<td>9G8</td>
<td>+</td>
<td>HIV-1</td>
<td>9G8 binding site</td>
<td>(108)</td>
</tr>
<tr>
<td>BARD1</td>
<td>−</td>
<td>Expected general function (tested on SV40 late and AdL3)</td>
<td></td>
<td>(135–137)</td>
</tr>
<tr>
<td>Cdc73</td>
<td>+</td>
<td>Specific subset of pre-mRNAs including Integrator complex subunit 6 (INT6)</td>
<td></td>
<td>(138)</td>
</tr>
<tr>
<td>CDK9</td>
<td>+</td>
<td>Pre-mRNAs with tandem pA signals including Metallopeptidase inhibitor 2 (TIMP-2), Syndecan 2 and Excision repair (ERCC6), Dihydrofolate reductase (DHFR)</td>
<td></td>
<td>(159)</td>
</tr>
<tr>
<td>CFIm25</td>
<td>−</td>
<td>CFIm68</td>
<td></td>
<td>(122)</td>
</tr>
<tr>
<td>CFIm68</td>
<td>−/+</td>
<td>CFIm68</td>
<td>3 UGUAA motifs</td>
<td>(31)</td>
</tr>
<tr>
<td>CFIm68</td>
<td>+</td>
<td>poly(A) polymerase α (PAPOLA)</td>
<td>UGUAA motifs</td>
<td>(32)</td>
</tr>
<tr>
<td>CFIm68</td>
<td>+</td>
<td>poly(A) polymerase γ (PAPOLG)</td>
<td>flanking and overlapping non-canonical poly(A) signal</td>
<td></td>
</tr>
<tr>
<td>CFIm68 CFIm25</td>
<td>+</td>
<td>Non-canonical pA signals in male germ cell transcripts including CFIm68 and CFIm25</td>
<td>UGUAA motifs</td>
<td>(121)</td>
</tr>
<tr>
<td>CFIm68 CFIm25</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSR1</td>
<td>+</td>
<td>Expected general function (tested on β-actin and GAPDH)</td>
<td></td>
<td>(129)</td>
</tr>
<tr>
<td>CstF64</td>
<td>+</td>
<td>IgM</td>
<td></td>
<td>(116,117)</td>
</tr>
<tr>
<td>eIF3f</td>
<td>−</td>
<td>HIV-1</td>
<td></td>
<td>(118)</td>
</tr>
<tr>
<td>ELL2</td>
<td>+</td>
<td>Immunoglobulin heavy chain locus</td>
<td>Promoter</td>
<td>(124)</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Factor</th>
<th>Role</th>
<th>RNA</th>
<th>Sequence element</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>hnRNP F</td>
<td>−</td>
<td>IgM Melanocortin receptor 1 (MC1-R)</td>
<td>DSE</td>
<td>(48)</td>
</tr>
<tr>
<td>hnRNP H</td>
<td>+</td>
<td>Expected to influence a specific subset of pre-mRNAs (tested on SV40 late and others)</td>
<td>G-rich AuxDSE</td>
<td>(49)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expected to influence a specific subset of pre-mRNAs including β-globin</td>
<td>G-rich AuxDSE</td>
<td>(45,46,104)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rous sarcoma virus (RSV)</td>
<td>Upstream negative regulator of splicing (NRS) element</td>
<td>(50)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Expected to influence a specific subset of pre-mRNAs including β-globin</strong></td>
<td><strong>CT/CGRP</strong></td>
<td></td>
</tr>
<tr>
<td>hnRNP I (PTB)</td>
<td>+</td>
<td>Cyclooxygenase-2 (COX-2) Complement C2 (C2) Glial fibrillary acidic protein (GFAP) α-Tropomyosin</td>
<td>U-rich USE</td>
<td>(40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expected to influence a specific subset of pre-mRNAs including β-globin</td>
<td>Upstream pyrimidine-rich element</td>
<td>(162,163)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>CT/CGRP</strong></td>
<td>Downstream enhancer (5′, 3′ splice sites and a pyrimidine-rich element)</td>
<td>(115)</td>
</tr>
<tr>
<td>hnRNP I (PTB)</td>
<td>−</td>
<td>A subset of pre-mRNAs including Prothrombin F2 (F2)</td>
<td>U-rich USE</td>
<td>(103)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expected to influence many pre-mRNAs including β-globin, α-globin and Complement C2 (C2)</td>
<td>DSE</td>
<td>(71)</td>
</tr>
<tr>
<td>hnRNP L</td>
<td>+</td>
<td>Herpes simplex virus thymidine kinase (HSV-TK)</td>
<td>Upstream pre-mRNA processing enhancer (PPE)</td>
<td>(164)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expected to influence a specific subset of pre-mRNAs including N-acylsphingosine amidohydrolase (acid ceramidase) 1 (ASAH1)</td>
<td>U-rich sequences close to AAUAAA</td>
<td>(165)</td>
</tr>
<tr>
<td>HSF1</td>
<td>+</td>
<td>Expected to influence Heat shock protein (HSP) pre-mRNAs including Hsp70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu</td>
<td>−</td>
<td>Expected to influence a specific subset of pre-mRNAs including CT/CGRP</td>
<td>Upstream pre-mRNA processing enhancer (PPE)</td>
<td>(165)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>CT/CGRP</strong></td>
<td>Downstream enhancer (5′, 3′ splice sites and a pyrimidine-rich element)</td>
<td>(115)</td>
</tr>
<tr>
<td>IRBIT</td>
<td>−</td>
<td>Expected to influence a specific subset of pre-mRNAs (tested on SV40 late)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NELF</td>
<td>+</td>
<td>Replication-dependent histones Mouse brain transcripts</td>
<td>5′ cap</td>
<td>(60)</td>
</tr>
<tr>
<td>Nova</td>
<td>+/−</td>
<td></td>
<td>YCAP-rich motifs in the 3′-UTR</td>
<td>(166)</td>
</tr>
<tr>
<td>P54</td>
<td>+</td>
<td>Cyclooxygenase-2 (COX-2)</td>
<td>U-rich USE</td>
<td>(40)</td>
</tr>
<tr>
<td>P54</td>
<td>+</td>
<td>Expected general functional (tested on model substrate SV40 late)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIPK2a</td>
<td>+</td>
<td>Detoxification and/or oxidative stress response transcripts</td>
<td>U-rich USE</td>
<td>(40)</td>
</tr>
<tr>
<td>PSF</td>
<td>+</td>
<td>Cyclooxygenase-2 (COX-2)</td>
<td>U-rich USE</td>
<td>(40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expected general function (tested on Gal5HIV2dsx reporter)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tested on SV40 late</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sam68</td>
<td>+</td>
<td>HIV-1</td>
<td>U-rich USE</td>
<td>(7)</td>
</tr>
<tr>
<td>Slim1, Slim2</td>
<td>+</td>
<td>HIV-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snc1</td>
<td>+</td>
<td>Expected to influence a specific subset of pre-mRNAs including Xiro1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR proteins</td>
<td>+</td>
<td>Gli al fibrillary acidic protein (GFAP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9G8, SRp40, SC35, ASF/SF2, Srp55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR proteins</td>
<td>+</td>
<td>Rous sarcoma virus (RSV)</td>
<td>Negative Regulator of Splicing element (NRS)</td>
<td>(170)</td>
</tr>
<tr>
<td>(SRp20, ASF, 9G8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRm160</td>
<td>+</td>
<td>Expected general regulator (tested on β-globin)</td>
<td>Downstream enhancer (5′ and 3′ splice sites and a pyrimidine-rich element)</td>
<td>(97,98,114)</td>
</tr>
<tr>
<td>SRp20</td>
<td>+</td>
<td>CT/CGRP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Factor</th>
<th>Role</th>
<th>RNA</th>
<th>Sequence element</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRp75</td>
<td>–</td>
<td>Expected to influence a specific subset of pre-mRNAs (tested on AdL3)</td>
<td>Artificial binding site</td>
<td>(86)</td>
</tr>
<tr>
<td>U1 snRNP</td>
<td>+</td>
<td>CT/CGRP</td>
<td>Downstream enhancer (5- and 3-splice sites and a pyrimidine-rich element)</td>
<td>(114)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expected general role (tested on SV40 late)</td>
<td>Upstream binding site</td>
<td>(171)</td>
</tr>
<tr>
<td>U1 snRNP</td>
<td>–</td>
<td>HIV-1</td>
<td>Downstream 5-splice site</td>
<td>(172,173)</td>
</tr>
<tr>
<td>(U170K)</td>
<td>–</td>
<td>Bovine papillomavirus (BPV)</td>
<td>Upstream 5–splice site</td>
<td>(85,173)</td>
</tr>
<tr>
<td>U1A</td>
<td>–</td>
<td>IgM</td>
<td>Between GU-rich elements at DSE</td>
<td>(73)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Upstream AUGC motifs</td>
<td>(84)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Upstream stem–loop structure</td>
<td>(12,82,83)</td>
</tr>
<tr>
<td>U1A snRNP</td>
<td>+</td>
<td>SV40 late</td>
<td>USE</td>
<td>(101)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expected general role (tested on SV40 late)</td>
<td>USE? (100)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>USE</td>
<td>(40)</td>
</tr>
<tr>
<td>U2 snRNP</td>
<td>+</td>
<td>Expected general role (tested on SV40 late)</td>
<td>Branchpoint at the upstream 3-splice site</td>
<td>(16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication-dependent histones</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Upstream C/GAAGAAAG motif</td>
<td>(96)</td>
</tr>
<tr>
<td>U2AF35</td>
<td>+</td>
<td>A subset of pre-mRNAs including Prothrombin F2 (F2)</td>
<td>U-rich USE</td>
<td>(103)</td>
</tr>
<tr>
<td>U2AF65</td>
<td>+</td>
<td>Expected general role (tested on model substrate β-globin and AdL3)</td>
<td>Pyrimidine tract at the upstream 3’ splice site</td>
<td>(14,15)</td>
</tr>
<tr>
<td>U2AF65</td>
<td>–</td>
<td>Expected to influence a specific subset of pre-mRNAs (tested on albumin)</td>
<td>Poly(A)-limiting element (PLE)</td>
<td>(174)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expected to influence a specific subset of pre-mRNAs (tested on AdL3)</td>
<td>Artificial binding site</td>
<td>(86)</td>
</tr>
<tr>
<td>tcstf64</td>
<td></td>
<td>Mouse testes transcripts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLPS3</td>
<td>+</td>
<td>Specific subset of pre-mRNAs including FCA</td>
<td>ARE</td>
<td>(176)</td>
</tr>
<tr>
<td>FCA</td>
<td>+</td>
<td>Specific subset of pre-mRNAs including FCA</td>
<td></td>
<td>(102,177)</td>
</tr>
<tr>
<td>PCFS4</td>
<td>+</td>
<td>Specific subset of pre-mRNAs including FCA</td>
<td></td>
<td>(67)</td>
</tr>
<tr>
<td>Yeast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cth2</td>
<td>+</td>
<td>Adenosine/uridine-rich elements (ARE)-containing transcripts including Membrane anchor subunit of succinate dehydrogenase (SDH4)</td>
<td>ARE</td>
<td>(178)</td>
</tr>
<tr>
<td>Nab4/hrpl</td>
<td>+</td>
<td>Specific subset of pre-mRNAs including Yeast homolog of human TF1HB (SUAT7), copper transporter 2 (CTR2) and others</td>
<td>UA-rich element upstream CS</td>
<td>(179)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specific subset of pre-mRNAs including Iso-1-cytochrome c (CYC1) and GAL7</td>
<td></td>
<td>(180)</td>
</tr>
<tr>
<td>Npl3</td>
<td>–</td>
<td>Npl3</td>
<td>U4 repeat and A-rich, UG motifs</td>
<td>(154)</td>
</tr>
<tr>
<td>Npl3</td>
<td>–</td>
<td>Expected general role (tested on GAL7)</td>
<td></td>
<td>(79)</td>
</tr>
<tr>
<td>Virus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICP27</td>
<td>+</td>
<td>Herpes Simplex Virus 1 (HSV-1) and other pA signals α-globin</td>
<td></td>
<td>(133,181)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(182)</td>
</tr>
<tr>
<td>NS1</td>
<td>–</td>
<td>Most cellular transcripts (tested also on β-actin)</td>
<td></td>
<td>(131,132)</td>
</tr>
<tr>
<td>SM</td>
<td>+</td>
<td>Epstein–Barr virus (EBV) DNA polymerase</td>
<td></td>
<td>(183)</td>
</tr>
<tr>
<td>Fly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELAV</td>
<td>-</td>
<td>Erect wing (Ewg)</td>
<td>Downstream tandem AU-rich motifs</td>
<td>(95)</td>
</tr>
<tr>
<td>SXL</td>
<td>-</td>
<td>Enhancer of rudimentary (et(r))</td>
<td>GU-rich DSE</td>
<td>(77)</td>
</tr>
</tbody>
</table>

