RNA Biology 6:5, 522-525; November/December 2009; © 2009 Landes Bioscience

An up-close look at the pre-mRNA 3'-end processing complex

Yongsheng Shi,* Serena Chan and Gabriel Martinez-Santibañez
Department of Microbiology and Molecular Genetics; University of California, Irvine; Irvine, CA USA

Polyadenylation is a nearly universal mRNA processing step in eukaryotic gene expression and it takes place in a macromolecular machinery termed the mRNA 3' processing complex. In a recent study, we reported the purification and comprehensive characterization of functional mammalian mRNA 3' processing complexes. Our study defined the protein composition of this machinery, revealed new potential links between mRNA 3' processing and other cellular processes, and characterized basic structural features of the "core" mRNA 3' processing complex. These results provided new insights into the mechanism of mRNA 3'-end formation.

A general mechanistic framework for mRNA 3' processing has emerged from a large body of biochemical and genetic studies in the last two decades (reviewed in refs. 1–3). In mammalian mRNAs, polyadenylation sites are mainly defined by an upstream AAUAAA element and a downstream G/U-rich element. Five trans-acting factors are required for 3' processing, including the poly(A) polymerase (PAP) and the multi-subunit protein complexes CPSF, CstF, CF Im and CF IIm. CPSF specifically recognizes the AAUAAA element at least in part through CPSF160. CstF binds to the downstream G/U-rich sequence via CstF64. CF Im is involved in additional RNA binding and contributes to specific recognition of poly(A) signals by the polyadenylation machinery. The functions of CF IIm in mRNA 3' processing remain poorly understood. These factors assemble onto mRNA polyadenylation signals to form the 3' processing complex. Within this macromolecular machinery, 3' processing takes place in two steps, an endonucleolytic cleavage catalyzed by CPSF73 followed by the addition of a poly(A) tail by PAP. Given the fundamental importance of mRNA 3' processing in eukaryotic gene expression, it is not surprising that the majority of the mRNA 3' processing factors are conserved from yeast to human despite the significant divergence between the mRNA polyadenylation signals in yeast and those of mammalian mRNAs.[1-3] Interestingly, however, at least 5 more mRNA 3' processing factors have been identified in yeast than in mammals,[2] raising the possibility that more factors remains to be discovered in the mammalian system.

Mammalian mRNA 3' processing has been studied mostly by biochemical purification and characterization of individual factors.[1,2] Although this reductionist approach will remain important in future studies, unique insights may be obtained by characterizing the mRNA 3' processing complex itself, including its composition, structure and dynamics. By combining glycerol gradient sedimentation and RNA tag-based affinity purification, we have recently purified the human mRNA 3' processing complex in its functional form and carried out proteomic, functional and structural analyses.[4] The major findings of these studies and their implications are discussed below.

Not All 3' Processing Factors are Created Equal

In our recent report,[4] we purified and carried out proteomic analysis of the mRNA 3' processing complexes at the "post-assembly" stage, during which the complex has been assembled but very little processing has occurred. In contrast to the current

Key words: mRNA 3' processing, cleavage, polyadenylation

Abbreviations: EM, electron microscopy

Submitted: 05/08/09

Accepted: 07/18/09

Previously published online:
www.landesbioscience.com/journals/rnabiology/article/9554

