



HAL
open science

Structural insights into the 3'-end mRNA maturation machinery: Snapshot on polyadenylation signal recognition

Stéphane Thore, Sébastien Fribourg

► **To cite this version:**

Stéphane Thore, Sébastien Fribourg. Structural insights into the 3'-end mRNA maturation machinery: Snapshot on polyadenylation signal recognition. *Biochimie*, 2019, 164, pp.105-110. 10.1016/j.biochi.2019.03.016 . hal-02472477

HAL Id: hal-02472477

<https://hal.science/hal-02472477>

Submitted on 20 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Title

Structural insights into the 3'-end mRNA maturation machinery: snapshot on Polyadenylation signal recognition

Stéphane Thore & Sébastien Fribourg*

INSERM U1212 – CNRS 5320 & Université de Bordeaux, 146 rue Léo Saignat, 33000 Bordeaux , France.

Correspondence should be addressed to :

e-mail: sebastien.fribourg@inserm.fr

tel: +33 (0)5 57 57 15 11

Abstract

Pre-mRNA 3'-end maturation is achieved **by** a mechanism requiring four different protein complexes **assembled from** approximately twenty factors. A global understanding of this essential process is still missing due to the inability to structurally characterize the **entire** complexes, even **though structures of the isolated** factors have been obtained. In this review, we summarize recent findings regarding the atomic description of one the major players, the Cleavage and Polyadenylation Specificity Factor complex (CPSF in human, CPF in yeast). These data provide information on the architecture adopted by **the** major components **of this complex**, and on its capacity to recognize the polyadenylation signal sequence.

Keywords: CPSF, polyadenylation

1. Introduction

As a gene is transcribed by RNA polymerase II (Pol II), the pre-mRNA undergoes multiple co-transcriptional processing steps to produce mature mRNA, including 5'-end capping, splicing, and 3'-end polyadenylation [1,2]. Most eukaryotic mRNA, with the exception of histone mRNA [3], follows a two-step cleavage and polyadenylation reaction for their 3'-end maturation [4]. Upon completion, the mRNA 3'-end carries a **polyadenylate tail with an average length of** 60 nucleotides in yeast up to 250 nucleotides in human cells [5].

The functional division of the 3'-end processing machinery into various sub-complexes dates from the late 1980s. At that time, a minimal set of complexes allowing efficient recapitulation of pre-mRNA 3'-end maturation *in vitro* were described. In human, four complexes were identified: Cleavage Factor I and II (CFIm and CFII_m), the Cleavage Stimulation Factor (CstF), and the Cleavage and Polyadenylation Specificity Factor (CPSF)[6–8]. Similar observations were **later** made in yeast showing that the mechanism relies on major complexes named: Cleavage Factor IA and IB (CF IA and CF IB /Hrp1), Cleavage Factor II (CF II), **and Polyadenylation Factor (PF) [10–13]. The composition of CF IA and CF IB was clearly defined, but the composition of (yeast) CF Ily and PF were more elusive. In the following decades, the precise protein composition of these complexes has been identified and showed a strong conservation of the yeast and human core subunits (Table I)[2,14–16]. These studies also demonstrated that CF Ily and PF were in fact part of a larger CPF complex [12].**

These complexes bind to the pre-mRNA by recognizing short RNA sequences. The CFIm in human or Hrp1 in yeast recognize a UGUA sequence or an AU-rich sequence respectively [17–20]. **CFIm and Hrp1 increase the precision of the cleavage site, and influence alternative polyadenylation (APA). However,** both CPSF and CPF are also able to recognize the polyadenylation site (PAS) independently of any other factors (Table I) [21,22]. Two other complexes, CstF and CFII_m, are thought to provide an additional hook onto the pre-mRNA to **improve efficiency [23–25](Figure 1).** CstF and CFII_m **also** regulate the coupling of **polyadenylation** to RNA polymerase II transcription termination [26]. Within the major CPSF **complex,** the CPSF73 protein carries the endonuclease activity **required** to cleave the pre-mRNA [27], while the poly(A)-polymerase (PAP) synthesizes the poly(A) tail.

From a structural perspective, the first data on the 3'-end processing factors came from the crystal structures of the **isolated** yeast and mammalian poly(A)-polymerases [28,29]. **These were followed by structures of** by their cognate complexes with RNA and hFip1 [30,31]. **Soon after, the** structures of CstF/CF IA and CFIm **provided insights into their** architecture, **including** the dimerisation of subunits such as CstF77, CstF50 and CFIm25 [32–37]. Also **available at this time were** structures of **the** isolated domains [38], **as well as** the structural basis for RNA recognition [19,39,40]. Altogether, these data provided a **partial** picture at the atomic scale of some pre-mRNA 3'-end processing factors.

The most important **complex is** CPSF, which **although containing** the cleavage and polyadenylation activities **remained** the least described for many years. **Nevertheless,** the structures of some individual subunits of CPSF were solved in the past decade revealing **details of** the enzymatic center carried by the CPSF73 protein, and the strong similarity between CPSF73 and CPSF100 [27]. Additional information came with the atomic details of the **interaction between the phosphorylated** RNA polymerase II Carboxy Terminal Domain (CTD) with **the CPSF-associated subunit** Ssu72 [41,42]. **More recently, the interaction between Ysh1 and Mpe1 was revealed** [43].

A bigger picture of **the** essential CPSF complex was **finally provided by** several **recent** independent studies, describing both the human and the yeast complexes [44–47]. In this review, we will focus on these recent **advances on the organization and architecture of** CPSF (and CPF).