HIV-1, human immunodeficiency virus type 1; SV40 late, simian virus 40 late; AdL3, adenovirus L3; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; IgM, immunoglobulin M heavy chain; CT/CGRP, calcitonin/calcitonin gene-related peptide; FCA, flowering time control protein; GAL7, galactose-1-phosphate uridyl transferase.
splicing factor binds in between two GU-rich regions downstream of the secretory poly(A) site of the immunoglobulin M (IgM) heavy chain pre-mRNA and inhibits both the binding of CstF64 to the GU-rich region and cleavage at this poly(A) site, resulting in polyadenylation switching at the membrane poly(A) site (73). The antagonistic binding of U1A and CstF64 can be explained by the observation that the U1A binding site extends to the adjacent distal GU-rich region, thus blocking the access of CstF64 to this element (73). Using a similar competition mechanism, the splicing factor hnRNP F hinders the assembly of a stable polyadenylation complex at the IgM secretory-specific poly(A) site through its binding near the downstream GU-rich element (48). HnRNP F can interfere with the ability of CstF64 to bind the RNA by impeding the conformational change occurring during RNA binding (74) or by multimerizing and protecting the adjacent GU-rich element. Unlike U1A or hnRNP F, the inhibitory function of neuron-specific members of a family of RNA-binding proteins, Hu
proteins, known to regulate mRNA stability and translation in the cytoplasm, depends not only on the binding at sequences adjacent to the AUAAA hexamer, i.e. the CPSF binding site, but also on their ability to interact with both CPSF and CstF (75). This mechanism may function in conjunction with other factors (see below) to block polyadenylation of calcitonin (CT)/calcitonin gene-related peptide (CGRP) exon 4 thus promoting the neuron-specific pathway in which exons 5 and 6 are included in the mature transcript to produce CGRP (76).

