Architecture of eukaryotic mRNA 3′-end processing machinery

Ana Casañal,1* Ananthanarayanan Kumar,1* Chris H. Hill,1 Ashley D. Easter,1 Paul Emsley,1 Gianluca Degliesposti,1 Yuliya Gordiyenko,1,2 Balaji Santhanam,1 Jana Wolf,1 Katrin Wiederhold,1 Gillian L. Dornan,1 Mark Skehel,1 Carol V. Robinson,2 Lori A. Passmore1†

1MRC Laboratory of Molecular Biology, Cambridge UK. 2Chemistry Research Laboratory, University of Oxford, Oxford, UK.

*These authors contributed equally to this work. †Corresponding author. Email: passmore@mrc-lmb.cam.ac.uk

Newly transcribed eukaryotic pre-mRNAs are processed at their 3′-ends by the ~1 MDa multiprotein machinery.

Here, we show that the nuclease, polymerase and phosphatase activities of yeast CPF are organized into three modules. Using cryo-EM, we determine a 3.5 Å resolution structure of the ~200 kDa polymerase module. This reveals four beta propellers in an assembly strikingly similar to other protein complexes that bind nucleic acid. Combined with in vitro reconstitution experiments, our data show that the polymerase module brings together factors required for specific and efficient polyadenylation, to help coordinate mRNA 3′-end processing.

Protein-coding genes in eukaryotes are transcribed by RNA polymerase II (Pol II) as precursor mRNAs, which undergo 5′ capping, splicing, and 3′-end processing. The 3′-end processing machinery includes the highly-conserved cleavage and polyadenylation factor (CPF). CPF cleaves pre-mRNAs, adds a poly(A) tail and triggers transcription termination but it is unclear how its different enzymes are coordinated and assembled.

To understand the architecture of CPF, we analyzed the overall stoichiometry, protein–protein interactions and composition of the purified complex using non-covalent nanoelectrospray ionization mass spectrometry (nanoESI-MS) (II). Ionization causes fragmentation of CPF into 38 different sub-complexes comprised of the 15 previously identified subunits of CPF (table S2). Computational analysis revealed a protein–protein interaction network where subunits are organized into three prominent modules centered around the CPF enzymatic activities: nuclease (Ysh1), phosphatase (Ssu72, Glc7), and polymerase (Pap1) (Fig. 1C).

The nuclease module is comprised of three subunits (Ysh1, Cft2 and Mpe1) while the phosphatase module contains seven subunits (Pta1, Ref2, Pti1, Swd2, Glc7, Ssu72 and Syd2). Syc1 was only observed when CPF was purified from a yeast strain with a tagged phosphatase module subunit (table S2). Syc1 bears homology to the C terminus of Ysh1, and they may occupy the same, mutually-exclusive binding site on Pta1 (12). Thus, CPF might contain fourteen, not fifteen, subunits while APT (associated with Pta1, (5)) may be a separate complex with six overlapping subunits.

The poly(A) polymerase module contains five subunits: Cft1, Pfs2, Pap1, Fip1 and Yth1 (Fig. 1C). Cft1 appears to play a central role as it is present in 14 of the 15 polymerase module sub-complexes (table S2). Pap1 and Fip1 can be present in up to two copies within the complex (fig. S2A). Fip1 is thought to tether Pap1 to CPF (13–16). We also observe sub-complexes that contain Pap1, but not Fip1 (table S2), suggesting that Pap1 may contact other CPF subunits (16, 17). The poly(A) polymerase module is analogous to a four-subunit mammalian complex which is necessary and sufficient for specific in vitro polyadenylation (18). This suggests that the architecture of yeast and mammalian complexes is highly similar.

NanoESI-MS did not reveal interactions between the three enzymatic modules of CPF. The modules may be held together by hydrophobic interactions that are stable in solution but weakened in the gas phase (II). Pulldowns with...
subunits from each module revealed potential connections between them (fig. S2B).

