Biophysical characterizations of the recognition
of the AAUAAA polyadenylation signal

Keith Hamilton,[1] Yadong Sun[1] & Liang Tong[1#]

[1]Department of Biological Sciences
Columbia University
New York, NY 10027, USA

corresponding author
       ltong@columbia.edu (L.T.)

Abstract

Most eukaryotic messenger RNA precursors must undergo 3′-end cleavage and polyadenylation for maturation. We and others recently reported the structure of the AAUAAA polyadenylation signal (PAS) in complex with the protein factors CPSF-30, WDR33 and CPSF-160, revealing the molecular mechanism for this recognition. Here we have characterized in detail the interactions between the PAS RNA and the protein factors using fluorescence polarization experiments. Our studies show that AAUAAA is recognized with ~1 nM affinity by the CPSF-160–WDR33–CPSF-30 ternary complex. Variations in the RNA sequence can greatly reduce the affinity. Similarly, mutations of residues that have van der Waals interactions with the bases of AAUAAA also lead to substantial reductions in affinity. Finally, our studies confirm that both CPSF-30 and WDR33 are required for binding the PAS RNA, and determine a ~7 nM affinity between CPSF-30 and the CPSF-160–WDR33 binary complex.

Short title: Recognition of the AAUAAA polyadenylation signal

Introduction

Most eukaryotic messenger RNA precursors (pre-mRNAs) must undergo extensive processing to become functional mRNAs, which includes 5′-end capping, splicing, and 3′-end cleavage and polyadenylation (1-5). The recognition of a polyadenylation signal (PAS) is a crucial step for 3′-end processing, which helps to define the position of cleavage in the pre-mRNA as the PAS is often located 10-30 nucleotides upstream of the cleavage site. The PAS is a hexanucleotide, and the most common motif is AAUAAA (~55% frequency) for mammalian pre-mRNAs, followed by the AUUAAA motif (~16% frequency) (6-8). Many other motifs can also support 3′-end processing, but are much rarer (<4% frequency). AAUAAA, AUUAAA and 10 other single nucleotide variants account for ~92% of PAS in human and mouse pre-mRNAs (8).

Many proteins are involved in pre-mRNA 3′-end processing (9-11), and several sub-complexes of this 3′-end processing machinery have been identified, including the cleavage and polyadenylation specificity factor (CPSF) and the cleavage stimulation factor (CstF). The 73-kDa subunit of CPSF (CPSF-73) is the endoribonuclease for the cleavage reaction (12), and two other CPSF subunits, WDR33 (11) and CPSF-30, are required for recognizing the PAS (13,14). CPSF-30 also interacts with Fip1, another subunit of CPSF, which helps to recruit the poly(A) polymerase to the 3′-end processing machinery. CstF recognizes a G/U-rich sequence motif downstream of the cleavage site, and it also has an important role in alternative polyadenylation (15-17).

We and others recently reported the structures of a quaternary complex of human CPSF-160, CPSF-30, WDR33 and an AAUAAA PAS RNA (Figs. 1A, 1B) (18,19), the structure of a ternary complex of the yeast protein homologs (Cft1, Yth1 and Pfs2, without RNA) (20), as well as the structure of a binary complex of human CPSF-160 and WDR33 (21). The structures of the quaternary complexes revealed extensive and specific interactions between the AAUAAA PAS and WDR33 and CPSF-30 (Fig. 1C), while

3

CPSF-160 serves a crucial scaffolding role in the complex. In addition, there is a Hoogsteen base pair between U3 and A6 of the PAS. A few aspects of the interactions between the ternary complex and PAS RNA have been studied by fluorescence polarization assays (21).

We report here detailed characterizations of the interactions between CPSF-160, WDR33, CPSF-30, and various PAS RNAs. We show that the ternary complex has high affinity for the AAUAAA PAS RNA, with K_d of ~1 nM, and the AUUAAA PAS RNA has a K_d of ~10 nM. In comparison, other sequence motifs that can also support 3′-end processing, as well as changes to the U3-A6 Hoogsteen base pair, lead to substantial reduction in the binding affinity. In addition, mutations of CPSF-30 residues that are in contact with the RNA bases can also give rise to reductions in binding affinity. The CPSF-160–WDR33 and CPSF-160–CPSF-30 binary complexes have much lower affinity for the RNA, confirming that both WDR33 and CPSF-30 are required for PAS recognition. The CPSF-160–WDR33 binary complex has high affinity for CPSF-30, with a K_d of ~7 nM.