3' end processing regulation following the above-described 'competition mechanism' is not restricted to mammalian factors. The Drosophila master sex-switch protein Sex-lethal (SXL), involved in both splicing and translation, competes with CstF64 for binding to the GU-rich DSE at the proximal poly(A) signal on the enhancer of rudimentary [e(r)] mRNA. The consequence of such regulation is to promote a female-specific poly(A) switching onto an otherwise distal non-responsive poly(A) signal. SXL-mediated alternative polyadenylation may provide a new mechanism for the retention of an SXL-binding site(s) (in the 3' UTR) in a female germline-specific manner for translation repression (77).

Similarly, the yeast hnRNP protein Npl3 inhibits mRNA 3' end formation by binding G+U-rich sequences (78), thus preventing the binding of Rna13, the Cstf 64 homolog, to the GAL-7 RNA (79). This regulatory mechanism, which depends on the phosphorylation status of Npl3, might be involved in masking weak or cryptic poly(A) sites thus ensuring recognition of the proper poly(A) signal (80).

**Inhibitory interaction between PAP and RNA-bound factors**

The RNA-binding activity of a regulatory factor can also be associated with repression of 3' end processing without interfering with the association of the core 3' end processing factors to the poly(A) signal. In that case, a protein bound to defined RNA elements in the vicinity of the AAUAAA hexamer interacts with PAP to block poly(A) tail addition (Figure 3B).

