Biochemistry

pubs.acs.org/biochemistry Article

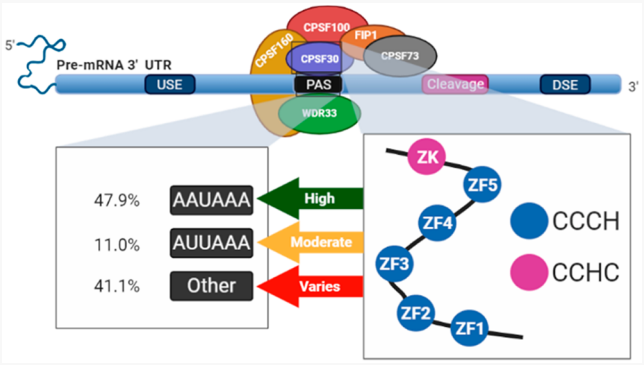
Understanding RNA Binding by the Nonclassical Zinc Finger Protein CPSF30, a Key Factor in Polyadenylation during Pre-mRNA Processing

Jordan D. Pritts, Abdulafeez A. Oluyadi, Weiliang Huang, Geoffrey D. Shimberg, Maureen A. Kane, Angela Wilks, and Sarah L. J. Michel*

Cite This: *Biochemistry* 2021, 60, 780−790 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

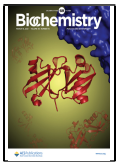
ABSTRACT: Cleavage and polyadenylation specificity factor 30 (CPSF30) is a zinc finger protein that regulates pre-mRNA processing. CPSF30 contains five CCCH domains and one CCHC domain and recognizes two conserved 3′ pre-mRNA sequences: an AU hexamer and a U-rich motif. AU hexamer motifs are common in pre-mRNAs and are typically defined as AAUAAA. Variations within the AAUAAA hexamer occur in certain pre-mRNAs and can affect polyadenylation efficiency or be linked to diseases. The effects of disease-related variations on CPSF30/pre-mRNA binding were determined using a construct of CPSF30 that contains just the five CCCH domains (CPSF30-5F). Bioinformatics was utilized to identify the variability within the AU hexamer sequence in pre-mRNAs. The effects of this sequence variability on CPSF30-5F/RNA binding affinities were measured. Bases at positions 1, 2, 4, and 5 within the AU hexamer were found to be important for RNA binding. Bioinformatics revealed that the three bases flanking the AU hexamer at the 5′ and 3′ ends are twice as likely to be adenine or uracil as guanine and cytosine. The presence of A and U residues in these flanking regions was determined to promote higher-affinity CPSF30-5F/RNA binding than G and C residues. The addition of the zinc knuckle domain to CPSF30-5F (CPSF30-FL) restored binding to AU hexamer variants. This restoration of binding is connected to the presence of a U-rich sequence within the pre-mRNA to which the zinc knuckle binds. A mechanism of differential RNA binding by CPSF30, modulated by accessibility of the two RNA binding sites, is proposed.



In eukaryotes, zinc finger proteins (ZFs) play important roles in transcription, translation, and regulation. It is estimated that 5% of all human proteins are ZFs, and these proteins share the common feature of having modular domains with a combination of cysteine and histidine residues that function as ligands to bind zinc.[1] Once zinc is bound, ZFs adopt distinct secondary structures that facilitate RNA, DNA, and protein binding.[2−9] ZFs are classified on the basis of their domain identity and architecture, and a class that has received considerable attention in recent years is the CCCH class.[3,10,11] The CCCH ZF motif has been identified in the sequence of numerous proteins with roles in RNA regulation, including RNA metabolism, splicing, decay, and translation.[2,3,10,11] CCCH type ZFs typically have between one and six domains, and there is emerging evidence that ZFs with CCCH domains preferentially recognize AU-rich RNA target sequences.[3,10−15]

One member of the CCCH class of ZFs is cleavage and polyadenylation specificity factor 30 (CPSF30 or CPSF4).[16] CPSF30 contains five CCCH ZF domains and a singular CCHC or zinc knuckle domain. CPSF30 is part of a complex of proteins called CPSF that regulate pre-mRNA processing.[13,17−21] During pre-mRNA processing, the CPSF complex facilitates recognition of the polyadenylation sequence (PAS), cleavage of the 3′ end, recruitment of poly(A) polymerase, and the subsequent addition of a polyadenosine tail, also known as polyadenylation (Figure 1A). Studies utilizing nuclear extracts that contain the CPSF proteins, the purified CPSF complex, and *in vitro*-translated CPSF proteins have identified some of the protein−protein and protein−RNA interactions within the CPSF complex.[16,22−24] The individual roles of the isolated CPSF proteins are not well understood. Our laboratory has previously reported the isolation of full length CPSF30 (CPSF30-FL) and a construct that contains just the five CCCH domains (CPSF30-5F).[13,25] ZFs with CCCH domains