To understand the molecular basis of subunit association, we used cryo-EM to study CPF isolated from yeast. This resulted in a 3D reconstruction at ~12 Å resolution (fig. S3). At this resolution, it is not possible to assign densities to subunits. Moreover, this structure is too small to represent the entire CPF complex.

Next, because of the central role of polyadenylation in 3′-end processing, we developed a strategy to overexpress the polymerase module in insect cells for structural and biochemical characterization. This could be purified with or without the Pap1 subunit, consistent with nanoESI-MS (fig. S4A). We imaged the ~200 kDa recombinant four-subunit complex (Cft1, Pfs2, Yth1, Fip1) using cryo-EM (fig. S4B, C). We determined a 3D reconstruction of the complex, at an overall resolution of 3.5 Å, allowing us to build atomic models into the density for 1,717 amino acids (table S3; fig. S4, S5). Prior to this, the only high resolution structure available for this complex was a crystal structure of 72 amino acids of CPSF30/Yth1 (19). The polymerase module is strikingly similar to the structure we obtained with the native CPF preparation (fig. S3).

In our cryo-EM map (Fig. 2; movie S1), three of the four subunits are well-ordered: Cft1 (residues 1–1356), Pfs2 (27–414) and Yth1 (1–97). The C-terminal half of Yth1 (zinc fingers 3–5, residues 98–208), all of Fip1, and several loops in Cft1 are not visible and are presumably disordered, consistent with predictions (fig. S6A).

Cft1 forms the core of the complex and is comprised of three seven-bladed beta propellers followed by a C-terminal helical domain. Beta propeller (BP) 1 and BP2 are each formed of contiguous sequences. In contrast, BP3 is predominantly C-terminal, but it also contains one beta strand from the N terminus, and three beta strands from the middle of the protein, creating an intertwined and rigid structural core (fig. S6B). The C-terminal helical domain of Cft1 is located at the nexus of the three beta propellers, further stabilizing the fold (Fig. 2B).

Our structure reveals an extensive interface between Cft1 and Pfs2, burying >4,200 Å² surface area (Fig. 3A). Almost 50 amino acids in the N-terminal region of Pfs2 are inserted into the cavity between Cft1-BP1 and -BP3, forming contacts with the tops of both beta propellers (Fig. 3B). A Pfs2 beta propeller (fig. S6C) is then positioned on the top of Cft1 stabilized by loops extending from BP1 and BP3. Many key interactions between these two proteins are conserved in the human orthologs CPSF160 and WDR33 (Fig. 3B, fig. S7).

Yth1 is anchored onto the complex by an extended N-terminal segment that binds in the central cavity of Cft1-BP3 and continues across a hydrophobic external face (Fig. 3C, D). Next, two of the five Cys-Cys-Cys-His (CCCH) zinc fingers pack into the interface between Cft1 and Pfs2 (Fig. 3E, F).

We performed cross-linking mass spectrometry to validate our structural model and to determine where Pap1 and Fip1 bind. Inter- and intra-molecular cross-links agree with our atomic models, and the crystal structure of Pap1 (fig. S8; table S4) (20). Fip1 crosslinks to the C-terminal part of Yth1 and the polymerase domain of Pap1 (fig. S8B). A previous crystal structure of Pap1 in complex with a peptide of Fip1 (15) revealed molecular details of their interaction but Pap1 also cross-links to the C-terminal helical domain of Cft1, ZnF1 of Yth1, and the C-terminal region of Pfs2 (fig. S8). Together, these data suggest that the flexible C-terminal half of Yth1 binds the intrinsically disordered protein Fip1, which in turn flexibly tethers Pap1 to the complex, allowing conformational freedom to add long poly(A) tails onto diverse RNA substrates.

The cryo-EM structure of the poly(A) polymerase module has a strikingly similar architecture to the eukaryotic DDB1-DDB2 and SF3b complexes (21–24) (Fig. 4A–C). DDB1-DDB2 recognizes UV-damaged DNA and acts as an adapter for a cullin-RING E3 ubiquitin ligase to trigger nucleotide excision repair. SF3b is a multi-protein assembly containing the Rse1/SF3b130 scaffold protein, which forms part of the U2snRNP complex essential for pre-mRNA splicing and branch site recognition.