2. CPSF modular organization

Mass spectrometry analysis using **either** native (noncovalent) nano-electrospray, or **with** disuccinimidyl suberate (DSS)-based cross-linking experiments, allowed a more refined view of the CPSF and CPF organization [44,46]. Native mass spectrometry analysis performed on purified yeast CPF complexes demonstrated the existence of three modules **that were named** the Nuclease, the Poly(A) polymerase and the Phosphatase module (Table I)[44]. Each module carries one of the three enzymatic activities required for efficient pre-mRNA 3'-end processing. The Phosphatase module is composed of **six** proteins in yeast, out of which four have a **human** homologue (i.e. **Pta1 (Symplekin), SSU72 (hSsu72), Swd2 (WDR82) and Glc7 (PP1a/b)**(Table 1)[48]. **The phosphatase module added to Syc1 forms the Associated**

with Pta1 complex (APT) involved in non-coding RNA transcription, a function that is independent of the 3'-end pre-mRNA processing [48]. The Nuclease module is made of three proteins: a dimer of Ysh1 (CPSF73) and Ydh1 (CPSF100), plus the Mpe1 protein (RBBP6)[49]. The Poly(A) polymerase module contains the Yhh1 (CPSF160) and Yth1 (CPSF30) proteins, Pfs2 (WDR33), Fip1 (hFip1), and the Poly(A) polymerase. This modular organization had been proposed for the nuclease part of the mammalian machinery specifically responsible for histone pre-mRNA 3'-end processing [50]. Another study of the homologous human complexes [45] is consistent with the biochemical definition of the Poly(A) polymerase module [21]. In addition, the latter study describes an extensive combinatorial analysis using different length constructs of four subunits from the CPSF, namely CPSF160, CPSF30, hFip1 and WDR33. Despite the existence of a new cross-link-based contact map, the atomic picture of the pre-mRNA cleavage and polyadenylation machinery was still missing. This overall atomic organization of CPSF/CPF finally found a first solution in a series of articles in which atomic models for the newly-defined yeast and human Poly(A) polymerase modules of CPSF were obtained, both in the RNA-free form and bound to the highly conserved polyadenylation signal [44–47].

3. Overall description and comparison

Numerous technical obstacles have hindered the search for an integrated description of CPSF/CPF, in particular difficulties in the recombinant production of well-behaving complexes. The identification of functional modules within CPF has helped in their recombinant production. Subsequently, several core component structures of the Poly(A) polymerase module were obtained using either cryo-electron microscopy or X-ray crystallography. Both techniques lead to models at resolutions ranging from 2.5 Å to 3.5 Å (Figure 2)[44–47]. The structures of homologous subunits from the human and the yeast complexes showed a high degree of similarities with a global root mean square deviation of 2.8 Å over more than 7,200 individual atoms. Altogether, these studies led to the conclusion that a similar function is supported by identical architectures of CPSF and CPF.

The reported structures correspond to three major components of the Poly(A) polymerase module: CPSF160, WDR33 and CPSF30 (Figure 2A). CPSF160 was predicted to contain three WD repeat domains although their precise boundaries were not easily predicted from the

primary sequence [51]. It turns out that the CPSF160 WD3 is composed of three non-consecutive stretches of residues distributed along its entire coding sequence. The first stretch is located just N-terminus to the WD1, the second stretch is found between the WD1 and WD2 domains, and the major part of the sequence for WD3 is located at the C-terminus of WD2 (Figure 2A and 2B). The three WD40 domains form a “Y” shape structure with a binding cavity that accommodates the WD40 domain of WDR33 (Figure 2A and 2B). Several key residues of CPSF160 responsible for docking WDR33 are highly conserved between species (Figure 2F). As expected, WDR33 folds as a WD40 domain. Critical residues involved in CPSF160 binding are located at the N-terminus of the WD-repeat domain and insert in the binding cavity located between the WD1 and the WD3 of CPSF160 (Figure 2A). The WDR33 N-terminus is not fully visible despite being present in the construct used for crystallization. Interestingly, part of this N-terminal sequence folds upon interaction with the PAS RNA (see next section; Figure 2C). The third component, CPSF30, is predicted to have five zinc finger domains connected by flexible linkers [52]. The tandem structure of zinc finger (ZF) 2 and 3 has been solved in complex with the NS1A protein in the context of preventing the antiviral cellular response [53]. Now, the structure of the N-terminal fragment of CPSF30 covering ZF1 to ZF3 (approximately from res. 1 to res. 130) is observed in complex with its native cellular partner CPSF160, as well as WDR33 and the PAS RNA (Figure 2C and 2D). The localization and orientation of the CPSF30 ZF3 domain is defined only in presence of the hexamer AAUAAA RNA (Figure 2D). Little is known about the Nuclease and the Poly(A) polymerase module interaction. Mapping sequence conservation onto the CPSF160 atomic surface highlights several large surface patches, which do not appear to contain residues from the CPSF160, CPSF30, WDR30 and hFip1 inter-subunit contacts (Figure 2E-2F)[45]. This high level of conservation usually indicates functionally relevant areas, and thus may indicate contact surfaces to the Nuclease module.