One of the best-characterized examples of this regulatory mechanism is the autoregulation of U1A pre-mRNA polyadenylation by the snRNP-free U1A. Cooperative binding of two U1A molecules to two stem–loop sequences in the 3' UTR of the U1A pre-mRNA induces a conformational change that allows a defined region of the U1A/U1A dimer to interact with the last 20 residues located at the C-terminus of PAP resulting in inhibition of polyadenylation (12,81–83). U1A binding to its own pre-mRNA prevents neither the binding of CPSF, thus excluding a steric block model, nor the cleavage reaction, indicating the specificity of the inhibitory mechanism for the polyadenylation step. This mechanism ensures that the activity of this enzyme, which is essential to the cell, is only downregulated when U1A, present in excess in the cell and therefore not engaged in U1 snRNP, binds the U1A pre-mRNA.

The RNA-dependent inhibition mechanism of PAP activity is not restricted to autoregulation. Indeed, U1A bound to three elements upstream of the secretory poly(A) site of the IgM pre-mRNA selectively inhibits polyadenylation in a developmental manner (84). Besides U1A, other splicing factors have been found to block poly(A) tail addition when bound upstream of the AAUAAA. The U170K subunit of the U1 snRNP (85) and the SR proteins, U2AF65 and SRp75 (86), share both a protein domain similar to the PAP regulatory domain of U1A (12) and a similar function in repressing PAP activity. In these factors the aforementioned domain can be present in a single or multiple copies and, for U2AF65 and SRp75, it resides in the arginine-serine-rich (RS) domain (12). In the case of U170K protein, the inhibitory mechanism involves the binding of the U1 snRNP to a cryptic 5' splice site located upstream of the BPV poly(A) signal and is used to repress late gene expression at early time of viral infection (85).

---

**Figure 3.** Mechanistic models of 3' end processing regulation of eukaryotic pre-mRNAs. Based on different examples of pre-mRNA 3' end processing regulation, eight mechanisms of action can be delineated. (A) Competition with pA factors for binding to a (C) Poly(A) signal (pA signal). (B) Inhibitory interaction between PAP and RNA-bound factors. (C) Poly(A) signal recognition in kinetic competition with splicing and transcription. (D) Recruitment of pA factors to a poly(A) signal. (E) Recruitment to alternative poly(A) sites. (F) Redistribution of pA factors to the cytoplasm. (G) Redistribution of pA factors in nuclear complexes. (H) Post-translational modification (PTM) of pre-mRNA 3' end processing factors and regulators. 3' end processing regulators are in red (negative regulation) or in green (positive regulation) while pA factors are in white. Protein–protein interactions are indicated with double-headed arrows while the RNA-binding activity of regulatory or pA factors is depicted with a thick, simple arrow in gray.
Given the importance of the extreme C-terminal region of PAP in polyadenylation repression by proteins sharing the PAP regulatory domain, the question is raised as to how this interaction influences the activity of PAP. Structural determination of this region of PAP alone and/or in complex with the regulatory proteins will provide additional information regarding the mechanistic details of this regulatory mechanism.

**Competition with other kinetically linked processes**

According to the kinetic model of 3' end processing regulation, the strength of a poly(A) signal is correlated with the rate of cleavage assembly compared to that of other temporally and physically linked processes (87), i.e. transcription elongation and splicing (Figure 3C).

This model predicts that protein factors or sequence/structural elements which modify the rate of processes competing with cleavage/polyadenylation can affect the efficiency of poly(A) signal recognition. In support of a kinetic view of poly(A) site utilization, it has been reported that transcriptional elongation, through effects on transcriptional pausing or arrest, affect poly(A) site recognition (2,88–90). Npl3 offers an interesting example of a protein factor which mediates the competition between pol II elongation and poly(A) site choice (discussed above and below). Conversely, factors modifying the efficiency of a poly(A) signal can indirectly influence processes which are temporally linked with 3' end processing. In agreement with this, the strength of the poly(A) signal affects the efficiency of transcriptional termination (91–94). The regulation by ELAV, a neuron-specific regulator of pre-mRNA processing of Drosophila, supports a kinetic link between polyadenylation and splicing. Indeed, ELAV binds to erect wing (ewg) RNA 3' of a poly(A) signal within the terminal intron and inhibits 3' end processing at this site. The RNA-binding activity of ELAV together with the presence of a functional poly(A) signal is required to promote splicing of the terminal intron. Since ELAV does not interfere with the recognition of the poly(A) signal by CPSF and CstF, it has been proposed that the inhibitory activity of ELAV may reside in its ability to slow down the recruitment of CFI, CFII or PAP, resulting in delay of the cleavage reaction. The role of ELAV could be to shift the kinetic competition between the two processes in favor of the recognition of the 3' splice site leading to neural-specific splicing of ewg pre-mRNA (95).

**Recruitment of 3' end processing factors to a 3' end processing signal**

Positive regulation of 3' end processing is mostly achieved by factors that bind the pre-mRNA and recruit the polyadenylation machinery through the association with a cleavage/polyadenylation factor. The interaction between the regulatory and the basal polyadenylation factor can be direct or mediated by a bridging factor (Figure 3D).

