Polyadenylation site selection: linking transcription and RNA processing via a conserved carboxy-terminal domain (CTD)-interacting protein

Marc Larochelle¹ · Judit Hunyadkürti¹ · François Bachand¹

Introduction

Polyadenylation is a cotranscriptional process that occurs at the 3' end of almost every eukaryotic mRNAs. Polyadenylation starts by the endonucleolytic cleavage of a pre-mRNA, generating a free hydroxyl group at the 3' end of the upstream cleavage product to which adenosine monophosphate (AMP) are successively added by poly(A) polymerases. Recent studies in fact indicate that a large proportion of eukaryotic genes use multiple polyadenylation sites during mRNA synthesis (Derti et al. 2012; Hoque et al. 2013; Mata 2013; Ozsolak et al. 2010; Schlackow et al. 2013). This phenomenon, known as alternative polyadenylation (APA), has emerged as an important layer of gene regulation by generating mRNA isoforms that contain or lack sequences that control RNA localization, stability, and translation (Tian and Manley 2013). Post-transcriptional gene regulation via APA is exploited during several biological processes, including embryonic development, cellular differentiation and proliferation, as well as neuronal activation (Curinha et al. 2014; Elkon et al. 2013). Thus, coordination between transcription and poly(A) site selection must be tightly regulated, as disproportion between alternative polyadenylation isoforms are implicated in a number of human disorders (Erson-Bensan and Can 2016; Ogorodnikov et al. 2016).

Seb1: an RNA-binding protein that promotes co-transcriptional polyadenylation site selection

Recent work by our laboratory identified a new player in the control of polyadenylation site selection. Using the fission yeast Schizosaccharomyces pombe, we showed that the Seb1 protein promotes polyadenylation site selection of...
both coding and non-coding genes (Lemay et al. 2016). At the amino acid level, Seb1 shares extensive homology with the well-studied *Saccharomyces cerevisiae* protein, Nrd1, as well as with human SCAF4 and SCAF8 proteins, all containing a CTD interaction domain (CID) and a RNA recognition motif (RRM) (Corden 2013; Meinhart and Cramer 2004). The CID is involved in binding the carboxy-terminal domain of the largest subunit of RNA polymerase II by recognizing a set of highly conserved heptad repeats of the consensus sequence Tyr<sup>1</sup>-Ser<sup>2</sup>-Pro<sup>3</sup>-Thr<sup>4</sup>-Ser<sup>5</sup>-Pro<sup>6</sup>-Ser<sup>7</sup> (Meinhart and Cramer 2004). Our analysis of Seb1-associated proteins by mass spectrometry identified a number of pre-mRNA 3′ end processing factors. Consistent with this, we found Seb1 to be strongly enriched at the 3′ end of both mRNA- and snoRNA-coding genes using chromatin immunoprecipitation assays coupled to high-throughput sequencing (ChIP-seq). The functional importance of Seb1 in poly(A) site selection was revealed by a global change in 3′ untranscribed region (UTR) length caused by a Seb1 deficiency, with a majority of affected genes showing a longer 3′ UTR, whereas the remaining genes displayed shorter 3′ UTRs. Interestingly, analysis of Seb1 domains indicated that both the CID and the RRM domains were required for poly(A) site selection and recruitment of Seb1 to the 3′ end of genes, suggesting a cooperative contribution of both CTD recognition and RNA binding in Seb1-dependent cotranscriptional 3′ end processing. Mapping of Seb1-RNA interactions at the transcriptome-wide level revealed that the majority of Seb1 RNA-binding mapped downstream of mRNA 3′ end cleavage sites and allowed the identification of a Seb1-recognition motifs consisting of a core GUA trinucleotide surrounded by A/U and A/G as 5′ and 3′ nucleotides, respectively. We also showed that binding of Seb1 to its consensus motif is critical for poly(A) site selection. Although our results showed that Seb1 is required for the recruitment of specific cleavage and polyadenylation factors, it appears that Seb1 is not essential for the 3′ end endonucleolytic cleavage reaction. First, binding of the fission yeast homolog of the human CPSF73 endonuclease, Ysh1, at the 3′ end of genes was not affected in cells deficient for Seb1. Second our analyses of RNA-seq data showed that mRNA abundance did not markedly change in a seb1 mutant. Collectively, our findings reveal that regulation of 3′ UTR length in fission yeast is a cotranscriptional process controlled by the recruitment of the Seb1 RNA-binding protein at the 3′ end of genes.

Yet, despite uncovering a role for the conserved Seb1 RNA-binding protein in cotranscriptional poly(A) site selection, many important questions remain to be elucidated. In this review, we will discuss some of the key points that need to be addressed for a better understanding of the molecular mechanism by which Seb1 controls cleavage site selection. How is Seb1 recruited to the 3′ end of genes? Can we observe redundancy between Seb1 and Pcf11, a component of the cleavage and polyadenylation complex that also contains a CID domain and that is involved in transcription termination? Can SCAF4 and SCAF8, putative homologue of *S. pombe* Seb1, promote polyadenylation site selection in human cells?