4. Polyadenylation signal recognition

The recent articles have also reported the structure of the CPSF Poly(A) polymerase module in complex with a short RNA molecule harboring the PAS sequence AAUAAA [46,47]. In the structure, the authors provide molecular details of the intimate recognition of the PAS by CPSF30 ZF2 and ZF3. This interaction is helped by the N-terminus and the WD40 domain of

WDR33 (Figure 3). These structures confirm that CPSF160 is not in direct contact with the PAS, contrary to what was believed prior to discovery of WDR33 as the PAS recognition factor [21,22]. Furthermore, CPSF160 is shielded from the PAS by the CPSF30 ZF2 and ZF3 domains. In these structures, the PAS signal contact with CPSF30 is stabilized by a short N-terminal peptide from WDR33 that folds onto the RNA to form a molecular embrace (Figure 3). WDR33 further contributes to the PAS interaction through two of the β -barrel blades from its WD-repeat domain (Figure 3). The recognition of the PAS sequence by the Poly(A) module involves numerous interactions. Specifically, the first two adenosines at position 1 and 2 are hydrogen-bonded to the ZF2 domain of CPSF30. The uridine at position 3 is also located between CPSF30 and WDR33 in such a conformation that its Watson-Crick edge interacts with the Hoogsteen edge of the adenosine at position 6 (Figure 3). Finally, the two “internal” adenosines of the PAS, A4 and A5, are interacting with the ZF3 domain of CPSF30. Around half of the PAS sequences in human and 40% in mouse do not match the canonical AAUAAA sequence although the third and the sixth positions are strictly conserved (Figure 1)[20,54]. These structures of the PAS recognition by CPSF factors explain at the molecular level why these two positions need to be conserved as they must be able to form a Hoogsteen base pair (Figure 3). Interestingly, mutations that specifically affect these positions are correlated with the appearance of human diseases such as α - and β -thalassemia [55,56].

Since the RNA sequence corresponds to a minimal PAS signature of 6 nucleotides, one should keep in mind that polyadenylation requires additional RNA sequence elements within the 3' UTR for efficient processing to occur (Figure 1). The distance between the PAS and the cleavage site varies between 10 to 30 nucleotides [57–59] and it is likely that Fip1 and CPSF160 helps stabilize the RNA in addition to roles as scaffolding proteins. Indeed, hFip1 was shown to bind RNA using different techniques such as gel shift assays, UV-cross linking with recombinant protein, and iCLIP-seq data [60,61]. CPSF160 was in fact wrongly attributed the role of PAS recognition subunit [6,54,55]. However, this conclusion was based on data from experiments using UV-cross linking. When those experiments were carried out, the presence of WDR33 as a CPSF component was unknown and unfortunately, WDR33 and CPSF160 co-migrate in SDS-PAGE. However, even though CPSF160 does not recognize the PAS, it was still shown to bind RNA using recombinant protein [64]. Other factors such CstF [23,39,65] and (mammalian) CFIm [25,66] also harbor RNA binding activity. These extra RNA

sequences are particularly important when one considers the case of alternative polyadenylation, which occurs in almost half of **human** mRNAs [67]. Indeed, it has been demonstrated that the UGUA sequence recognized by CFIm plays an important role in the choice of the polyadenylation site [68,69]. It **is also possible** that very few contacts exist between RNA and the various proteins of the 3' processing machinery outside the key signature sequences **already** known for decades. **This** situation would ensure plasticity required for **this** single machinery to process several thousand different mRNAs. The final answer to whether additional RNA interactions are required to properly cleave and polyadenylate pre-mRNA will likely **require** the use of longer RNA sequences in addition to a more complete reconstitution of the system.

5. Concluding remarks

With these **recent** outstanding studies, the scientific community has gained considerable **atomic** insight into polyadenylation signal recognition by the core proteins **of** the Poly(A) polymerase module **from** the pre-mRNA 3'-end maturation machinery. Nevertheless, a bigger picture is still needed to completely understand how the newly defined modules are integrated into CPSF, **what are their respective roles in RNA recognition and to determine the** interplay that undoubtedly exists between CPSF and the CstF, CFIm, and CFII complexes. It is also unclear whether the **current** definition of CPF in yeast is fully shared **with** human **CPSF**, apart from the **previously proposed** conservation of the Poly(A) polymerase **and cleavage modules** [21]. Another exciting topic **awaiting structural details is** the connection between 3'-end pre-mRNA processing factors and the maturation of histone mRNA or certain non-coding RNA **that are not polyadenylated. It is clear that** numerous mechanistic questions are still pending. **Nevertheless, we anticipate in the coming years** an enormous amount of exciting data **thanks** to the development of advanced technologies for cryo-EM and for expression and purification of recombinant materials.

Acknowledgements

We would like to thank Cameron Mackereth and Lionel Minvielle-Sébastien for their critical reading of the manuscript. This work was funded by **an** ANR grant (ANR Blanc 2010 grant

RIBOPRE40S (ANR-10-BLAN-1224) and RIBOMAN ANR-16-CE11-0029 to S.F.), the INSERM, the CNRS, and the University of Bordeaux.

Figure Legend

Figure 1: Schematic representation of the mammalian factors required for the pre-mRNA 3'-end maturation.

The various subcomplexes are labeled and their protein composition indicated according to current **information**. The cis-acting RNA elements **and** consensus sequences are displayed in black (or grey) and labeled. DSE stands for Downstream Sequence Element, Aux DES for Auxiliary Distal Sequence Element. **There is a specific focus on** the PolyAdenylation Signal (PAS), for which the strictly conserved positions have been highlighted in red.

Figure 2: Overall structure of the CPSF Poly(A) polymerase module.

The model shown in panels A, B and C is the PAS-bound form of the core components of CPSF corresponding to the PDB code 6FBS. The PAS-unbound CPSF30 structure comes from the PDB entry 6EOJ.

A) CPSF160 structure is shown as cartoon, colored by domain (green for WD40 domains, blue for the C-terminal helical domain (CTD), left panel) or represented as green surface (right panel). **In the left panel**, CPSF30 and WDR33 structures are shown as colored cartoons (gold for CPSF30, light pink for WDR33) with **a** transparent surface on top. **In the right panel** WDR33 is shown as cartoon and colored **by domain** (magenta for the WD40 domain, red for the N-terminal helix). The PAS RNA is shown in black sticks.