The direct recruitment mechanism is used by splicing factors to mediate the functional interplay between the splicing and 3' end processing machinery. Three main actors have been described to be directly involved in this coupling: U2AF65, U2 snRNP and SRm160. U2AF65, the large subunit of the U2AF factor, bound to the pyrimidine tract at the last intron 3' splice site, stimulates both cleavage and polyadenylation by recruiting the heterodimeric cleavage factor CFIy m 59/25 at the poly(A) signal (15). This interaction involves the RS region present in both factors and mediates the ability of the last intron pyrimidine tract to positively influence mRNA 3' end formation ([14,16] and references inside). The U2AF65 RS region is also involved in the reciprocal regulation, i.e. the stimulation of splicing by the polyadenylation machinery, but in this case U2AF65 is tethered on the RNA at the 3' splice site through its interaction with PAP (13). Coupling between the two processes also requires the functional interaction between the SF3b subunit of the splicing factor U2 snRNP bound to the branch point site of the upstream intron and CPSF associated to the poly(A) signal (16). The U2 snRNP-dependent function in 3' end processing is conserved between polyadenylated and histone pre-mRNAs. Indeed, the SF3b subunit of U2 snRNP, in conjunction with hPrp43, a DEAH-box helicase, contacts directly a conserved motif within the histone transcript and stimulates U7 snRNP-dependent cleavage. This may occur through the interaction between U2 snRNP and CPSF, which contains CPSF73, the endonuclease for U7-snRNP-dependent cleavage (96). Targeting CPSF to stimulate polyadenylation machinery assembly is the mechanism used also by the SR protein, SRm160, a splicing coactivator and component of the splicing-dependent exon junction complex (97). The functional interaction between SRm160 and the 3' end processing machinery is evolutionary conserved and does not involve the splicing-dependent exon junction complex (98) but depends on the interaction between the PWI domain of SRm160 and the pre-mRNA (99).

Besides coupling between splicing and 3' end processing, the association between CPSF and the AAUAAA hexamer can be increased by the splicing factor U1A, in its U1 snRNP-free form (100), probably when bound to upstream auxiliary sequences playing a positive role in 3' end processing regulation (101). The recruiting mechanism is also used to modulate poly(A) site selection in specific biological processes. The interaction between the WW domain of the Arabidopsis RNA-binding factor FCA and the polyadenylation factor FY promotes the choice of the promoter-proximal polyadenylation site within the FCA pre-mRNA to produce a transcript encoding a non-functional protein. This negative autoregulatory loop contributes to control the Arabidopsis floral transition (102).

The indirect mechanism consists in recruiting a protein factor, which in turn interacts with one component of the polyadenylation machinery. This mechanism is exemplified by the splicing factor PTB, that in addition to playing a repressor role when competing with CstF binding at DSE (71), is able to stimulate 3' end processing when associated to upstream elements (35,40,50,103). The positive function of PTB is mediated by another splicing and 3' end processing factor, hnRNP H, which binds to G-rich AuxDSE to stimulate cleavage and
polyadenylation (45,46,49,50,104–106). PTB increases the RNA-binding activity of hnRNP H and, in turn, hnRNP H recruits either CstF (46,104) or PAP (50) facilitating the 3' end processing reaction. Similarly, the f subunit of the eukaryotic initiation factor 3 (eIF3f), which co-purifies with the 3' end processing complex (34), interacts with both the cyclin-dependent kinase 11 (CDK11) and the SR splicing regulator 9G8, and modulates cleavage of the 3' end of the HIV-1 RNA by regulating the sequence-specific recognition of 9G8. Since 9G8 interacts with CFIm (107), this regulatory mechanism is thought to directly affect the assembly of the 3' end machinery (108).

An indirect recruitment mechanism can also explain how the 5' cap structure influences the efficiency of 3' end processing (109–111). This stimulatory effect is mediated by the physical association between the nuclear cap binding complex (CBC) and, in turn, CBC associates with SLBP to determine whether 3' end processing will follow the stem–loop pathway or the aberrant polyadenylation pathway (60).

Another interesting example is provided by the recognition of the human CT/CGRP exon 4 poly(A) signal by factors bound to a downstream intronic enhancer. This cis-element increases 3' end processing of the exon 4 poly(A) signal by stimulating the binding of CstF64 to the RNA (113). A number of factors could mediate this effect, including the splicing factors U1 snRNP, ASF/SF2, SRp20 and PTB (113–115). While the mechanism of regulation by PTB could involve competition between PTB and U2AF65 for binding to the enhancer pyrimidine tract (115), SRp20 may function directly by recruiting factors at the exon 4 poly(A) signal or indirectly by stabilizing the binding of U1 snRNP which in turn stimulates poly(A) signal recognition (114).

pA factor-mediated redirection of the 3' end processing machinery to alternative poly(A) sites

Regulated processing at the pre-mRNA 3' ends can be induced by the differential expression of constitutive pA factors. In most cases, the consequence of this regulatory mechanism is to promote the selection of alternative poly(A) signals that insufficiently recruit the polyadenylation machinery due to the presence of suboptimal cis-acting elements (Figure 3E).