Received: December 3, 2020
Revised: February 9, 2021
Published: February 22, 2021



https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

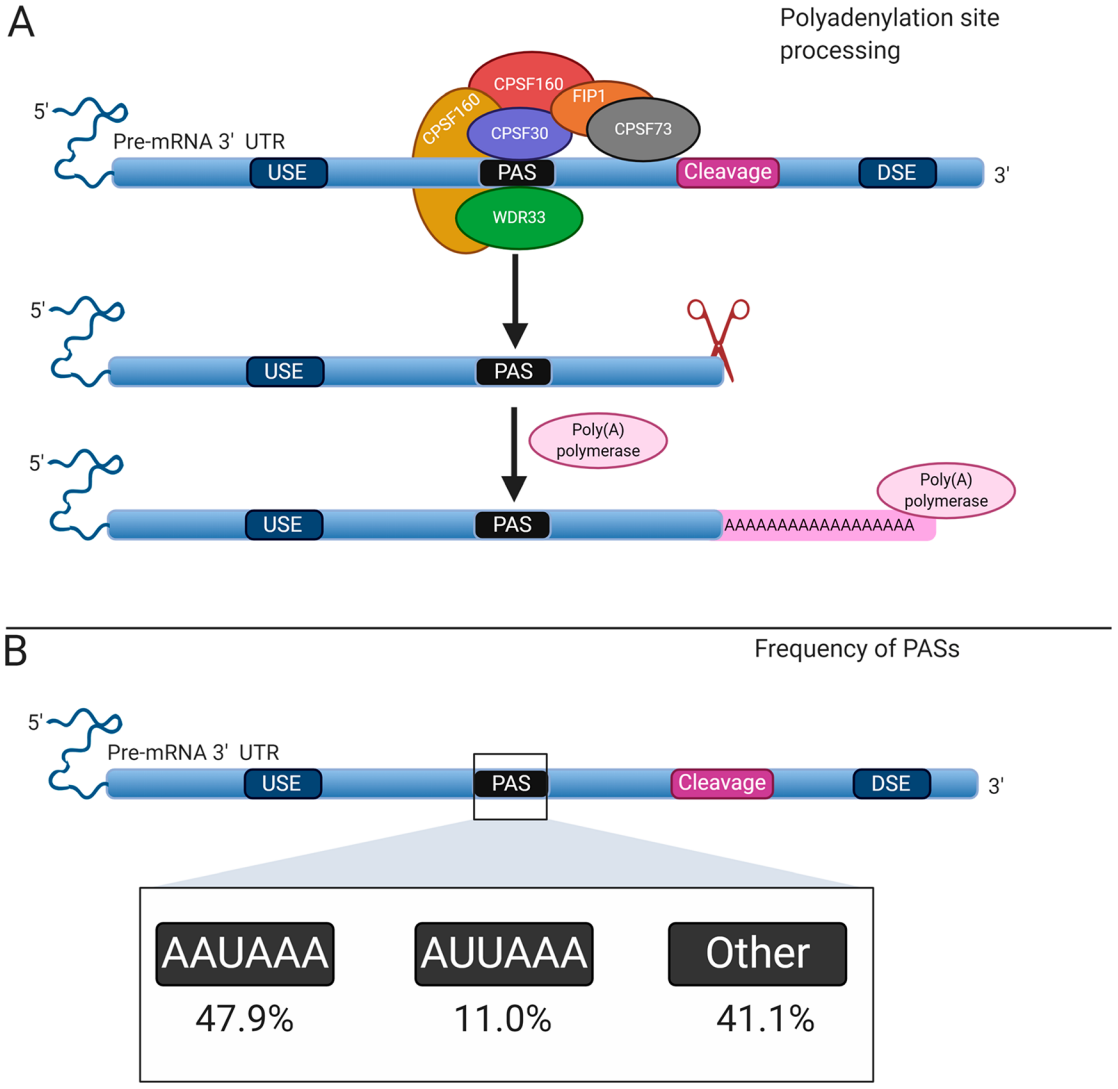


Figure 1. (A) Schematic representation of polyadenylation site recognition and processing. The upstream element or USE is most commonly described as U-rich; the cleavage site most commonly occurs at a cytosine-adenosine dinucleotide sequence located approximately 15−30 nucleotides downstream from the central AU hexamer (PAS), and the downstream element (DSE) is most commonly rich in U or G/U. (B) Polyadenylation site occurrence based on mRNA expression frequencies determined in this study.