B) Domain organization for CPSF160, CPSF30 and WDR33.

C) Closeup view **of the** CPSF30 structure, **represent** as **a** gold-colored cartoon. CPSF160 and WDR33 models are shown as colored cartoon (green for CPSF160, light pink for WDR33) with transparent surface on top. Gold spheres correspond to the zinc ions present in the zinc finger domains. The PAS is shown as in panel A.

D) Superposition of the CPSF30 structures from yeast (in blue) and human (in gold) obtained in absence and in the presence of PAS RNA, respectively (in black). The third zinc finger

domain is only visible in the presence of the PAS RNA.

E and F) Side and top views of the CPSF160 model surface, colored according to sequence conservation as calculated using the ConSurf [webservice](#) (white: low conservation; magenta: high or strict conservation)[70]. Several areas show a high degree of conservation. The conservation has been calculated using 144 sequences with a sequence identity between 35% and 95%.

Figure 3: Polyadenylation signal sequence recognition by the Poly(A) polymerase module.

Closeup view of the PAS RNA binding site between CPSF30 and WDR33. Proteins and ions are shown as in [Figure 2](#). PAS RNA is depicted as sticks colored according to atom type (white for carbon, blue for nitrogen, red for oxygen, and orange for phosphate). CPSF30 side chains that coordinate the zinc ions are shown as sticks. Individual bases, zinc finger domains and RNA chain extremities are labeled.

References

- [1] A. Visvanathan, K. Somasundaram, mRNA Traffic Control Reviewed: N6-Methyladenosine (m6A) Takes the Driver's Seat, *BioEssays*. 40 (2018) 1700093. doi:10.1002/bies.201700093.
- [2] A. Tudek, M. Lloret-Llinares, T. Heick Jensen, The multitasking polyA tail: Nuclear RNA maturation, degradation and export, *Philos. Trans. R. Soc. B Biol. Sci.* 373 (2018) 20180169. doi:10.1098/rstb.2018.0169.
- [3] W.F. Marzluff, K.P. Koreski, Birth and Death of Histone mRNAs, *Trends Genet.* (2017) 1494–1508. doi:10.1016/j.tig.2017.07.014.
- [4] N.J. Proudfoot, Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut, *Science*. 352 (2016). doi:10.1126/science.aad9926.
- [5] D. Zheng, B. Tian, Sizing up the poly(A) tail: Insights from deep sequencing, *Trends Biochem. Sci.* 39 (2014) 255–257. doi:10.1016/j.tibs.2014.04.002.
- [6] W. Keller, S. Bienroth, K.M. Lang, G. Christofori, Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA, *EMBO J.* 10 (1991) 4241–4249. doi:10.1002/j.1460-2075.1991.tb05002.x.
- [7] Y. Takagaki, L.C. Ryner, J.L. Manley, Four factors are required for 3'-end cleavage of

- pre-mRNAs., *Genes Dev.* 3 (1989) 1711–1724. doi:10.1101/gad.3.11.1711.
- [8] Y. Takagaki, L.C. Ryner, J.L. Manley, Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation, *Cell.* 52 (1988) 731–742. doi:10.1016/0092-8674(88)90411-4.
- [9] G. Christofori, W. Keller, 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP, *Cell.* 54 (1988) 875–889. doi:10.1016/S0092-8674(88)91263-9.
- [10] J. Chen, C. Moore, Separation of Factors Required for Cleavage and Polyadenylation of Yeast Pre-messenger-RNA, *Mol. Cell. Biol.* 12 (1992) 3470–3481. doi:10.1128/MCB.12.8.3470.Updated.
- [11] P.J. Preker, M. Ohnacker, L. Minvielle-Sebastia, W. Keller, A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor, *Embo J.* 16 (1997) 4727–4737. doi:10.1093/emboj/16.15.4727.
- [12] E. Wahle, U. Rügsegger, 3'-End processing of pre-mRNA in eukaryotes, *FEMS Microbiol. Rev.* 23 (1999) 277–295. doi:10.1016/S0168-6445(99)00008-X.
- [13] E. Nedeá, X. He, M. Kim, J. Pootoolal, G. Zhong, V. Canadien, T. Hughes, S. Buratowski, C.L. Moore, J. Greenblatt, Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends, *J. Biol. Chem.* 278 (2003) 33000–33010. doi:10.1074/jbc.M304454200.
- [14] S. Chan, E.A. Choi, Y. Shi, Pre-mRNA 3'-end processing complex assembly and function, *Wiley Interdiscip. Rev. RNA.* 2 (2011) 321–335. doi:10.1002/wrna.54.
- [15] Y. Shi, D.C. Di Giammartino, D. Taylor, A. Sarkeshik, W.J. Rice, J.R. Yates, J. Frank, J.L. Manley, Molecular Architecture of the Human Pre-mRNA 3' Processing Complex, *Mol. Cell.* 33 (2009) 365–376. doi:10.1016/j.molcel.2008.12.028.
- [16] K. Xiang, L. Tong, J.L. Manley, Delineating the Structural Blueprint of the Pre-mRNA 3'-End Processing Machinery, *Mol. Cell. Biol.* 34 (2014) 1894–1910. doi:10.1128/MCB.00084-14.
- [17] K.M. Brown, G.M. Gilmartin, A Mechanism for the Regulation of Pre-mRNA 3' Processing by Human Cleavage Factor Im, *Mol. Cell.* 12 (2003) 1467–1476. doi:10.1016/S1097-2765(03)00453-2.