Results

Variation of the AAUAAA RNA length

The structures of the quaternary complexes show that WDR33 and CPSF-30 primarily recognize the PAS hexanucleotide itself (Fig. 1C). Although a 17-mer RNA was used for the structural study, only the PAS was found to be well ordered (18). The nucleotide directly following the PAS was weakly ordered, and the other nucleotides were disordered. To assess this structural observation, we used RNA oligos of various lengths, 17-mer (FAM-AACCUCCAAUAAACAAC), 11-mer (FAM-CCUCCAAUAAACA) and 6-mer (AAUAAA-FAM), and carried out fluorescence polarization binding assays. All these oligo RNAs carry a FAM fluorescent label at the 5′ or 3′ end, allowing direct observation of the fluorescence polarization signal.

4

The experimental data confirmed that the 17-mer and 11-mer RNAs have nearly the same binding affinity to the CPSF-160–WDR33–CPSF-30 ternary complex, with K_d values of 0.56 nM and 1.4 nM, respectively (Fig. 2A). The 17-mer RNA was used in titration experiments numerous times, and the observed K_d values ranged between 0.6 and 1.2 nM, consistent with that reported in an earlier study (21). Fip1 (residues 137-243) was included at 10 μM concentration in all the assays to stabilize the C-terminal segment of CPSF-30, although it did not have an effect on RNA binding.

In comparison, the 6-mer oligo had a higher K_d value of 29 nM. This could be due to interference by the FAM label or contribution from nucleotides outside the PAS. For example, the phosphate group of the nucleotide directly following the PAS has some interactions with WDR33 (Fig. 1C). To assess these different scenarios, we used an unlabeled AAUAAA 6-mer oligo and an unlabeled CAAUAAAC 8-mer oligo and carried out competition fluorescence polarization binding assays against the 17-mer FAM-labeled oligo (Fig. 2B). The K_d value determined from this assay for the AAUAAA 8-mer oligo was 72 nM, 24-fold higher than the 11-mer oligo, while the 6-mer oligo showed some binding only at 5000 nM concentration of the ternary complex. These data indicate that nucleotides outside the AAUAAA hexamer also contribute significantly to the binding.

The phosphate group directly following the PAS is likely a major factor enhancing the binding affinity. The FAM label in the 6-mer oligo is located at the 3′-end, which contains a phosphate group at that position, consistent with the higher affinity of this oligo (Fig. 2A). We carried out the competition assay with the unlabeled 11-mer oligo (Fig. 2B), and obtained a K_d value of 3.0 nM, indicating that the FAM label did not have a significant effect on the binding affinity of this longer RNA (the K_d for the FAM-labeled 11-mer RNA was 1.4 nM, Fig. 2A). Overall, these assays demonstrate that AAUAAA plus the phosphate group of the following nucleotide is crucial for high-affinity binding. Other nucleotides make smaller contributions to the interaction.

5

Variation of the U3-A6 Hoogsteen base pair

A U3-A6 Hoogsteen base pair was observed in the PAS when bound to the CPSF-160–WDR33–CPSF-30 ternary complex (Fig. 1C) (18,19). The binding mode of the PAS suggests that other Hoogsteen base pairs, such as C3-G6, could be accommodated, while a wobble U3-G6 base pair would not fit. To assess whether the alternative Hoogsteen base pair can support binding to the ternary complex, we determined the binding affinity of unlabeled 11-mer RNA with AACAAG (variations from the AAUAAA PAS are indicated by underlines) as the equivalent of the AAUAAA PAS by competition fluorescence polarization assays. The experimental data showed no binding of the oligo even at 5000 nM concentration of the ternary complex (Fig. 2C), indicating that the alternative Hoogsteen base pair could not be accommodated in the binding site. The 11-mer RNA with AAAAAU as the PAS showed no binding either.

We also characterized the binding of FAM-labeled 17-mer RNAs with AAGAAA and AACAAA as the PAS, to test the effect of breaking the U3-A6 Hoogsteen base pair. These RNAs did show binding to the ternary complex, with K_d values of 69 and 180 nM, respectively (Fig. 2A), roughly 120- and 320-fold higher than AAUAAA. The K_d value for the AAGAAA oligo is consistent with that reported in an earlier study (21).