often recognize AU-rich RNA target sequences.[3,13,26] AU-rich sequences are present in most pre-mRNAs; therefore, we investigated whether isolated CPSF30 binds to AU-rich RNA sequences. We reported that both CPSF30-FL and CPSF30-5F bind to the AU-rich sequence present within α-synuclein pre-mRNA with high affinity in a cooperative manner.[13,25] These results identified a role for CPSF30 in direct RNA binding.[13,25,27] The AU-rich sequence to which CPSF30 binds in pre-mRNA has the motif AAUAAA. This sequence is called the AU hexamer and is centrally located in the polyadenylation signal (PAS).[19] The AU hexamer is present in a majority of pre-mRNAs, suggesting that this is a general target for CPSF30 (Figure 1B).

In our work to identify the binding partner of CPSF30, we made the discovery that one of the CCCH domains has a 2Fe-2S cluster cofactor in lieu of zinc, with the four remaining CCCH domains binding zinc.[13] This cofactor must be present for optimal RNA binding by CPSF30, in addition to zinc, indicating a functional role for both metals.[13] We also reported that both CPSF30-FL and CPSF30-5F (which contains only the five CCCH domains) bind to the AU hexamer with similar affinities, suggesting that the CCCH domains are important for RNA recognition.[25] Although there are not yet any structures of CPSF30 bound to RNA alone, two cryo-EM structures of the CPSF complex provide additional support for the role of the CCCH domains in binding to the AU hexamer.[28,29] In these structures, only ZF1−ZF3 are visible with ZF2 and ZF3 directly binding to a short fragment of RNA (10 and 17 mer) at adenosines 1/2 and 4/5 of the AAUAAA sequence.[28,29] Notably, ZF2 and ZF3 of CPSF30 are not sufficient to bind to RNA in isolation. We reported that a construct of CPSF30 containing just ZF2 and ZF3 does not bind to the AU hexamer pre-mRNA sequence.[25]

CPSF30 also binds to polyuracil RNA.[16,25] This binding is observed only when the CCHC "zinc knuckle" domain is present. CPSF30-5F, which contains just the five CCCH domains, does not bind polyuracil RNA, indicating that the zinc knuckle present in CPSF30-FL is involved in direct binding.[25] Taken together, these biochemical results reveal that CPSF30 plays a role in pre-mRNA processing by binding to specific pre-mRNA sequences facilitating the transition from pre-mRNA to mRNA.

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

A number of diseases have been linked to aberrant polyadenylation.[20,30−37] Of these diseases, at least five are associated with mutated PAS hexamer sequences within their pre-mRNA sequences. This led us to hypothesize that variants within the PAS hexamer would result in abrogation of binding of CPSF30 to pre-mRNA. To test this hypothesis, we utilized fluorescence anisotropy (FA) to determine the effects of these variants on CPSF30/RNA binding. Herein, we report that certain variants have a profound effect on CPSF30/RNA binding, while others do not. We link these findings to specific hydrogen bonding interactions between amino acid residues on CPSF30 and its partner RNA. We expand these studies to interrogate the general role of the AU hexamer in CPSF30/RNA binding, using both experimental and bioinformatics approaches. Studies to understand how CPSF30 discriminates between the AU hexamer and polyuracil RNA sequences are also described, and a model for CPSF30/AU hexamer/polyU binding is proposed.

MATERIALS AND METHODS

Cloning, Expression, and Purification of Holo-CPSF30-5F and Holo-CPSF30-FL. Cloning, expression, and purification followed our published methods.[13,25,27] Briefly, CPSF30 DNA from the *Bos taurus* homologue was obtained from G. Martin and W. Keller (University of Basel, Basel, Switzerland). CPSF30-5F (UniProtKB O19137) (amino acids 33−170) and full length (UniProtKB O19137) (amino acids 1−243) constructs were cloned into the pMAL-c5e plasmid utilizing NdeI and BamHI restriction sites and confirmed at the Biopolymer-Genomics core facility of the University of Maryland, Baltimore. BL21-DE3 cells were transformed via heat shock with either holo-CPSF30-5F or holo-CPSF30-FL plasmids. Transformed cells were incubated at 37 °C for 45 min. Overnight cultures containing 50 mL of Lennox modified LB broth supplemented with 100 μg/mL ampicillin were inoculated with 150 μL of transformed cells and allowed to grow overnight at 37 °C while being shaken. The following day, 10−15 mL of overnight culture was used to inoculate 4 L culture flasks containing 1 L of Lennox modified LB broth supplemented with 100 μg/mL ampicillin and 0.2% (w/v) glucose. Cell cultures were grown at 37 °C while being shaken to an optical density of approximately 0.3 where flasks were supplemented with 0.8 mM $ZnCl_2$ and 0.6 mM $FeCl_3$. Once cultures reached an optical density between 0.5 and 0.6, they were supplemented with 0.4 mM $Na_2S{\cdot}9H_2O$ and induced with 1 mM IPTG. Flasks were incubated at 37 °C for 3 h while being shaken, centrifuged at 7800*g* for 20 min at 4 °C, and stored at −20 °C.