- [18] M.M. Kessler, M.F. Henry, E. Shen, J. Zhao, S. Gross, P.A. Silver, C.L. Moore, Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast, *Genes Dev.* 11 (1997) 2545–2556. doi:10.1101/gad.11.19.2545.
- [19] J.M. Pérez-Cañadillas, Grabbing the message: Structural basis of mRNA 3'UTR recognition by Hrp1, *EMBO J.* 25 (2006) 3167–3178. doi:10.1038/sj.emboj.7601190.
- [20] B. Tian, J. Hu, H. Zhang, C.S. Lutz, A large-scale analysis of mRNA polyadenylation of human and mouse genes, *Nucleic Acids Res.* 33 (2005) 201–212. doi:10.1093/nar/gki158.
- [21] L. Schönemann, U. Kühn, G. Martin, P. Schäfer, A.R. Gruber, W. Keller, M. Zavolan, E. Wahle, Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by WDR33, *Genes Dev.* 28 (2014) 2381–2393. doi:10.1101/gad.250985.114.
- [22] S.L. Chan, I. Huppertz, C. Yao, L. Weng, J.J. Moresco, J.R. Yates, J. Ule, J.L. Manley, Y. Shi, CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing, *Genes Dev.* 28 (2014) 2370–2380. doi:10.1101/gad.250993.114.
- [23] Y. Takagaki, J.L. Manley, C.C. MacDonald, J. Wilusz, T. Shenk, A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs, *Genes Dev.* 4 (1990) 2112–2120. doi:10.1101/gad.4.12a.2112.
- [24] H. de Vries, U. Rügsegger, W. Hübner, A. Freidlin, H. Langen, W. Keller, Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors, *EMBO J.* 19 (2000) 5895–5904. doi:10.1093/emboj/19.21.5895.
- [25] P. Schäfer, C. Tüting, L. Schönemann, U. Kühn, T. Treiber, N. Treiber, C. Ihling, A. Graber, W. Keller, G. Meister, A. Sinz, E. Wahle, Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors, *RNA.* 24 (2018) 1721–1737. doi:10.1261/rna.068056.118.
- [26] S. McCracken, N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S.D. Patterson, M. Wickens, D.L. Bentley, The C-terminal domain of RNA polymerase II couples mRNA processing to transcription, *Nature.* 385 (1997) 357–360. doi:10.1038/385357a0.
- [27] C.R. Mandel, S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J.L. Manley, L. Tong, Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease,

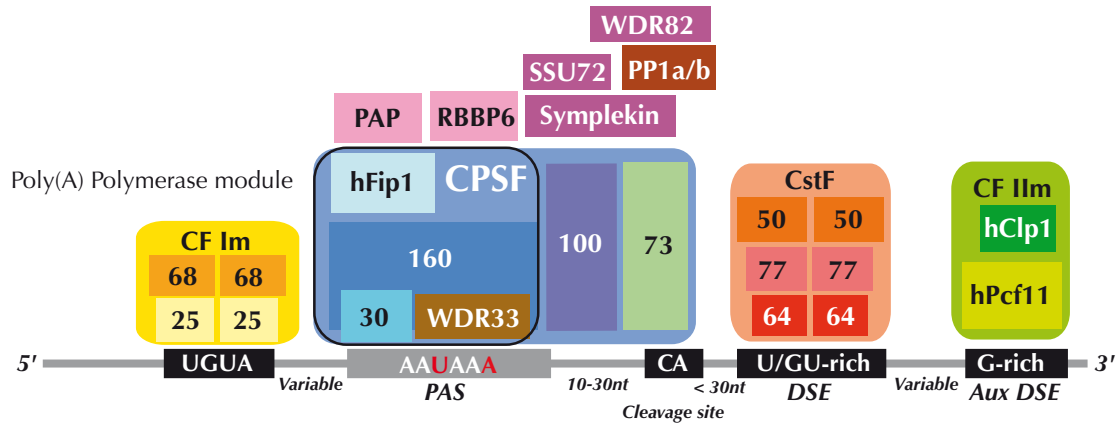
- Nature. 444 (2006) 953–956. doi:10.1038/nature05363.
- [28] J. Bard, A.M. Zhelkovsky, S. Helmling, T.N. Earnest, C.L. Moore, A. Bohm, Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP, *Science*. 289 (2000) 1346–1349. doi:10.1126/science.289.5483.1346.
- [29] G. Martin, W. Keller, S. Doublé, Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP, *EMBO J.* 19 (2000) 4193–4203. doi:10.1093/emboj/19.16.4193.
- [30] P.B. Balbo, A. Bohm, Mechanism of Poly(A) Polymerase: Structure of the Enzyme-MgATP-RNA Ternary Complex and Kinetic Analysis, *Structure*. 15 (2007) 1117–1131. doi:10.1016/j.str.2007.07.010.
- [31] G. Meinke, C. Ezeokonkwo, P. Balbo, W. Stafford, C. Moore, A. Bohm, Structure of yeast poly(A) polymerase in complex with a peptide from Fip1, an intrinsically disordered proteid, *Biochemistry*. 47 (2008) 6859–6869. doi:10.1021/bi800204k.
- [32] Y. Bai, T.C. Auperin, C.Y. Chou, G.G. Chang, J.L. Manley, L. Tong, Crystal Structure of Murine CstF-77: Dimeric Association and Implications for Polyadenylation of mRNA Precursors, *Mol. Cell*. 25 (2007) 863–875. doi:10.1016/j.molcel.2007.01.034.
- [33] P. Legrand, N. Pinaud, L. Minvielle-Sébastien, S. Fribourg, L. Minvielle-Sebastia, S. Fribourg, The structure of the CstF-77 homodimer provides insights into CstF assembly, *Nucleic Acids Res.* 35 (2007) 4515–4522. doi:10.1093/nar/gkm458.
- [34] M. Moreno-Morcillo, L. Minvielle-Sébastien, C. Mackereth, S. Fribourg, Hexameric architecture of CstF supported by CstF-50 homodimerization domain structure, *RNA*. 17 (2011) 412–418. doi:10.1261/rna.2481011.
- [35] Q. Yang, G.M. Gilmartin, S. Doublé, Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3' processing, *Proc. Natl. Acad. Sci.* 107 (2010) 10062–10067. doi:10.1073/pnas.1000848107.
- [36] Q. Yang, M. Coseno, G.M. Gilmartin, S. Doublé, Crystal structure of a human cleavage factor CFIm25/CFIm68/RNA complex provides an insight into poly(A) site recognition and RNA looping, *Structure*. 19 (2011) 368–377. doi:10.1016/j.str.2010.12.021.
- [37] A.R. Paulson, L. Tong, Crystal structure of the Rna14-Rna15 complex, *RNA*. 18 (2012) 1154–1162. doi:10.1261/rna.032524.112.
- [38] X. Qu, J.M. Perez-Canadillas, S. Agrawal, J. De Baecke, H. Cheng, G. Varani, C. Moore, The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a