**Recruitment of Seb1 to the 3′ end of genes**

Analysis of Seb1 amino acid sequence indicates that it contains an N-terminal CID domain that exhibits homology to the CID domains of Nrd1, Pcf11, and SCAF4/8 proteins (Meinhart and Cramer 2004). Importantly, experimental evidence support that the CID domain of Seb1 is functionally relevant: (1) GST pull-down assays indicate that Seb1 associates with the CTD of *S. pombe* Rpb1 (Lemay et al. 2016); and (2) substitutions of conserved residues in the Seb1 CID that are key for CID-CTD recognition (Kubicek et al. 2012; Lunde et al. 2010; Meinhart and Cramer 2004) abolished Seb1 recruitment to the 3′ end of genes and impaired polyadenylation site selection (Lemay et al. 2016). Yet, as the RNAPII CTD is dynamically modified during the transcription cycle, the detailed mechanism of CTD recognition by Seb1 still remains to be determined. Each consensus heptad repeat of the RNAPII CTD can be subjected to post-translational modifications, including phosphorylation, methylation, acetylation, and ubiquitination (Hsin and Manley 2012). The most studied CTD modifications are phosphorylation events that target Ser<sup>2</sup> and Ser<sup>5</sup> of the consensus heptad repeat. A general model of CTD phosphorylation during the transcription cycle proposes that a hypo-phosphorylated form of RNAPII is recruited to promoters and becomes phosphorylated on Ser<sup>5</sup> during transcription initiation. As transcription elongation progresses towards the 3′ end of genes, the levels of Ser<sup>5</sup> phosphorylation decrease while the levels of Ser<sup>2</sup> phosphorylation gradually increase, reaching a peak near the poly(A) sites, beyond which the levels of Ser<sup>2</sup> phosphorylation decrease (Srivastava and Ahn 2015). Accordingly, functional links have been established between Ser<sup>2</sup> phosphorylation and 3′ end processing in both yeast and humans (Ahn et al. 2004; Davidson et al. 2014; Licatalosi et al. 2002).

Given the specific recruitment of Seb1 to the 3′ end of genes (Lemay et al. 2016), one reasonable prediction would be that Seb1 specifically recognizes a form of the CTD that is hyper-phosphorylated on Ser<sup>2</sup>, which is predominant at or near poly(A) sites. However, this simple correlation can be misleading, as it appears to be the case for budding yeast Pcf11. Accordingly, although Pcf11 levels concentrate at the 3′ end of genes, Pcf11-CTD interactions are not solely mediated by Ser<sup>2</sup> phosphorylation (Lunde et al. 2010).
fact, it was recently found that *pcf11* mutants with substitutions in the CID domain have reduced levels of CTD Ser\(^2\) phosphorylation (Grzechnik et al. 2015), suggesting that Pcf11-CTD interactions stimulate phosphorylation of CTD Ser\(^2\). It is also possible that Seb1 recruitment is influenced by CTD Tyr\(^1\) phosphorylation. Similar to Ser\(^2\) phosphorylation, levels of Tyr\(^1\) phosphorylation are low at gene promoters but increase as RNApolyII progresses towards the 3′ end of genes (Mayer et al. 2012). Yet, in comparison to phosphorylated Ser\(^2\), Tyr\(^1\) phosphorylation levels decrease before RNApolyII reaches the polyadenylation site. Importantly, it was found that Tyr\(^1\) phosphorylation impairs the recruitment of specific CID-containing proteins, including Pcf11 (Mayer et al. 2012), and that Tyr\(^1\) dephosphorylation at or near poly(A) sites produces a transcription window in which the ratio between Ser\(^2\) phosphorylation relative to unphosphorylated Tyr\(^1\) is highest, which has been proposed to facilitate binding of 3′ end processing and termination factors (Schreieck et al. 2014). Yet, phosphorylation dynamics of the RNApolyII CTD in *S. pombe* are poorly understood and are likely to be different than from *S. cerevisiae* (Schwer and Shuman 2011). Unlike *S. cerevisiae*, Ser2 phosphorylation in *S. pombe* is lethal (West and Corden 1995), the analogous Tyr-to-Phe CTD mutant is viable in *S. pombe* (Schwer and Shuman 2011), suggesting that Tyr\(^1\) phosphorylation is not required for growth in fission yeast. Alternatively, a CTD-independent recruitment cannot be excluded. Indeed, a recent structural analysis of *S. pombe* Lsg1, which is the homologue of *S. cerevisiae* Ctk3 that functions as a subunit of the CTD kinase CTDK-1, revealed a structure that closely resembles the CID domains found in the *S. cerevisiae* Rtt103, Pcf11, and Nrd1 as well as in human SCAF8 (Muhlbacher et al. 2015). However, a complete CTD-binding fold was not conserved in Lsg1, which explains its inability to bind to unphosphorylated or phosphorylated CTD-derived peptides. Therefore, the possibility exists that Seb1 might be recruited to the 3′ end of genes via a non-canonical CID domain that may promote different protein–protein interactions than the expected Seb1-CTD recognition.