Variation of other positions of the AAUAAA PAS

Besides the U3-A6 Hoogsteen base pair, the structures show that A1 and A4 of the AAUAAA PAS are specifically recognized by CPSF-30, while A2 and A5 also establish favorable hydrogen-bonding interactions (18,19). To assess the binding affinity of other PAS hexamers that can also support 3′-end processing, we selected from those identified in mammalian pre-mRNAs (7,8), changing each of these four positions at a time. The unlabeled 11-mer RNAs that we studied included GAUAAA (~1% frequency, first position), AUUAAA (~16%, second position), AAUGAA (~1%, fourth position), and AAUACA (~2%, fifth position) as the PAS.

6

The experimental data showed that these variant PAS RNAs have much lower affinity for the ternary complex, except for AUUAAA, which is the second most frequently observed PAS (Fig. 2D). The K_d value for AUUAAA 11-mer oligo is 10.3 nM, only about 3-fold higher than the corresponding 11-mer AAUAAA oligo (K_d of 3.0 nM). The GAUAAA 11-mer oligo has a K_d value of 170 nM (55-fold higher), while the AAUGAA and AAUACA 11-mer oligos showed only minor binding at 5000 nM concentration (Fig. 2D). AAGAAA is another PAS hexamer with low frequency (~3%), and it had K_d of 69 nM (120-fold higher) (Fig. 2A).

Mutations of CPSF-30

Besides hydrogen-bonding interactions between the adenine bases and the backbone of CPSF-30, there are also extensive van der Waals interactions. Specifically, the A1, A2, A4 and A5 bases are each involved in π-stacking interaction with an aromatic side chain of CPSF-30: A1 with Phe84, A2 with His70, A4 with Phe112, and A5 with Phe98 (Fig. 1C). In addition, A2 is flanked on the other face by two Lys side chains, Lys77 and Lys78, and Arg73 has ionic interactions with the phosphate connecting nucleotides U3 and A4.

To test the importance of these interactions for PAS RNA binding, we produced the H70A, R73A, K77A/K78A, F84A and F112A mutants of CPSF-30. Each mutant and the wild-type CPSF-30 was mixed with the CPSF-160–WDR33 binary complex at 10-fold molar excess, which we showed was sufficient to achieve nearly maximal binding to the RNA (see next). The affinity of these mixtures for the FAM-labeled 17-mer RNA was then determined (Fig. 3). The K77A/K78A mutant had roughly 3-fold higher K_d (2.8 nM) compared to wild-type CPSF-30. On the other hand, the F84A and R73A mutations had larger effects on the binding, with 35- and 55-fold higher K_d values (32 and 49 nM). Finally, the H70A and F112A mutants showed the largest effects, with 440- and 590-fold higher K_d values (390 and 530 nM).

Requirement of both WDR33 and CPSF-30 for RNA binding

7

The AAUAAA RNA is bound at the interface between WDR33 and CPSF-30 (Fig. 1C), and our earlier electrophoretic mobility shift assays showed that CPSF-30 alone, or the CPSF-160–WDR33 and CPSF-160–CPSF-30 binary complexes could not bind the RNA (18). To characterize these interactions more quantitatively, we mixed CPSF-30 at increasing molar ratios (0-, 0.5-, 1-, 2-, 5-, 10-, 20- and 30-fold) with the CPSF-160–WDR33 binary complex, and observed a clear enhancement of the apparent affinity of the mixture for the RNA when CPSF-30 concentration was increased (Fig. 4). Above 10-fold molar ratio of CPSF-30 relative to CPSF-160–WDR33, nearly maximal RNA binding was obtained. On the other hand, in the absence of CPSF-30, no binding was observed even at 125 nM concentration of the CPSF-160–WDR33 binary complex. Similarly, no RNA binding was observed for CPSF-30 alone at up to 125 nM concentration, consistent with both WDR33 and CPSF-30 being required for RNA binding.

Assuming that each mixture contains two equilibria: A+B=AB and AB+C=ABC, where A represents the CPSF-160–WDR33 binary complex, B represents CPSF-30, and C represents the RNA, the total concentrations of CPSF-160–WDR33, CPSF-30 and RNA in the mixture and the K_d values for the two equilibria would determine the shape of each titration curve, which varied the concentration of CPSF-160–WDR33 (A) while keeping the molar ratio of CPSF-30 (B) constant. The K_d of the ternary complex for the RNA (ABC) was fixed at 0.6 nM based on the titration data for this complex (Fig. 2A). We carried out a global fit of all the titration curves, which determined the K_d for CPSF-30 binding to CPSF-160–WDR33 as 7.1 nM (Fig. 4).