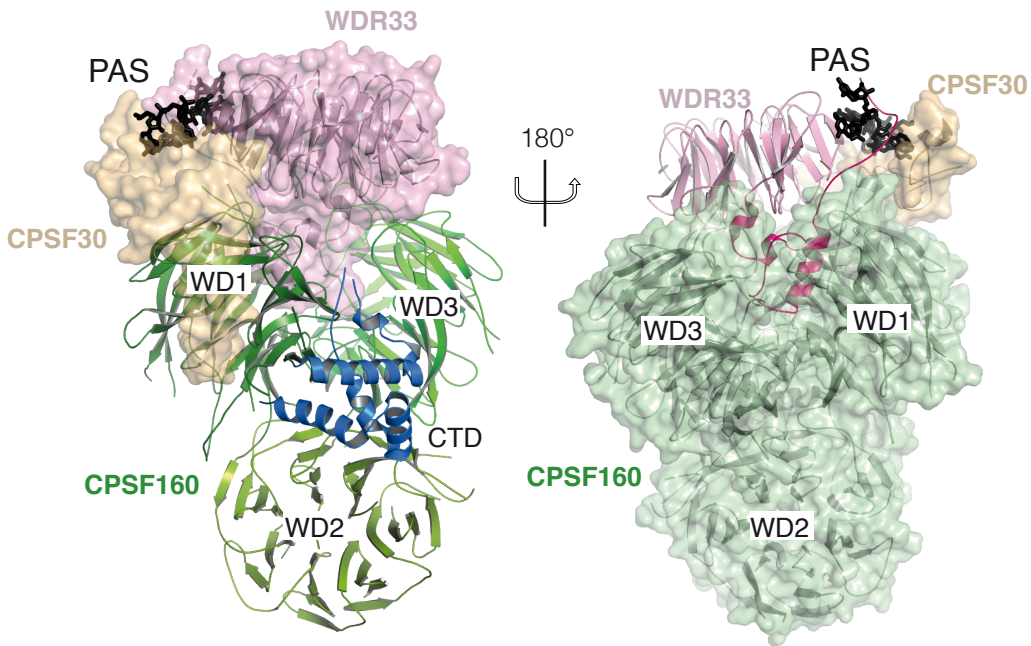
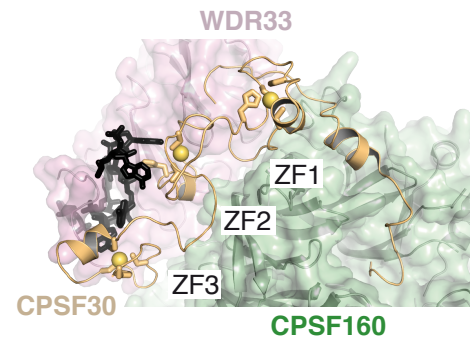
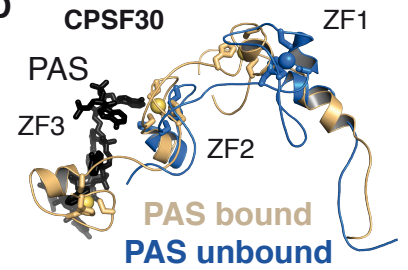
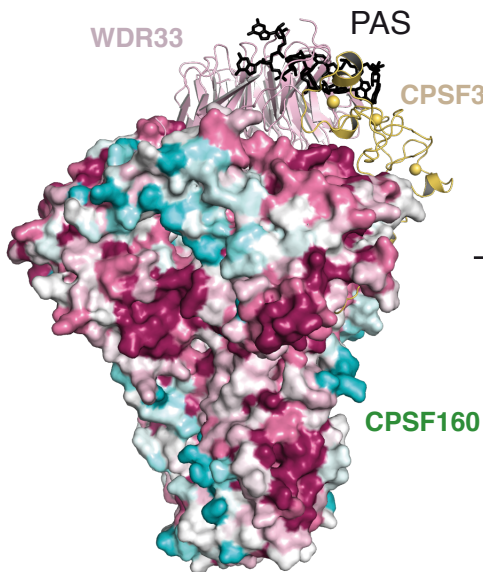
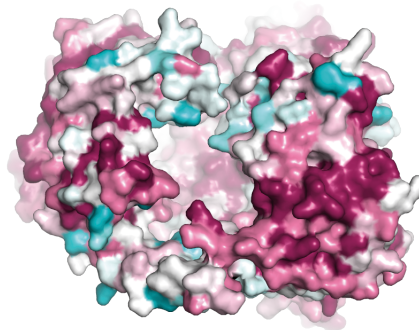
- new structure critical for mRNA 3'-end processing, *J. Biol. Chem.* 282 (2007) 2101–2115. doi:10.1074/jbc.M609981200.
- [39] J.M. Pérez Cañadillas, G. Varani, Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein, *EMBO J.* 22 (2003) 2821–2830. doi:10.1093/emboj/cdg259.
- [40] C. Pancevac, D.C. Goldstone, A. Ramos, I.A. Taylor, Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors, *Nucleic Acids Res.* 38 (2010) 3119–3132. doi:10.1093/nar/gkq002.
- [41] K. Xiang, T. Nagaike, S. Xiang, T. Kilic, M.M. Beh, J.L. Manley, L. Tong, Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex, *Nature.* 467 (2010) 729–733. doi:10.1038/nature09391.
- [42] K. Xiang, J.L. Manley, L. Tong, An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72, *Genes Dev.* 26 (2012) 2265–2270. doi:10.1101/gad.198853.112.
- [43] C.H. Hill, V. Boreikaité, A. Kumar, A. Casañal, P. Kubik, G. Degliesposti, S. Maslen, A. Mariani, O. von Loedelholz, M. Girbug, M. Skehel, L.A. Passmore, Activation of the Endonuclease that Defines mRNA 3' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex, *Mol Cell.* 73 (2019) 1–15. doi:10.1016/j.molcel.2018.12.023.
- [44] A. Casañal, A. Kumar, C.H. Hill, A.D. Easter, P. Emsley, G. Degliesposti, Y. Gordiyenko, B. Santhanam, J. Wolf, K. Wiederhold, G.L. Dornan, M. Skehel, C. V. Robinson, L.A. Passmore, Architecture of eukaryotic mRNA 3'-end processing machinery, *Science.* 358 (2017) 1056–1059. doi:10.1126/science.aao6535.
- [45] M. Clerici, M. Faini, R. Aebersold, M. Jinek, Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex, *Elife.* 6 (2017). doi:10.7554/eLife.33111.
- [46] M. Clerici, M. Faini, L.M. Muckenfuss, R. Aebersold, M. Jinek, Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex, *Nat. Struct. Mol. Biol.* 2 (2018) 135–138. doi:10.1038/s41594-017-0020-6.
- [47] Y. Sun, Y. Zhang, K. Hamilton, J.L. Manley, Y. Shi, T. Walz, L. Tong, Molecular basis for the recognition of the human AAUAAA polyadenylation signal, *Proc. Natl. Acad. Sci.* 115 (2017) 1419–1428. doi:10.1073/pnas.1718723115.