**Interplay between Seb1 and Pcf11 in *S. pombe* 3′ end processing**

Pcf11 is a conserved subunit of the cleavage and polyadenylation machinery that plays key functions in coupling transcription with 3′ end processing via interaction with the CTD of RNA polymerase II (Barilla et al. 2001; Licatalosi et al. 2002; Sadowski et al. 2003). Pcf11 can also bind RNA (Hollingworth et al. 2006; Zhang et al. 2005), but does not appear to contain a conserved RRM domain. Interestingly, recent work from the group of Nicholas Proudfoot showed that, similar to *S. pombe* Seb1, *S. cerevisiae* Pcf11 is involved in transcription termination of coding and non-coding RNAs (Grzechnik et al. 2015). However, expression of *S. pombe* Seb1 into a *pcf11* mutant of *S. cerevisiae* did not rescue the growth defect phenotype of the *pcf11* mutant (Lemay et al. 2016), suggesting that Seb1 and Pcf11 contribute to different roles in 3′ end processing. Consistent with this view, the fission yeast genome contains an homologue of *S. cerevisiae* Pcf11. Similar to Seb1, *S. pombe* Pcf11 is also enriched at the 3′ end of genes (Lemieux and Bachand 2009). To begin to address the functional relationship between Seb1 and Pcf11 in 3′ end processing, we recently generated a strain in which *seb1* and *pcf11* alleles were both expressed under the control of the thiamine-sensitive *nmt1* promoter, as both genes are essential for cell viability. The simultaneous depletion of Seb1 and Pcf11 in *S. pombe* exacerbated the growth defect phenotypes seen in either single mutants (Fig. 1a). This observation was supported by RNA analyses, which revealed that the levels of *fba1* and *rps2* mRNAs in the double mutant were markedly decreased relative to either single mutant strains (Fig. 1b, compares lanes 4 to lanes 2–3). Interestingly, the depletion of Seb1 and Pcf11 resulted in different effects on mRNA expression: whereas the depletion of Seb1 caused the accumulation of 3′-extended transcripts (lane 2), a deficiency in Pcf11 resulted in reduced mRNA levels, but no detectable 3′ extensions (lane 3). Furthermore, despite reduced total RNA levels in the double mutant, the ratio between proximal and distal isoforms was unchanged between the single Seb1 depletion and double Seb1/Pcf11 depletion (compare lanes 2 and 4). These preliminary results suggest that Seb1 and Pcf11 contribute differently to mRNA 3′ end processing. Future studies will be required to dissect the functional interplay between Seb1 and Pcf11 proteins in *S. pombe* 3′ end processing of coding and non-coding genes.

**Human SCAF4 and SCAF8**

By using yeast two-hybrid screening with the mammalian CTD as a bait, Yuryev and coworkers identified a number of CTD-binding proteins (Yuryev et al. 1996). Two of the newly identified proteins contain a CID domain at the N-terminus, an RRM domain, and a serine/arginine(SR)-rich region positioned just in front of the RRM domain. These SR-like CTD-associated factors, SCAF4 and SCAF8, share common domain architecture and sequence homology with *S. cerevisiae* Nrd1 and *S. pombe* Seb1 (Lemay et al. 2016; Meinhardt and Cramer 2004; Yuryev et al. 1996). Whereas SCAF8 has been shown to bind phosphorylated CTD (Becker et al. 2008; Patturajan et al. 1998), little is known about SCAF4. RNAi-mediated depletion of
SCAF8 in a human cell line does not induce transcription termination defects at U1 and U2 snRNA genes as well as at the gene encoding β-actin mRNA (O’Reilly et al. 2014). Despite the fact that a deficiency in SCAF8 did not appear to cause defects in RNA processing, it should be noted that SCAF4 and SCAF8 are 43% identical (65% similar) at the amino acids level, and as such, that redundancy between these two closely related proteins might not be unexpected. It is also possible that SCAF4 and SCAF8 have gene and/or cell type-specific functions, such that regulation of the CTD post-translational modification status could influence the spatial and temporal recruitment of these factors. Indeed, specific RNAPII CTD modification patterns have been associated with environmental stresses, various types of cancers, neurological disorders, and the inflammatory response (Srividastava and Ahn 2015). Interestingly, a recent study identified SCAF4 and SCAF8 using a combination of proteomic and genomic screens designed to isolate factors with a role in the transcription-related DNA damage (Boeing et al. 2016). Although the mechanism by which SCAF4 and SCAF8 contribute to this pathway remains to be defined, gene regulation via APA could represent a mean by which these factors orchestrate a proper DNA damage response.

Conclusions

Seb1 is a fascinating protein that is necessary for polyadenylation site selection in fission yeast. On the basis of similarities identified between S. pombe and human polyadenylation signals, including use of the canonical AAUAAA hexamer (Mata 2013; Schlackow et al. 2013), we anticipate that an understanding of the mechanism by which Seb1 coordinates transcription and polyadenylation site selection will provide key insights into how alternative polyadenylation is controlled in human cells.

Acknowledgments

Work on polyadenylation site selection and transcription termination is supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to F.B. F.B. is the Canada Research Chair in Quality Control of Gene Expression.

References

Davidson L, Muniz L, West S (2014) 3′ End formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are