DDB1 and Rse1 both contain three beta-propellers followed by a C-terminal alpha helical domain, with the same fold as Cft1, and the three proteins show weak sequence homology (~15% sequence identity). Thus, all three complexes use similar scaffold proteins (DDB1, Rse1 or Cft1) to assemble a rigid and structurally stable complex. Their interaction partners (DDB2, Hsh155/Rds3 or Pfs2) bind in the same cavity between BP1 and BP3, with a similar binding mode that involves alpha helices, but the exact interaction mechanism is not conserved (Fig. 4A–C). Like DDB1-DDB2 and SF3b, Cft1 may bind additional subunits through its beta propellers.

DDB1-DDB2 and SF3b directly bind nucleic acid. The polymerase module also binds RNA in a gel shift assay (fig. S9A). The surface of Pfs2 equivalent to the DDB2 DNA binding site contains a cluster of conserved lysines, arginines and aromatic residues that could form an RNA interaction surface (fig. S9B–D). This Pfs2 surface lies adjacent to the RNA-binding ZnF2 of Yth1 (25–27) and together, they might comprise a composite RNA binding platform (Fig. 4A) that is disrupted in viral infections (fig. S9E).

Intact CPF requires the accessory cleavage factors (CF) IA and IB for efficient and specific polyadenylation (28–30). Addition of recombinant CF IA (but not CF IB) stimulates the polyadenylation activity of the polymerase module and intact CPF (Fig. 4D, fig. S10A–C). CF IA has no effect on iso-
ated Pap1 (fig. S10D), underscoring the functional importance of the other polymerase subunits.

CF IA is comprised of four different protein subunits—a heterotrimer of Rna14–Rna15 and a heterodimer of Pcf11–C1p1. Rna14–Rna15 was sufficient to stimulate polyadenylation (fig. S10E). Moreover, the CF IA complex, and specifically the Rna14–Rna15 subcomplex, binds to the polymerase module in pulldown assays (Fig. 4E; fig. S10F).

The arrangement of CPF, where its enzymatic activities are segregated into three modules, suggests that coupling between the enzymes is not through intimate, stable contacts, and may be dynamic. CF IA likely stimulates polyadenylation by contributing additional RNA binding sites. Together, Pfs2, Yth1 and CF IA could stably bring specific RNA sequences to the complex. This would allow Pap1, which is flexibly tethered to the complex by the intrinsically disordered protein Fip1, to access a variety of different RNA substrates for efficient and controlled polyadenylation. Thus, the polymerase module acts as a hub to bring together Pap1, substrate RNA and CF IA (Fig. 4F). Moreover, by tethering these components together with the nuclease and phosphatase modules of CPF, it would facilitate accurate 3′-end processing and co-ordination with transcription.

REFERENCES AND NOTES

First release: 26 October 2017


...


ACKNOWLEDGMENTS

We thank Michael Webster, Christos Savva, Andreas Boland, Rafael Fernández-Leiro, Thomas Martin, Garib Murshudov, Sjors Scheres, Alan Brown, Chris Russo, Katerina Naydenova, Amy Yewdall, Ann Kelley, LMB EM facility and LMB scientific computation for assistance and advice; and David Barford, Andrew Carter and Madan Babu for reviewing the manuscript. This work was supported by an EMBO Long-Term Fellowship ALTF66-2015 co-funded by the European Commission (LTFCOFUND2013, GA-2013-609409) through Marie Curie Actions (to A.C); Gates Cambridge (to A.K.); the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant 261151 (to L.A.P); the European Union’s Horizon 2020 research and innovation programme (ERC grant 725685) (to L.A.P); ERC Grant No. 695511 (ENABLE) (to C.V.R.); and Medical Research Council grants MC_U105192715 (L.A.P) and MC_U105185859 (to Madan Babu). We acknowledge Diamond for access to eBIC (proposal EM15622) funded by the Wellcome Trust, MRC and BBSRC. Cryo-EM density maps are deposited in the EMDB (EMD-3908) and atomic coordinates are deposited in the PDB (6EOJ). Reagents are available upon request from L.A.P. under a material transfer agreement with MRC. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aaf6535/DC1