Discussion

Overall, our studies have provided detailed knowledge on PAS recognition by the CPSF-160–WDR33–CPSF-30 ternary complex, extending beyond the binding study reported earlier (21). The results confirm that both hydrogen-bonding and van der Waals

interactions are important for the recognition of the A1, A2, A4 and A5 bases. Changing the identity of these bases generally has strong deleterious effects on the binding affinity, with the exception of the AUUAAA sequence, consistent with it also being frequently observed for 3′-end processing. On the other hand, loss of π-stacking interactions with these bases is also detrimental for the recognition. The structure suggests that the U2 base of AUUAAA could maintain the hydrogen-bond with the main-chain amide of Lys78 in CPSF-30 with a small conformational change, as well as the π-stacking with His70 (Fig. 1C).

Our studies demonstrate the importance of the U3-A6 Hoogsteen base pair for high-affinity binding to the ternary complex. The AA<u>A</u>AA<u>U</u> hexamer, swapping the positions of the U3 and A6 nucleotides, cannot maintain a Hoogsteen base pair as the AAUAAA hexamer, explaining the lack of binding for this RNA. On the other hand, the AA<u>C</u>AA<u>G</u> hexamer appears to fit nicely into the binding site, with the guanine base flanked on either side by Phe43 and Phe153 of WDR33 and picking up a hydrogen-bond between its 2-amino group and the main-chain carbonyl of Thr115 in WDR33. The exact reason why this hexamer cannot bind with high affinity is not clear, although it is consistent with the fact that it is not frequently observed for 3′-end processing (8). At the same time, the AA<u>G</u>AAA and AA<u>C</u>AAA hexamers are able to bind the ternary complex, albeit with substantially reduced affinity, indicating that a base pair here may not be absolutely required for binding.

The less frequently observed PAS motifs studied here appear to have much lower affinity for the ternary complex. In fact, the AAUAAA hexamer is most often associated with the last PAS in human and mouse pre-mRNAs, while the less frequently observed hexamers are associated with upstream PAS of pre-mRNAs with two or more processing sites (8). Especially, the AA<u>G</u>AAA hexamer is often found as the PAS in an upstream exon. This suggests that these less frequently observed PAS hexamers have a more prominent role in alternative polyadenylation, and that other protein factors (such as CstF)

9

as well as recognition of auxiliary sequence motifs may be important for the processing at

these sites.

Methods

Protein expression and purification. A human CPSF-160–WDR33 binary complex was expressed in baculovirus-infected Hi5 insect cells and a full-length human CPSF-30 was expressed in *E. coli* as an MBP fusion protein as described earlier (18). The components were mixed together for the binding assays, allowing the variation of the molar ratios of CPSF-30 relative to CPSF-160–WDR33.

A CPSF-160–WDR33–CPSF-30 ternary complex was expressed in Hi5 insect cells with the Multibac expression system (22) (Geneva Biotech). WDR33 (residues 1-425) carried an N-terminal His-tag, followed by MBP and a TEV protease cleavage site. CPSF-160 and CPSF-30 are untagged.

Human Fip1 (residues 137-243) was cloned into the pET28a vector (Novagen), with N-terminal His and SUMO tags, and expressed in *E. coli* BL21(DE3) cells.

The CPSF-160–WDR33–CPSF-30 and the CPSF-160–WDR33 complexes were purified following the same protocol. The insect cells were lysed by sonication in a buffer containing 50 mM Tris (pH 8.0), 500 mM NaCl, 30 mM imidazole, 10 mM beta-mercaptoethanol, and one SIGMAFAST protease inhibitor cocktail tablet. The lysate was mixed with 2 mL Ni-NTA beads (Qiagen), washed with 15 mL buffer containing 2 M NaCl to remove bound nucleic acids, then the protein was eluted with 5 mL buffer containing 250 mM imidazole. 1 mg TEV protease was added, and the sample was incubated overnight at 4°C. It was then run over a Superdex200 16/60 column (GE Healthcare), using a buffer containing 20 mM (Tris 8.0), 350 mM NaCl, and 10 mM DTT.

For Fip1, the cells were lysed by sonication in a buffer containing 50mM Tris (pH 8.0), 200 mM NaCl, 30 mM imidazole, 10 mM beta-mercaptoethanol, and 2 mM PMSF. The lysate was incubated with 5 mL Ni-NTA, washed with 30 mL buffer, then Fip1 was eluted with 10 mL buffer containing 250 mM imidazole. 100 μg UlpI protease was added, and the sample was incubated at 4°C for one hour. It was then purified further using a 5 mL Fastflow MonoQ column followed by a Superdex 200 16/60 column using a buffer containing 20 mM Tris (pH 8.0), 100 mM NaCl, and 10 mM DTT.