- [48] M. Lidschreiber, A.D. Easter, S. Battaglia, J.B. Rodríguez-Molina, A. Casañal, M. Carminati, C. Baejen, P. Grzechnik, K.C. Maier, P. Cramer, L.A. Passmore, The APT complex is involved in non-coding RNA transcription and is distinct from CPF, *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky845.
- [49] D.C. Di Giammartino, W. Li, K. Ogami, J.J. Yashinski, M. Hoque, B. Tian, J.L. Manley, RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs, *Genes Dev.* 28 (2014) 2248–2260. doi:10.1101/gad.245787.114.
- [50] K.D. Sullivan, M. Steiniger, W.F. Marzluff, A Core Complex of CPSF73, CPSF100, and Symplekin May Form Two Different Cleavage Factors for Processing of Poly(A) and Histone mRNAs, *Mol. Cell.* 34 (2009) 322–332. doi:10.1016/j.molcel.2009.04.024.
- [51] B. Dichtl, D. Blank, M. Sadowski, W. Hübner, S. Weiser, W. Keller, Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination, *EMBO J.* 21 (2002) 4125–4135. doi:10.1093/emboj/cdf390.
- [52] S.M.L. Barabino, W. Hübner, A. Jenny, L. Minvielle-Sebastia, W. Keller, The 30-kd subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are rna-binding zinc finger proteins, *Genes Dev.* 11 (1997) 1703–1716. doi:10.1101/gad.11.13.1703.
- [53] K. Das, L.-C. Ma, R. Xiao, B. Radvansky, J. Aramini, L. Zhao, J. Marklund, R.-L. Kuo, K.Y. Twu, E. Arnold, R.M. Krug, G.T. Montelione, Structural basis for suppression of a host antiviral response by influenza A virus, *Proc. Natl. Acad. Sci.* 105 (2008) 13093–13098. doi:10.1073/pnas.0805213105.
- [54] E. Beaudoin, D. Gautheret, Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST Data, *Genome Res.* 11 (2001) 1520–1526. doi:10.1101/gr.190501.
- [55] D.R. Higgs, S.E.Y. Goodbourn, J. Lamb, J.B. Clegg, D.J. Weatherall, N.J. Proudfoot, α -Thalassaemia caused by a polyadenylation signal mutation, *Nature.* 306 (1983) 398–400. doi:10.1038/306398a0.
- [56] S.H. Orkin, T.C. Cheng, S.E. Antonarakis, H.H. Kazazian Jr., H.H. Kazazian, Thalassaemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene, *EMBO J.* 4 (1985) 453–456. doi:10.1002/J.1460-2075.1985.TB03650.X.
- [57] N.J. Proudfoot, G.G. Brownlee, 3' Non-coding region sequences in eukaryotic