For purification of holo-CPSF30-5F, pellets were resuspended in 25 mL of a buffer [20 mM Tris and 200 mM NaCl (pH 7.5)]. A protease inhibitor tablet was added to prevent protein degradation. Sonication was utilized for cell lysis, and the cell lysate was then centrifuged at 17710*g* for 20 min at 4 °C. The sonicated supernatant was loaded onto an amylose resin gravity column (15 mL bed volume) and allowed to flow through. The resin was washed four times with 45 mL of cold buffer [20 mM Tris and 200 mM NaCl (pH 7.5)]. Protein was eluted three times with 15 mL of cold elution buffer [20 mM Tris, 200 mM NaCl, and 30 mM maltose (pH 7.5)], producing a total of 45 mL of isolated protein. The ultraviolet−visible spectrum of each elution was recorded, and the protein was concentrated utilizing a 30 kDa molecular weight cutoff spin filter. Holo-CPSF30-5F was buffer exchanged for fluorescence anisotropy studies [20 mM Tris and 100 mM NaCl (pH 8)]. The protein concentration was determined utilizing a calculated extinction coefficient of 85400 M^{-1} cm^{-1} at 278 nm; the purity was assessed with sodium dodecyl sulfate−polyacrylamide gel electrophoresis, and the metal occupancy was determined utilizing ICP-MS. Holo-CPSF30-FL was expressed and purified like holo-CPSF30-5F with the following exceptions. The sonicated supernatant was supplemented with 300 mM NaCl and added to 15 mL of amylose resin. The supernatant and resin were incubated for 15−20 min while being shaken and then allowed to flow through. The salt concentration was decreased with subsequent washes, and protein was eluted utilizing a 20 mM Tris, 200 mM NaCl, and 30 mM maltose buffer (pH 7.5). The calculated extinction coefficient utilized for holo-CPSF30-FL was 88200 M^{-1} cm^{-1}.

Inductively Coupled Plasma Mass Spectrometry (ICP-MS). The protein metal occupancy was determined utilizing ICP-MS as previously described.[25,27] Protein samples were diluted to 1 μM with 6% nitric acid with a final volume of 2 mL. A mixing T was utilized to run the sample and an internal standard containing Rh, Bi, Ge, and Sc to the nebulizer. Samples were measured in He mode to prevent interference with argon oxide.

Fluorescence Anisotropy (FA). An FA binding assay was utilized to measure the affinities of CPSF30 (5F and full length) for RNA variants. RNA oligonucleotides with a 3′ 6-FAM (6-carboxyfluorescein)-conjugated fluorescent molecule were purchased from Sigma at HPLC-purified grade. ISS K2 and PC-1 spectrofluorometers were configured in the L format and utilized for FA binding studies. FA experiments were conducted by employing an excitation wavelength and a slit width of 495 nm and 2 mm, respectively, and an emission wavelength and a slit width of 517 nm and 1 mm, respectively. FA experiments were performed in 20 mM Tris, 100 mM NaCl, 0.3 mg/mL polyC, and 0.1 mg/mL bovine serum albumin (pH 8) with 5 nM fluorescently labeled RNA in 5 mm Spectrosil far-UV quartz window fluorescence cuvettes (Starna Cells). Either CPSF30-FL or CPSF30-5F was titrated into the cuvette until saturation was achieved. The anisotropy and total fluorescence intensity were then recorded. Raw anisotropy values were volume corrected and then corrected for changes in fluorophore quantum yield utilizing the following equation:

$$r_c = \frac{r_0(r_b - r) + rQ(r - r_0)}{r_b - r + Q(r - r_0)}$$

where r_c is the corrected anisotropy, r_0 is the anisotropy of the free fluorescently labeled RNA, r_b is the anisotropy of the protein/RNA complex at saturation, and r is the raw anisotropy.