Materials and Methods

Figs. S1 to S10

Tables S1 to S4

References (31–76)

Movie S1

Additional Data S1

11 August 2017; accepted 12 October 2017

Published online 26 October 2017

10.1126/science.aaf6535
Fig. 1. Architecture of native yeast CPF. (A) Coomassie-stained SDS-PAGE showing CPF/APT purified from a yeast strain containing TAPS-tagged Ref2 (marked with an asterisk). (B) Coupled cleavage and polyadenylation assay of purified CPF analyzed by denaturing urea-PAGE. CYC1 is the substrate RNA. Cleavage products are CYC1-5’ and CYC1-3’. Aberrant cleavage products are marked with asterisks. (C) Interaction network between subunits of CPF and APT complexes from computational analysis of nanoESI-MS data (solid lines) and from pull-downs (dotted lines). Black lines indicate confirmed binary interactions; grey lines are used for interacting proteins where the direct interaction partner could not be confirmed (see Methods). The yellow dashed line denotes that Syc1 and Ysh1 likely bind Pta1 in a mutually exclusive manner. Protein symbols are scaled to have an area proportionate to their molecular weights. Yellow stars denote enzymes.
Fig. 2. Cryo-EM structure of the polymerase module of CPF. (A) Cryo-EM map and (B) cartoon representation of the atomic model of polymerase module. Yth1 (magenta), Pfs2 (yellow), Cft1 (green) and zinc ions (pale cyan) are depicted. The three beta-propeller domains of Cft1 (BP1, BP2 and BP3) are colored in different shades. (C) Schematic representation of polymerase module subunits, with domain boundaries. Grey regions are not visible in the cryo-EM map. Human orthologs are given with the percent sequence identity (similarity).
Fig. 3. Protein–protein interactions in the polymerase module. (A–B) Details of the Cft1 (green)–Pfs2 (yellow) interaction where the cryo-EM map of Cft1 is shown (A). Selected interactions with residues conserved in human orthologs labeled in blue (B). (C–F) Details of Yth1 (magenta) interaction with Pfs2 and Cft1 where a surface representation of the Cft1 model is shown (C). Selected electrostatic and hydrophobic interactions are depicted (D). Zinc ions are in cyan (E). ZnF2 is stabilized by pi stacking between Yth1-H85 and -W70, as well as several hydrogen-bonding interactions between side chains of Pfs2 and backbone atoms of Yth1 (F).
**Fig. 4. The polymerase module acts as a hub to bring together Pap1, RNA and accessory factors.**  
(A–C) The polymerase module (A) is structurally similar to the DDB1–DDB2 DNA repair (PDB: 3ei3; B) and SF3b splicing (PDB: 5gm6; C) complexes.  
(D) Polyadenylation of a fluorescently labeled 42 nt pre-cleaved (pc) CYC1 RNA by the polymerase module (with and without CF IA) analyzed by 15% denaturing urea PAGE.  
(E) Coomassie-stained SDS-PAGE of pulldown experiment showing immobilized polymerase module after incubation with purified maltose-binding protein (MBP), CF IA or Rna14–Rna15.  
(F) Model for the 3′-end processing machinery obtained by combining data from nanoESI-MS, cryo-EM, cross-linking MS and in vitro pull downs. Yellow stars denote enzymes.
Architecture of eukaryotic mRNA 3'-end processing machinery

Ana Casañal, Ananthanarayanan Kumar, Chris H. Hill, Ashley D. Easter, Paul Emsley, Gianluca Degliesposti, Yuliya Gordiyenko, Balaji Santhanam, Jana Wolf, Katrin Wiederhold, Gillian L. Dornan, Mark Skehel, Carol V. Robinson and Lori A. Passmore

published online October 26, 2017 originally published online October 26, 2017