Fluorescence polarization binding assays. The assays were performed using a Neo2S plate reader (Biotek). The buffer for all the assays contained 20 mM Tris (pH 8.0), 150 mM NaCl, 10 mM DTT, and 10 μM Fip1. A 17-mer oligonucleotide that included a polyadenylation site and the surrounding bases from the SV40 virus and a 6-carboxyfluorescein (6-FAM) 5'-end label was used as the probe at a concentration of 1 nM in all experiments. For assays that involved titration with CPSF-30, various molar ratios of CPSF-30 were added to CPSF-160–WDR33 and the mixture was allowed to incubate on ice for 1 h. Oligonucleotides were purchases from IDT. The data from competition experiments were fitted using an analytical equation (23).

Figure legends

Figure 1. Overall structure of the human CPSF-160–WDR33–CPSF-30–PAS RNA quaternary complex. **(A)**. Schematic drawing of the quaternary complex. CPSF-160 (yellow), WDR33 (blue) and CPSF-30 (green) are shown as molecular surfaces. The PAS RNA is shown as a sphere model (orange). **(B)**. Structure of the quaternary complex, viewed after a 90° rotation of around the vertical axis. **(C)**. Recognition of the AAUAAA PAS (orange) by CPSF-30 zinc fingers ZF2-ZF3 (green) and WDR33 (blue). Hydrogen-bonds in the U3-A6 Hoogsteen base pair are indicated with dashed lines in red. Side chains of CPSF-30 that contact the RNA bases are shown as stick models, and those selected for mutagenesis studies are labeled in red. Zinc atoms are shown as spheres in pink. Produced with PyMOL (www.pymol.org).

Figure 2. Effects of variations in the RNA on PAS recognition. **(A)**. Fluorescence polarization binding assays of the CPSF-160–WDR33–CPSF-30 ternary complex with labeled 17-mer, 11-mer and 6-mer AAUAAA PAS RNAs, as well as labeled 17-mer RNAs with AA<u>G</u>AAA and AA<u>C</u>AAA as the PAS (nucleotides distinct from AAUAAA are indicated by underline). The curves represent theoretical fitting to the binding data. **(B)**. Competition fluorescence polarization binding assays of the CPSF-160–WDR33–CPSF-30 ternary complex with unlabeled 11-mer, 8-mer and 6-mer AAUAAA PAS RNAs. Labeled 17-mer AAUAAA PAS RNA was used as the reporter. **(C)**. Competition fluorescence polarization binding assays of the CPSF-160–WDR33–CPSF-30 ternary complex with unlabeled 11-mer RNAs containing variations of the U3-A6 Hoogsteen base pair, AA<u>A</u>AA<u>U</u> and AA<u>C</u>AA<u>G</u>. **(D)**. Competition fluorescence polarization binding assays of the CPSF-160–WDR33–CPSF-30 ternary complex with unlabeled 11-mer RNAs containing variations of the A1, A2, A4 and A5 bases. Error bars are standard deviations from two repeats.

Figure 3. Effects of mutations in CPSF-30 on PAS recognition. Fluorescence polarization binding assays between the labeled 17-mer AAUAAA PAS RNA and mixtures of the CPSF-160–WDR33 binary complex and CPSF-30 wild-type and mutants at 10-fold molar ratio. Error bars are standard deviations from two repeats.

Figure 4. Both CPSF-30 and WDR33 are required for PAS RNA binding. Fluorescence polarization binding assays between the labeled 17-mer AAUAAA PAS RNA and mixtures of the CPSF-160–WDR33 binary complex and CPSF-30 at increasing molar ratios. Error bars are standard deviations from two repeats.