According to this model, increased expression of CstF64 has been associated to alternative polyadenylation of several pre-mRNAs in various biological situations (116–120). During B-cell differentiation, upregulation of CstF64 is proposed to result in a switch of IgM heavy-chain mRNA from membrane-bound to a secreted form (116,117). A similar CstF64 dose-dependent switch from distal to proximal poly(A) signal selection in the transcript encoding the transcription factor NF-Atc occurs during T-cell differentiation (119). In macrophages, lipopolysaccharide stimulation increases CstF levels resulting in alternative polyadenylation of several pre-mRNAs (118). More recently, the differential expression of CFIm subunits, CFIm25 and CFIm68, in mouse male germ cells has been correlated to the utilization of alternative promoter proximal poly(A) signals in a number of transcripts, including those encoding for the regulatory factors, suggesting autoregulation of both CFIm subunits (121). In agreement with this study, knocking down CFIm25 results in alternative poly(A) signal selection (122). The emerging view is that poly(A) signals lacking the A(A/U)UAAA hexamer but containing the CFIm binding site (i.e. the UGUA element) would be favored under conditions of high CFIm levels, whereas the distal, often canonical poly(A) signals are used when the concentration of CFIm is low (121). As in the case of CstF64, the underlying mechanism is the recruitment of the polyadenylation machinery to an unfavorable poly(A) site by the increased binding of a constitutive factor.

pA factor-mediated poly(A) switch can be directed by mechanisms other than an increase of pA factors’ expression. The use of the promoter-proximal secretory poly(A) signal of the immunoglobulin heavy-chain locus is accompanied by more binding of phosphorylated pol II CTD and of the transcription elongation factor ELL2 to the transcription start site region along with more loading of CstF64 onto pol II. The binding of ELL2 and CstF-64 to pol II is dependent on serine 2 phosphorylation on the pol II CTD (123). The proposed model is that ELL2 promotes CstF64 binding to phosphorylated pol II and, as a consequence of this loading, pA factors present at high local concentration act on the weak secretory-specific poly(A) site to direct its recognition and cleavage (124).

Redistribution of pA factors to the cytoplasm

Regulation of the subcellular partitioning of mRNA binding proteins is an important aspect of the post-transcriptional control of gene expression. Recent reports suggest that this regulatory mechanism influences not only the splicing, stability and translation processes (125) but also the polyadenylation status of the transcripts (Figure 3F).

The first evidence supporting a role for the redistribution mechanism in controlling 3' end processing was the regulation of PAP by 14-3-3e, a member of the 14-3-3 protein family (126). This regulation is mediated by a direct, phosphorylation-dependent interaction between this factor and the C-terminal region of PAP. The consequence of this interaction is the inhibition of the polyadenylation activity of PAP and the increase in its cytoplasmic localization. 3' end processing regulation by 14-3-3e may also lead to a more gene specific regulation of mRNA expression by targeting an auxiliary factor of the polyadenylation machinery. Indeed, activation of 14-3-3e by the extracellular signal-regulated protein kinase (ERK) during the heat-shock response induces cytoplasmic sequestration of the heat-shock transcription factor 1 (HSF1) (127). In addition to inducing transcription of heat-shock protein (HSP) genes, HSF1 plays a role in enhancing the polyadenylation efficiency of this class of genes by interacting with symplekin in a stress-induced
manner (128). Therefore, 14-3-3ε-mediated sequestration of HSF1 in the cytoplasm may contribute to the attenuation of HSF1 nuclear functions in upregulating HSP genes expression, a process which could take place once the acute phase of the response to stress is over (127).

More recently, two other factors have been demonstrated to influence polyadenylation by binding 3' end processing factors and inducing their translocation from the nucleus to the cytoplasm. The cellular stress response 1 (CSR1) is a tumor-suppressor protein that interacts with CPSF73 inducing its redistribution to the cytoplasm and, as a consequence, inhibition of polyadenylation (129). It has been proposed that inhibition of CPSF activity may be the mechanism by which the tumor-suppressor CSR1 mediates cell death (129). The mechanism of action of the second factor, IP3R-binding protein released with inositol 1,4,5-triphosphate (IRBIT), is a protein involved in calcium signaling and regulation of intracellular and extracellular pH, closely resembles that of 14-3-3ε. Indeed, IRBIT binds PAP and the hFip1 subunit of CPSF in a phosphorylation-dependent manner and causes both inhibition of polyadenylation and redistribution of hFip1 into the cytoplasm (130). This regulatory mechanism takes place in response to oxidative stress, which leads to modification of IRBIT phosphorylation status (130).

These examples make clear the importance of the redistribution mechanism in downregulating polyadenylation in the nucleus. An important unsolved issue is to establish whether redistribution of nuclear pA factors have consequences on other nuclear (i.e. transcription, capping, splicing, histone 3' end processing and snRNA processing) or cytoplasmic (i.e. cytoplasmic polyadenylation) processes where 3' end processing factors play also an important function.

Redistribution of pA factors in nuclear complexes

The physical association of regulatory factors to the core processing machinery assembled onto the bipartite poly(A) signal and subsequent redistribution of these factors in new protein complexes is a mechanism that further contributes to modulate 3' end processing efficiency in a positive or in a negative manner (Figure 3G).

Influenza A infection provides an interesting example of regulation of 3' end processing by formation of regulatory complexes with basal pA factors. The effector domain of influenza A virus NS1 protein interacts with CPSF30 in influenza virus-infected cells and inhibits cellular pre-mRNAs 3' end cleavage and polyadenylation by preventing the binding of CPSF to the RNA. The RNA-binding activity of NS1 neither affects the interaction between the two factors, nor influences the function of NS1 in polyadenylation regulation or the ability of NS1 to block the accessibility of CPSF (131). The NS1 effector domain interacts also with PABP forming a NS1–CPSF–PABP trimeric complex. The physical association of NS1 with PABP leads to polyadenylation inhibition of cellular pre-mRNAs that escaped cleavage inhibition by NS1. This may occur by blocking the interaction between PABP and PAP that is required for processive addition of A residues. Consequently, via the two-pronged attack against CPSF and PABP, the NS1 protein blocks 3' end processing of cellular pre-mRNAs in infected cells. An additional consequence of the interaction with PABP is to inhibit its nuclear-cytoplasmic shuttling, thus allowing NS1 to control not only the synthesis of mature transcripts but also their export to the cytoplasm (132). In other types of viral infection, the redistribution of factors in RNA processing complexes could be involved in the polyadenylation-mediated switch from early to late gene expression. Indeed, HSV1 infection causes an increase in ICP27/IE63 expression and a concomitant reorganization of splicing components, but not of pA factors, at the site of transcription. Both events could be responsible of an enhanced protein binding, including CstF, to weak poly(A) sites of late genes and as a consequences, increase processing at these sites (133).