- messenger RNA, *Nature*. 263 (1976) 211–214. doi:10.1038/263211a0.
- [58] M. Zavolan, A.J. Gruber, R. Schmidt, M. Belmadani, S. Ghosh, A.R. Gruber, G. Martin, W. Keller, A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation, *Genome Res.* 26 (2016) 1145–1159. doi:10.1101/gr.202432.115.
- [59] N. Yurko, M. Hoque, F. Bachand, M. Larochelle, J.L. Manley, X. Liu, B. Tian, J.-F. Lemay, Comparative analysis of alternative polyadenylation in *S. cerevisiae* and *S. pombe*, *Genome Res.* 27 (2017) 1685–1695. doi:10.1101/gr.222331.117.
- [60] I. Kaufmann, G. Martin, A. Friedlein, H. Langen, W. Keller, Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase, *EMBO J.* 23 (2004) 616–626. doi:10.1038/sj.emboj.7600070.
- [61] B. Lackford, C. Yao, G.M. Charles, L. Weng, X. Zheng, E.A. Choi, X. Xie, J. Wan, Y. Xing, J.M. Freudenberg, P. Yang, R. Jothi, G. Hu, Y. Shi, Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal, *EMBO J.* 33 (2014) 878–889. doi:10.1002/emboj.201386537.
- [62] C.L. Moore, J. Chen, J. Whoriskey, Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding, *EMBO J.* (1988).
- [63] G.M. Gilmartin, J.R. Nevins, An ordered pathway of assembly of components required for polyadenylation site recognition and processing, *Genes Dev.* 3 (1989) 2180–2190. doi:10.1101/gad.3.12b.2180.
- [64] K.G.K. Murthy, J.L. Manley, The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation, *Genes Dev.* 9 (1995) 2672–2683. doi:10.1101/gad.9.21.2672.
- [65] Y. Takagaki, J.L. Manley, RNA recognition by the human polyadenylation factor CstF., *Mol. Cell. Biol.* 17 (1997) 3907–3914. doi:10.1128/MCB.17.7.3907.
- [66] D. Hollingworth, C.G. Noble, I.A. Taylor, A. Ramos, RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11, *RNA.* 12 (2006) 555–560. doi:10.1261/rna.2304506.
- [67] B. Tian, J.L. Manley, Alternative polyadenylation of mRNA precursors, *Nat. Rev. Mol. Cell Biol.* 18 (2016) 18–30. doi:10.1038/nrm.2016.116.
- [68] Y. Zhu, X. Wang, E. Forouzmand, J. Jeong, F. Qiao, G.A. Sowd, A.N. Engelman, X. Xie,

- K.J. Hertel, Y. Shi, Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation, *Mol. Cell.* 69 (2018) 62–74.e4. doi:10.1016/j.molcel.2017.11.031.
- [69] G. Martin, A.R. Gruber, W. Keller, M. Zavolan, Genome-wide Analysis of Pre-mRNA 3' End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3' UTR Length, *Cell Rep.* 1 (2012) 753–763. doi:10.1016/j.celrep.2012.05.003.
- [70] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, *Nucleic Acids Res.* 44 (2016) W344-50. doi:10.1093/nar/gkw408.



A**C****B****D****E****F**

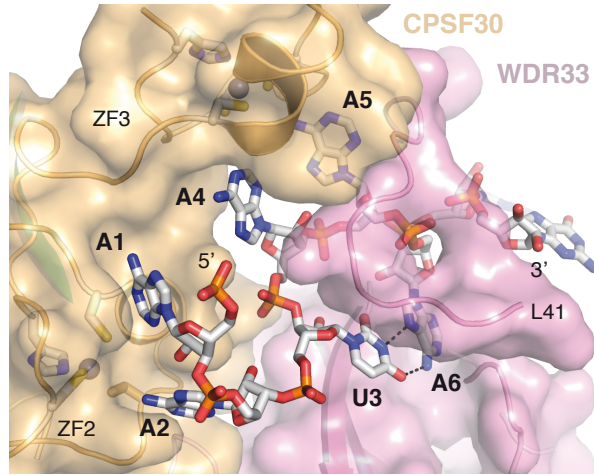


Table I: Yeast and human pre-mRNA 3'-end maturation proteins.

Human			Yeast		
Complex	Subunit		Complex	Subunit	
CPSF	CPSF160	Poly(A) polymerase module	CPF	Cft1/Yhh1	Poly(A) polymerase module
	WDR33			Pfs2	
	hFip1			Fip1	
	CPSF30			Yth1	
	PAPOLA*			Pap1	
	CPSF100	Nuclease module		Cft2/Ydh1	Nuclease module
	CSPF73			Ysh1/Brr5	
	RBBP6*			Mpe1	Phosphatase module
	Symplekin*			Pta1	
PP1a/b		Glc7			
SSU72		Ssu72			
WDR82		Swd2/Cps35			
		Pti1			
		Ref2			
		Syc1			
CstF	CstF77		CF IA	Rna14	
	CstF64			Rna15	
	CstF50			-	
CFIIm	hPcf11			Pcf11	
	hClp1			Clp1	
CFIm	CFIm 68/59				
	CFIm 25				
			CF IB	Hrp1	

* Note: Currently there is no exact correspondence between the human and the yeast factors organization into modules, although partial overlap has been documented. A cleavage module was proposed for human CPSF that includes CPSF73 and CPSF100, but not Symplekin and RBBP6²¹. Similarly, PAPOLA is independent of CPSF. WDR82 and PP1a/b were not shown to be involved in the 3'-end pre-mRNA processing, and Syc1 is not a component of CPF but defines the APT complex (Associated with Pta1) in association to the phosphatase module⁴⁸. Finally, one should not confuse mammalian CFIIm (defined as hPcf11 and hClp1) and yeast CF II (defined as Cft1, Ydh1, Ysh1 and an unknown factor)¹².

E. Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by WDR33. *Genes Dev* 2014; 28:2381–93.

12. Wahle E, Rügsegger U. 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.*1999; 23:277–95.