13

References

1. Proudfoot, N.J. (2011) Ending the message: poly(A) signals then and now. *Genes Develop.*, **25**, 1770-1782.
2. Yang, Q. and Doublie, S. (2011) Structural biology of poly(A) site definition. *Wiley Interdiscip. Rev. RNA*, **2**, 732-747.
3. Xiang, K., Tong, L. and Manley, J.L. (2014) Delineating the structural blueprint of the pre-mRNA 3' end processing machinery. *Mol. Cell. Biol.*, **34**, 1894-1910.
4. Curinha, A., Oliveira Braz, S., Pereira-Castro, I., Cruz, A. and Moreira, A. (2014) Implications of polyadenylation in health and disease. *Nucleus*, **5**, 508-519.
5. Shi, Y. and Manley, J.L. (2015) The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Develop.*, **29**, 889-897.
6. Proudfoot, N.J. and Brownlee, G.G. (1976) 3' non-coding region sequences in eukaryotic messenger RNA. *Nature*, **263**, 211-214.
7. Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M. and Gautheret, D. (2000) Patterns of variant polyadenylation signal usage in human genes. *Genome Res.*, **10**, 1001-1010.
8. Tian, B., Hu, J., Zhang, H. and Lutz, C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucl. Acid Res.*, **33**, 201-212.
9. Zhao, J., Hyman, L. and Moore, C.L. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, **63**, 405-445.
10. Mandel, C.R., Bai, Y. and Tong, L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.*, **65**, 1099-1122.
11. Shi, Y., di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates III, J.R., Frank, J. and Manley, J.L. (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell*, **33**, 365-376.
12. Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L. and Tong, L. (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*, **444**, 953-956.
13. Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates III, J.R., Ule, J., Manley, J.L. and Shi, Y. (2014) CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Develop.*, **28**, 2370-2380.
14. Schonemann, L., Kuhn, U., Martin, G., Schafer, P., Gruber, A.R., Keller, W., Zavolan, M. and Wahle, E. (2014) Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Develop.*, **28**, 2381-2393.
15. Elkon, R., Ugalde, A.P. and Agami, R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, **14**, 496-506.
16. Gruber, A.R., Martin, G., Keller, W. and Zavolan, M. (2014) Means to an end: mechanism of alternative polyadenylation of messenger RNA precursors. *Wiley Interdiscip. Rev. RNA*, **5**, 183-196.
17. Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, **18**, 18-30.
18. Sun, Y., Zhang, Y., Hamilton, K., Manley, J.L., Shi, Y., Walz, T. and Tong, L. (2018) Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proc. Natl. Acad. Sci. USA*, **115**, E1419-E1428.

19. Clerici, M., Faini, M., Muckenfuss, L.M., Aebersold, R. and Jinek, M. (2018) Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. *Nat. Struct. Mol. Biol.*, **25**, 135-138.

20. Casanal, A., Kumar, A., Hill, C.H., Easter, A.D., Emsley, P., Degliesposti, G., Gordiyenko, Y., Santhanam, B., Wolf, J., Wiederhold, K. *et al.* (2017) Architecture of eukaryotic mRNA 3'-end processing machinery. *Science*, **358**, 1056-1059.

21. Clerici, M., Faini, M., Aebersold, R. and Jinek, M. (2017) Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. *eLife*, **6**, e33111.

22. Sari, D., Gupta, K., Thimiri Govinda Raj, D.B., Aubert, A., Drncova, P., Garzoni, F., Fitzgerald, D. and Berger, I. (2016) The MultiBac baculovirus/insect cell expression vector system for producing complex protein biologics. *Adv. Exp. Med. Biol.*, **896**, 199-215.