*Correspondence to:
Yongsheng Shi; Email: yongshes@uci.edu

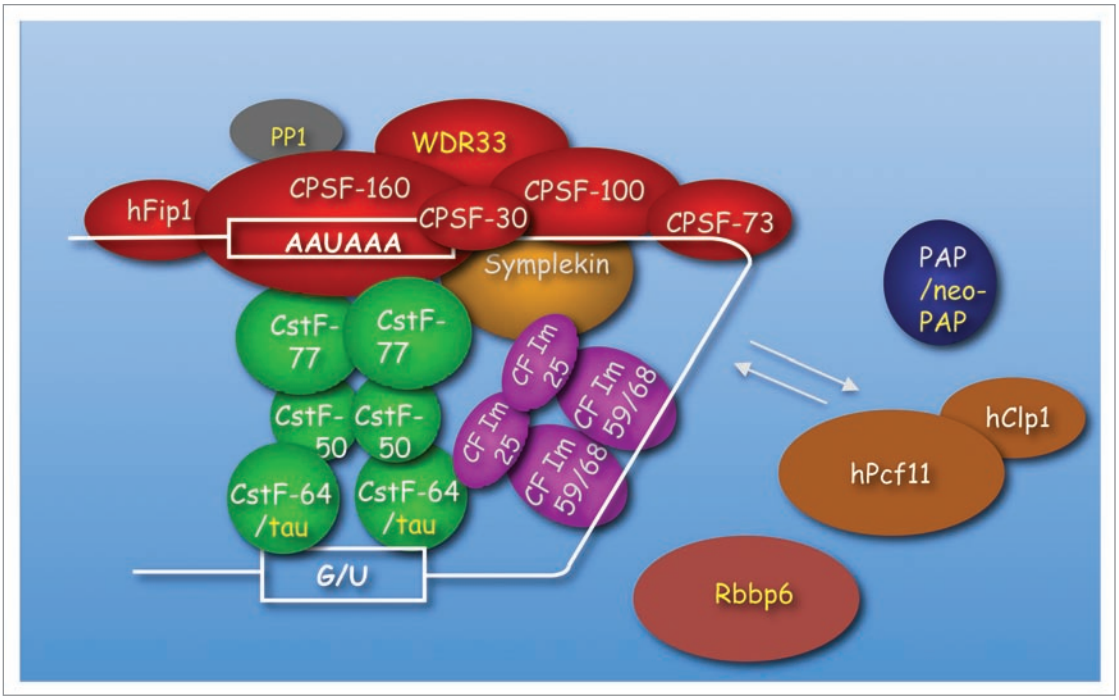

Figure 1. A new model for the mammalian mRNA 3' processing complex. This model incorporates results from recent studies (discussed in the text) and has two new features. (1) CPSF, CstF, and CF Im form the "core" mRNA 3' processing complex. CF IIm, PAP, and other potential mRNA 3' processing factors such as Rbbp6 are transiently and/or weakly associated with the core complex. (2) Both CstF and CF Im function as dimers in the complex and, as a result, each of these subcomplexes contributes two RNA-binding subunits.

model of the 3' processing complex,[1,2] it is clear from our proteomic analyses that not all 3' processing factors are equally represented in the 3' processing complex. First, although CPSF, CstF and CF Im were all detected at high levels, the CF IIm components were either barely detectable (hPcf11) or completely missing (hClp1).[4] This suggests that CPSF, CstF and CF Im form a stable "core" mRNA 3' processing complex, in which specific interactions between the polyadenylation signal in the mRNA and the mRNA 3' processing machinery have been established. CF IIm, however, associates with the "core" complex only weakly and/or transiently, indicating that it may primarily function in catalysis. Secondly, PAP is absent from the "post-assembly" mRNA 3' processing complex[4]. Instead, the related neo-PAP was detected at low levels.[4] These observations raise a couple of important questions: (1) How and when is PAP recruited to the 3' processing complex? Previous studies suggest that PAP is recruited to the 3' processing machinery via direct interaction with CPSF160.[5] However, we failed to detect PAP in our proteomic analysis of the CPSF complex purified under relatively mild conditions by immunoprecipitation.[4] These data suggest that PAP is not tightly associated with CPSF and is perhaps not stably recruited to the mRNA 3' processing complex until later stages. (2) Is PAP the primary poly(A) polymerase in mammalian cells?[1-3] As the canonical PAP and neo-PAP display essentially identical activities in vitro,[6] they may play redundant roles in vivo. An alternative but mutually non-exclusive possibility is that each PAP may be responsible for processing a specific subset of mRNAs. Further studies are necessary to discern these scenarios.

Even within the "core" 3' processing complex, the stoichiometry of CPSF, CstF and CF Im is almost certainly not 1:1:1 as originally thought.[1,2] First, our study provided indirect evidence that CstF functions as a dimer.[4] For example, when the purified mRNA 3' processing complexes were resolved by SDS-PAGE and stained by silver, the CstF77 band was consistently and significantly more intense than the CPSF73 band.[4] Furthermore, close to twice as many unique peptides were detected for CstF77 than CPSF73 despite their similar sizes.[4] These observations, although not strictly quantitative, are consistent with the notion that CstF forms a dimer in the 3' processing complex. This conclusion is consistent with earlier biochemical studies showing that CstF77 and CstF50 can both associate with themselves[7] and with the recently solved X-ray structure of CstF77,[8] showing that the HAT domain of CstF77 self-dimerizes.[3,8] Secondly, we have obtained evidence that strongly suggests that CF Im also forms a dimer (unpublished data), a conclusion that is again supported by recent structural studies demonstrating that the 25 kDa subunit of CF Im forms a homodimer.[9,10] Based on our results and other recent studies, we propose a new model for the 3' processing complex (**Fig. 1**), in which CPSF, CstF and CF Im form the "core" 3' processing complex at a 1:2:2 stoichiometry and CF IIm dynamically associates with the "core" complex. PAP or neo-PAP is recruited to the complex at a later stage.