The association between BRCA1-associated protein BARD1 and CstF50 provides an important link between 3' end processing regulation by this mechanism and DNA damage. This interaction requires the linker between the ankyrin and BRCT domains of BARD1 (134). BARD1, like CstF50 (9), interacts with the pol II CTD while the protein partner of BARD1, BRCA1, co-immunoprecipitates with both CstF64 and BARD1, suggesting the formation of a BRCA1–BARD1–CstF trimeric complex connected to the pol II holoenzyme. BARD1, probably associated with BRCA1, inhibits mRNA 3' end formation in vitro functional assays (135). DNA-damage-inducing agents promote BARD1 phosphorylation (136) and increase the formation of the BRCA1–BARD1–CstF trimeric complex. This in turn leads to BARD1-dependent inhibition of 3' end processing (137). On the basis of the properties of BARD1, BRCA1 and CstF, it has been proposed that BARD1, in conjunction with pol II, senses the sites of DNA damage and the inhibitory function of the trimeric complex ensures that nascent RNAs are not erroneously polyadenylated (137).

Positive regulation of 3' end formation by this mechanism is exemplified by the tumor suppressor Cdc73. This factor associates with the cleavage/polyadenylation machinery (34) and in particular, with the CPSF-CstF complex (138). This physical interaction is necessary for in vitro 3' end processing of model substrates and is involved in the positive gene-specific regulation of polyadenylation in vivo. Cdc73 facilitates the binding of the two basal factors to actively transcribed chromatin regions. Since Cdc73 is a component of the pol II and of the chromatin-associated human Pat1 complex which orchestrates co-transcriptional histone modification, the functional association of Cdc73 with CstF and CPSF may help to coordinate transcription and RNA processing of specific genes (138).

The HSF1 transcription factor offers another example that underscores the importance of the functional interplay between the transcriptional and 3' end processing machineries. As described previously, in cells exposed to stress conditions the transcriptional factor HSF1 forms a complex with CstF64 and symplekin and this correlates with increased efficiency of Hsp70 mRNA polyadenylation (128). The underlying mechanism could be similar...
to that described for the transcription factor TFIIID which recruits CPSF to the promoter for the formation of mRNA 3' ends (6). Regulated loading of pA factors as a means to affect 3' end processing may be the mechanism used by Smicl, a CPSF-interacting protein that translocates from the cytoplasm to the nucleus at the midblastula transition in Xenopus. Smicl (Smad-interacting CPSF 30-like) is required for phosphorylation at serine 2 of pol II and regulates 3' end processing probably by allowing docking of proteins required for transcription and cleavage/polyadenylation (139). These and other molecular interactions may contribute to the physical tethering between both ends of the gene observed in yeast and mammalian cells (140).

Redistribution of pA factors by 3' end processing regulators can also lead to inclusion of additional pA factors. The nuclear phosphoinositide signaling pathway involves the nuclear type I phosphatidylinositol 4-phosphate 5-kinases (PIPKIz) and Star-PAP, a novel enzyme which possesses PAP activity. The two enzymes co-localize at nuclear speckles, interact with each other and control the expression of genes involved in detoxification and/or oxidative stress response. The underlying mechanism of regulation by the oxidative stress involves an increased association of Star-PAP with PIPK1z and components of the polyadenylation machinery, and an improved Star-PAP enzymatic activity leading to a rapid initiation of 3' end formation of this class of genes (141).

Post-translational modification of 3' end processing factors and regulators

Post-translational modification is emerging as an important mechanism that contributes to 3' end processing regulation. It affects the activity of both basal components of the 3' end processing machinery and regulatory factors. It contributes to modulate the activity, the nucleus-cytoplasm partitioning, the stability of the basal factors and their ability to interact within the core machinery and with regulatory factors.

Post-translational modification of 3' end processing factors offers an additional layer of 3' end processing regulation (Figure 3H). Every component of the cleavage/polyadenylation machinery, though not every subunit, is affected by this modification, including methylation, sumoylation, acetylation and phosphorylation (142). The current knowledge on the functional consequences of these modifications focuses mainly on PAP. Phosphorylation of this enzyme occurs throughout the cell cycle by the cdc2-cyclinB kinase (143,144); however, in the M-phase, hyperphosphorylation of PAP results in the downregulation of its activity, which in turn appears to be important for normal cell growth (143,145). This regulatory mechanism is inactivated by the HIV-1 Vpr accessory protein which blocks the activity of the cdc2-cyclinB complex, leading to PAP hyperactivity and contributing to HIV-1 pathogenesis (146). Phosphorylation of PAP can also be a requirement to interact with a regulatory protein, as shown for 14-3-3e, a protein factor which influences both PAP localization and activity. Hyperphosphorylation of PAP is not relevant for this interaction, suggesting that the PAP-14-3-3e association may function in phases of the cell cycle other than the M-phase (126). PAP is also subject to acetylation in its C-terminal region but this modification does not modify its polyadenylation activity. Instead, acetylated PAP loses its ability to associate with CFIm25, which is also acetylated via an interaction between the 68-kDa subunit of CFIm and the CBP (CREB-binding protein) acetyltransferase. Acetylation of PAP inhibits also the nuclear localization of PAP by inhibiting the binding to the importin α/β complex. It has been proposed that this post-translational modification plays a role in the reversible assembly of the 3' end processing complexes (147).