23. Wang, Z.X. (1995) An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Lett.*, **360**, 111-114.
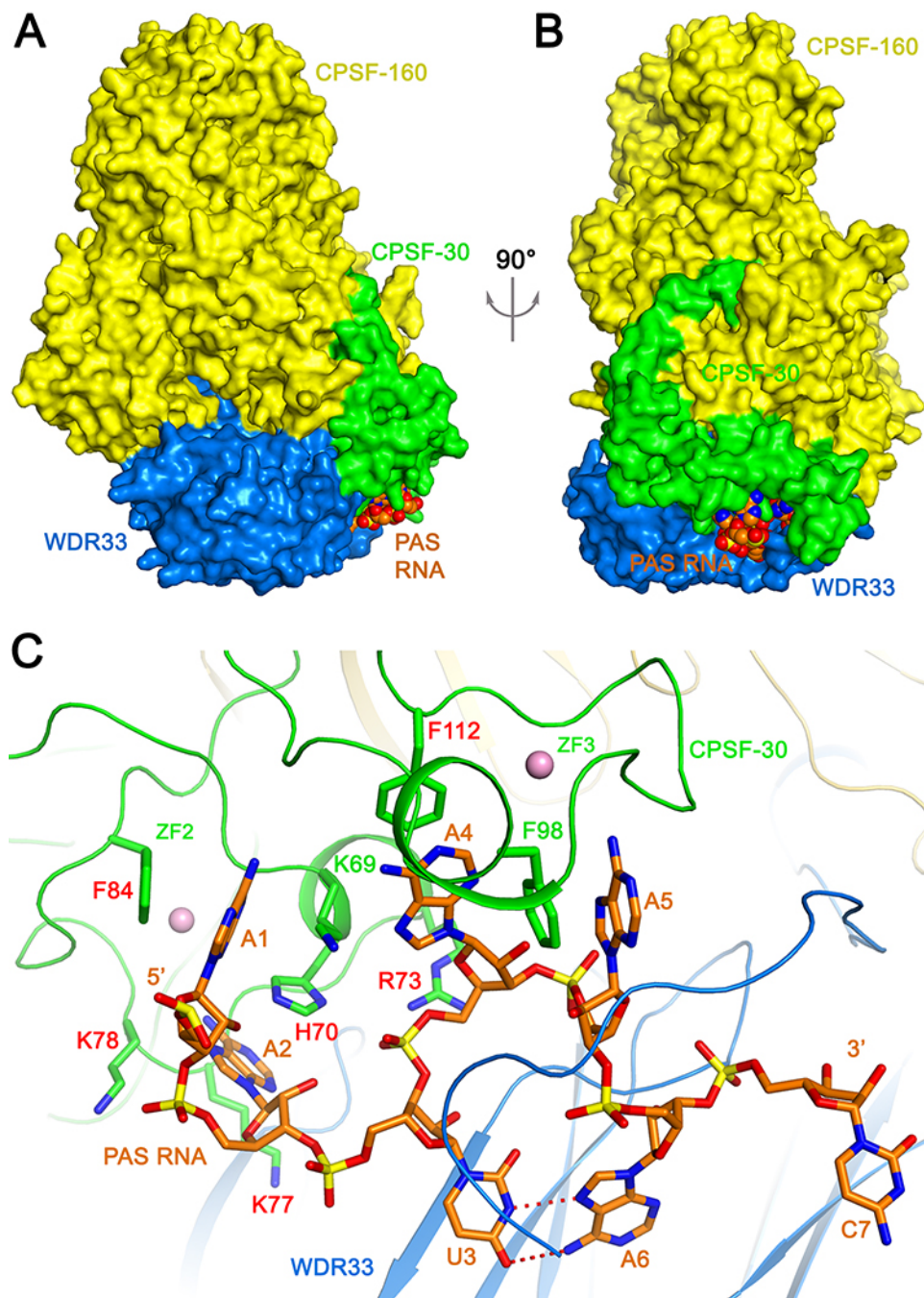