These results have direct implications on the mechanisms of mRNA 3' processing.

The CF Im monomer contains either CF Im59 or 68, both contain RRM (RNA Recognition Motif) domains.[1,2] Similarly, CstF contains the RNA-binding proteins CstF64 or the related CstF64τ.[1,2,4] Thus, dimerization of the CF Im and CstF can lead to significantly more potential RNA-protein contacts within the 3' processing complex than previously thought, which may be necessary to achieve high RNA-binding affinity as well as specificity. Furthermore, the alternative RNA-binding subunits in CstF and CF Im may have distinct RNA-binding specificities.[11] Thus different combination of these factors achieved through dimerization may recognize distinct polyadenylation signals. Broader sequence specificity may be advantageous for CstF and CF Im as their respective targets, the G/U-rich elements and the auxiliary elements, are both highly divergent.[1,2]

Function by Association: New Mammalian mRNA 3' Processing Factors

One of our goals for characterizing the mRNA 3' processing complex was to identify new mammalian 3' processing factors. In addition to known 3' processing factors, our proteomic analyses of the mRNA 3' processing complex detected over 50 other proteins.[4] First, we identified several factors that are the (putative) homologues of essential yeast polyadenylation factors, including Wdr33, the phosphatase PP1, and Rbbp6/PACT. The subsequent functional studies showed that Wdr33 and PP1 are associated with CPSF and play essential roles in mammalian mRNA 3' processing. PP1, like its yeast counterpart Glc7p, is specifically required for polyadenylation, but not for cleavage. Interestingly, Rbbp6/PACT, homologue of the yeast polyadenylation factor Mpe1, was originally identified as an Rb- and p53-binding protein,[12,13] and was recently shown to regulate p53 stability through modulating its ubiquitination.[14] Thus Rbbp6 provides an intriguing link between mRNA 3' processing and the cell cycle and the p53 pathway. Secondly, our study identified CstF64τ as a stable component of the CstF and the mRNA 3' processing complex.[4] CstF64τ is a paralogous variant of CstF64 that is reportedly expressed only in the testis.[15] Our results indicate that CstF64τ is much more widely expressed (unpublished data) and a putative fission yeast homologue was identified.[4] These results indicate that the mRNA 3' processing machinery is even more conserved than previously appreciated. Finally, many factors detected in the mRNA 3' processing complex have putative or known functions outside polyadenylation,[4] and we believe that at least some of them may mediate crosstalk between mRNA 3' processing and a wide range of other cellular processes, including transcription, splicing, RNA turnover, translation, and the DNA damage response. Sorting out the extensive coupling between mRNA 3' processing and other cellular processes will be a major challenge for future studies, but a necessary step toward understanding gene expression and its regulation at the systems level.

A (EM) Picture is Better than a Thousand Words: Structural Studies

Detailed structural information is essential for understanding the mechanisms of gene expression at the molecular level, as exemplified by the studies of RNAP II[16] and the ribosomes.[17] Although crystal structures have been obtained for a number of individual mRNA 3' processing factors (reviewed in ref. 3), higher order structures are necessary for understanding how mRNA 3' processing factors are spatially organized to carry out RNA recognition and catalysis. Structural analysis of the entire mRNA 3' processing complex, however, faces a number of obstacles. The quantities of the purified mRNA 3' processing complex that can be currently obtained are still far below what is needed for crystallization trials. And the potentially dynamic nature of the complex poses additional challenges. Under these circumstances, single particle cryo-electron microscopy (cryo-EM) analysis[18,19] is the method of choice for structural analysis of the mRNA 3' processing complex. In this method, samples are flash frozen in a thin layer of vitrified ice and subsequently subject to EM and computational analysis for reconstruction of a 3-D model. It requires minimal amount of materials and the samples remain hydrated in their native state. With recent advances in specimen preparation and computational techniques, increasingly high resolution (up to ~4 Å) structures has been obtained for a wide variety of macromolecular complexes.[19] As an initial step, we visualized the negative stained mRNA 3' processing complexes using EM and carried out single-particle analyses.[4] The results of these experiments provided the first glimpse of the mRNA 3' processing complex, a ~250 Å long "kidney"-shaped particle. Given the relatively simple biochemical nature of mRNA 3' processing, it is surprising that the size of the mRNA 3' processing complex rivals that of the bacterial ribosome large subunit and the mammalian spliceosomal A complex.[4] The relative homogeneous nature and the large size of the purified mRNA 3' processing complexes make it suitable for cryo-EM analysis in the future.