Sumoylation of PAP C-terminus was also found to have multiple effects on PAP function. In particular, sumoylation increases the nuclear localization and the protein stability of PAP but attenuates its enzymatic activity (148). Sumoylation targets other factors of the 3' end processing machinery, specifically CPSF73 and symplekin, thus suggesting that it plays an important role in 3' end processing complex formation implying a regulation at different levels and multiple protein factors (149). As demonstrated for sumoylated PAP (148), post-translational modification can regulate the cellular availability of factors which are required to accomplish the 3' end processing reaction. During the cell cycle, replication-dependent histone pre-mRNA 3' end processing plays a critical role in regulating histone mRNAs levels (150,40) and much of this regulation depends on the levels of SLBP (151). Degradation of SLBP by the proteasome at the end of the S-phase is activated by the sequential action of two kinases, first the cyclin A/Cdk1, which is activated near the end of S-phase, and then the CK2 (152). A different mechanism of regulation of the levels of pA factors is used during the picornavirus infection. This consists in the proteolytic degradation of CstF64 by the picornavirus 3C protease during the viral infection and leads to inhibition of host 3' end processing. The consequence of this regulation is to produce less cellular transcripts and to force the cellular machineries to process and express the viral RNAs (153).

Recent reports highlight the importance of phosphorylation for the regulatory function of auxiliary factors, regardless of the specific molecular mechanism used (Figure 3H). Stress-induced phosphorylation of IRBIT appears as the major determinant of its interaction with both hFip1 and PAP. In addition, phosphorylation of IRBIT controls the ability of this factor to translocate hFip1 to the nucleus and to inhibit polyadenylation (130). DNA-damage activated ataxia-telangiectasia mutated (ATM) kinase phosphorylates BARD1 in the presence of BRCA1 and this modification is relevant for BARD1 function in mediating repression of polyadenylation and pol II degradation. Importantly, mutation of BARD1 Tyr734 affects not only DNA-damage functions but also the ability of BARD1 to interact with CstF, suggesting that formation of the CstF–BARD1 complex plays an important role in the genotoxic stress-activated BARD1 functions (136). Another interesting example is Npl3, a yeast protein that competes with Rna15 for...
binding to a polyadenylation precursor and inhibits cleavage and polyadenylation in vitro (79). Npl3 plays also a function in stimulation of the transcriptional elongation activity by interacting with the pol II CTD. Casein kinase 2 (CK2) was found to be required for the phosphorylation of Npl3 and to reduce the ability of Npl3 to compete with Rna15 for binding to poly(A) signals and to interact with the CTD. This study suggests that phosphorylation of Npl3 promotes its dissociation from the mRNA/pol II, and contributes to the association of the polya/termination factor Rna15 (80). Increased levels of Npl3 also result in a negative-feedback loop in which phosphorylation of Npl3, probably by a different kinase, suppresses efficient recognition of the productive processing signals in its own transcript (154).

**PERSPECTIVES**

Studies over the past 20 years contributed to better characterize the constitutive factors of the basal eukaryotic 3′ end processing machineries (depicted in Figure 2) and to identify many other factors that modulate the efficiency and specificity of poly(A) signal recognition by these machineries (Table 1). For some of these factors, the mechanisms underlying the regulatory function have been described and overlapping models of regulation can be proposed (Figure 3 and Supplementary Table S1). Regulation of 3′ end processing efficiency and alternative 3′ end processing (Figure 1) are increasingly considered as important steps in gene regulation. Several examples highlight the essential contribution of 3′ end processing in physiological (e.g. immunity and inflammation) or pathological processes (e.g. cancer and viral infection) (25). Furthermore, evidences for global regulation of alternative 3′ end processing have recently been illustrated. Strikingly, alternative 3′ UTR events show an even higher frequency of tissue-specific regulation than other forms of alternative splicing events (155). A global change in 3′ UTR length, very likely by means of alternative 3′ end processing, occurs during T cell activation (156), neuronal activation (156), embryonic development (157), spermatogenesis (121) or oncogene activation (158). A possible scenario has been proposed in which some trans-acting factors act globally, some act tissue specifically, and some act gene specifically, with the combinatorial expression of all the different trans-acting factors determining the probability of using each proximal poly(A) signal (158). Since mechanistic details, that may help in both the understanding of gene expression regulation and the management of the various diseases in which 3′ end processing is associated (25), are often lacking, future work should be directed toward providing novel examples to feed the mechanistic models of mRNA 3′ end formation proposed here or novel paradigms of pre-mRNA 3′ end processing in eukaryotes.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We thank A. Decorsiere for critical reading of the manuscript and the referees for helpful comments.

**FUNDING**

INSERM; Institut Claudius Regaud; and FRM (Equipe FRM, soutenue par la Fondation Recherche Medicale). Funding for open access charges: INSERM; Institut Claudius Regaud and FRM.

**Conflict of interest statement.** None declared.

**REFERENCES**