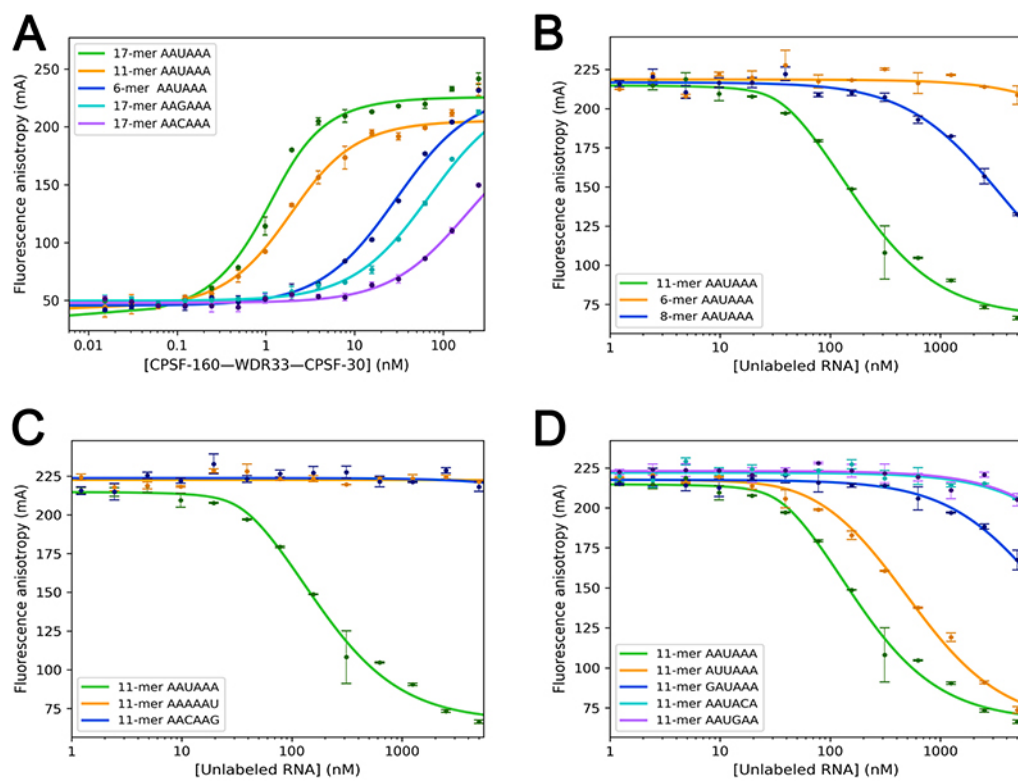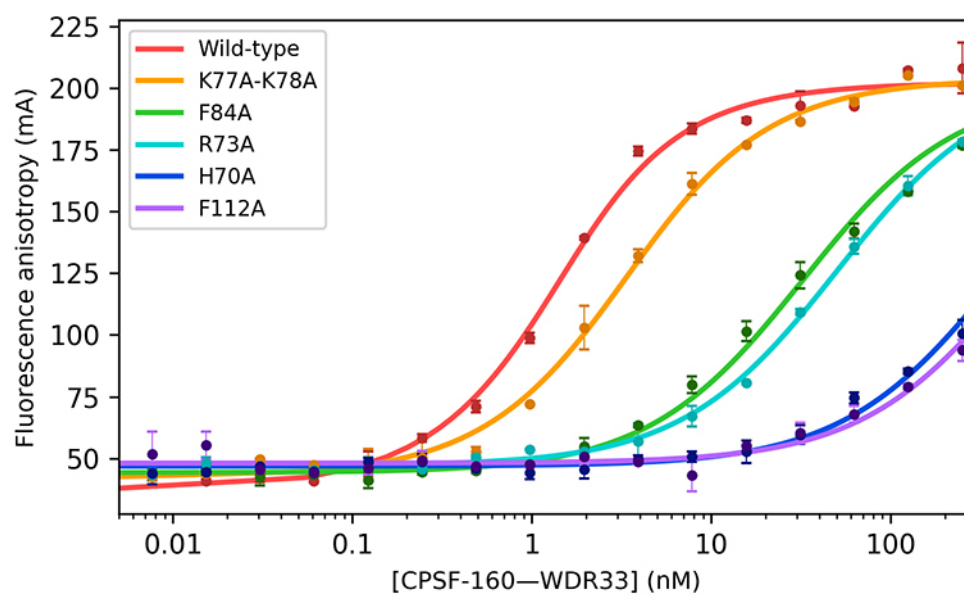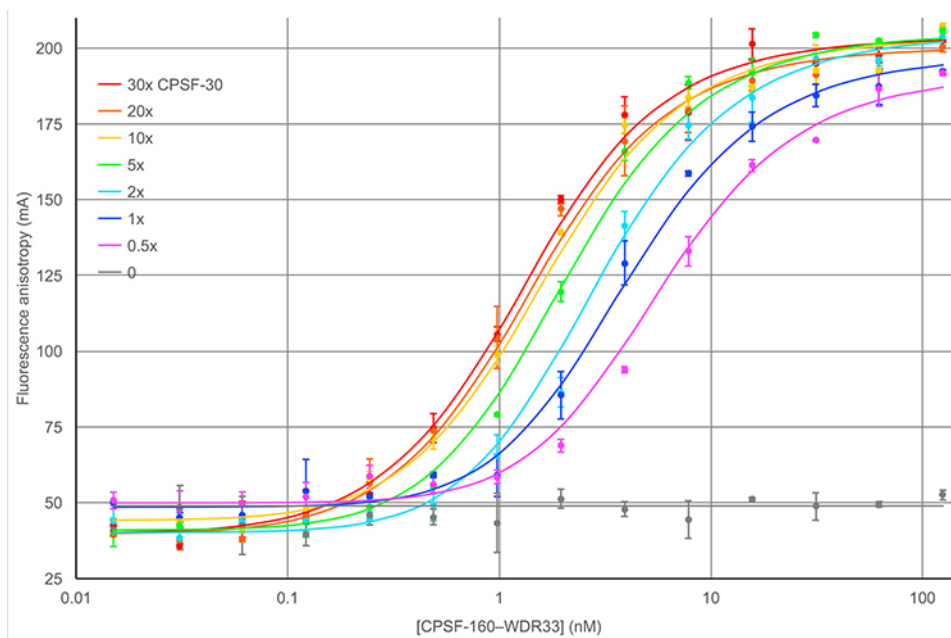
Fig. 1

Fig. 2

Fig. 3

18

Fig. 4

19