What's Next?

The availability of highly purified mRNA 3' processing complex and its subcomplexes makes it possible to address a number of fundamental and long-standing questions in the field. First, how are the polyadenylation signals specifically recognized by the mRNA 3' processing machinery? Although CPSF160 and CstF64 have been shown to bind key cis-elements,[1,2] a more extensive RNA-protein interaction network is almost certainly required for specific recognition of polyadenylation signals by the mRNA 3' processing machinery. A systematic characterization of RNA-protein interactions within the mRNA 3'processing complex will reveal such a network. Secondly, it will be of great interest to characterize the mRNA 3' processing complexes at later stages, such as the post-cleavage complexes as well as the post-polyadenylation mRNPs. These studies will provide new insights into the dynamic nature of the mRNA 3' processing complexes and help address a number of critical mechanistic questions. For examples, is there a conformational and/or compositional change between the cleavage and polyadenylation steps? How is PAP recruited to the mRNA 3' processing complex and how is the timing of its

recruitment controlled? How does mRNA 3' processing influence later steps of gene expression? Finally, although our initial structural characterization of the mRNA 3' processing complex provided a tantalizing first glimpse of its general features, single particle cryo-EM analysis is clearly the next step toward building a high-resolution 3-D model and mapping its key components. In the coming years, as more of the delicate inner workings of the mRNA 3' processing complex are teased out, we will be able to better understand how this fascinating molecular machine really works.

Acknowledgements

I would like to acknowledge Dr. James Manley for his generous and continued support. Research in my laboratory is supported by a start-up fund from the University of California, Irvine.

References

1. Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. Genes Dev 1997; 11:2755-66.
2. Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev 1999; 63:405-45.
3. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci 2008; 65:1099-122.
4. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, et al. Molecular architecture of the human pre-mRNA 3' processing complex. Mol Cell 2009; 33:365-76.
5. Murthy KG, Manley JL. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes Dev 1995; 9:2672-83.
6. Topalian SL, Kaneko S, Gonzales MI, Bond GL, Ward Y, Manley JL. Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. Mol Cell Biol 2001; 21:5614-23.
7. Takagaki Y, Manley JL. Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol Cell Biol 2000; 20:1515-25.
8. Bai Y, Auperin TC, Chou CY, Chang GG, Manley JL, Tong L. Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. Mol Cell 2007; 25:863-75.
9. Coseno M, Martin G, Berger C, Gilmartin G, Keller W, Doublié S. Crystal structure of the 25 kDa subunit of human cleavage factor Im. Nucleic Acids Res 2008; 36:3474-83.
10. Tresaugues L, et al. The crystal structure of human cleavage and polyadenylation specific factor-5 reveals a dimeric Nudix protein with a conserved catalytic site. Proteins 2008; 73:1047-52.
11. Monarez RR, MacDonald CC, Dass B. Polyadenylation proteins CstF-64 and tauCstF-64 exhibit differential binding affinities for RNA polymers. Biochem J 2007; 401:651-8.
12. Sakai Y, Saijo M, Coelho K, Kishino T, Niikawa N, Taya Y. cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. Genomics 1995; 30:98-101.
13. Simons A, Melamed-Bessudo C, Wolkowicz R, Sperling J, Sperling R, Eisenbach L, et al. PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb. Oncogene 1997; 14:145-55.
14. Li L, Deng B, Xing G, Teng Y, Tian C, Cheng X, et al. PACT is a negative regulator of p53 and essential for cell growth and embryonic development. Proc Natl Acad Sci USA 2007; 104:7951-6.
15. Wallace AM, Dass B, Ravnik SE, Tonk V, Jenkins NA, Gilbert DJ, et al. Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. Proc Natl Acad Sci USA 1999; 96:6763-8.
16. Kornberg RD. The molecular basis of eukaryotic transcription. Proc Natl Acad Sci USA 2007; 104:12955-61.
17. Steitz TA. A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol 2008; 9:242-53.
18. Frank J. Single-particle imaging of macromolecules by cryo-electron microscopy. Annu Rev Biophys Biomol Struct 2002; 31:303-19.
19. Stark H, Luhrmann R. Cryo-electron microscopy of spliceosomal components. Annu Rev Biophys Biomol Struct 2006; 35:435-57.