Data were then analyzed utilizing Graphpad Prism 5 fit to a cooperative binding model as follows:

$$nP + R \rightleftharpoons P_nR$$

$$K = \frac{[P_nR]}{[P]^n[R]}$$

$$r_{tc} = r_0 + (r_b - r_0)\frac{\left(\frac{[P]}{[P]_{1/2}}\right)h}{1\left(\frac{[P]}{[P]_{1/2}}\right)h}$$

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

Table 1. RNA Oligomers Utilized in This Study[a]

RNA Name	RNA Sequence
ASYN-24	CACUUUAAUAAUAAAAAUCAUGCU
ASYN-30	UCUCACUUUAAUAAUAAAAAUCAUGCUUAU
α-thalassemia	UCUCACUUUAAUAAUAAGAAUCAUGCUUAU
β-thalassemia	UCUCACUUUAAUAACAAAAAUCAUGCUUAU
IPEX	UCUCACUUUAAUAAUGAAAAUCAUGCUUAU
LUPUS	UCUCACUUUAAUAAUAGAAAUCAUGCUUAU
FABRY	UCUCACUUUAAUAUUAAGAAUCAUGCUUAU
P2AtoG	UCUCACUUUAAUAGUAAAAAUCAUGCUUAU
P2AtoU	UCUCACUUUAAUAUUAAAAAUCAUGCUUAU
P4AtoC	UCUCACUUUAAUAAUCAAAAUCAUGCUUAU
P5AtoC	UCUCACUUUAAUAAUACAAAUCAUGCUUAU
P6AtoC	UCUCACUUUAAUAAUAACAAUCAUGCUUAU
Most freq fl	UCUCACUUUAAAAAUAAAAAACAUGCUUAU
Least freq fl	UCUCACUUUUCGAAUAAAUCGCAUGCUUAU
ARE-16	AUUAUUUAUUUAUUUA
ARE-18	UUAUUAUUUAUUUAUUUA
ARE-20	UUUAUUAUUUAUUUAUUUAG
ARE-24	AUUUAUUUAUUAUUUAUUUAUUUA
ARE-30	UUUAUUAUUUAUUUAUUAUUUAUUUAUUUA
ARE-38	GUGAUUAUUUAUUAUUUAUUUAUUAUUUAUUUAUUUAG

[a]The polyadenylation sequence is underlined, and PAS variants are colored red.

where h is the Hill coefficient, r_{tc} is the total corrected anisotropy, [P] is the protein concentration, and $[P]_{1/2}$ is the protein concentration at half of the saturation value. All titrations were performed in at least triplicate and conducted in tandem with a positive control of either α-syn$_{24}$ or α-syn$_{30}$ to confirm protein/RNA binding activity.

Fluorescence Anisotropy Competition Assays. A forward titration of CPSF30-FL protein to fluorescently labeled RNA was conducted as described in Fluorescence Anisotropy (FA) to determine protein concentrations needed to reach 75% complex formation. Competition assays were performed by first adding CPSF30-FL protein to achieve a starting anisotropy value equal to approximately 75% complex formation (301 nM protein for the α-syn$_{24}$ complex and 198 nM for the polyU complex). Unlabeled RNA was titrated into the cuvette until anisotropy values reached saturation equal to free fluorescently labeled RNA.

RNA Oligonucleotides. All RNA oligonucleotides were purchased desalted and HPLC purified, and their sequences are listed in Table 1.

PAS Hexamer Bioinformatics. Polyadenylation signal sequences of *Homo sapiens* were obtained from PolyASite Atlas V2.0 (GRCH38.96).[38] Up to 432,444 clusters containing one of the following poly(A) signals residing in a region of 60 nucleotides upstream to 10 nucleotides downstream of a representative poly(A) site of the cluster were included in the analysis (AAUAAA, AUUAAA, UAUAAA, AGUAAA, AAUACA, CAUAAA, AAUAUA, GAUAAA, AAUGAA, AAUAAU, AAGAAA, ACUAAA, AAUAGA, AUUACA, AACAAA, AUUAUA, AACAAG, and AAUAAG), following the same criteria as Herrmann et al.[38] The frequencies of each hexamer were calculated in three forms of analysis: (1) site-based frequency, (2) cluster-based frequency, and (3) mRNA expression-weighted frequency:

$$\text{site-based frequency} = \frac{\text{count of specific hexamer sequence}}{\text{count of all hexamers}}$$

$$\text{cluster-based frequency} = \frac{\text{count of clusters containing a specific hexamer}}{\text{count of all clusters}}$$

$$\text{mRNA expression-weighted frequency} = \frac{\text{sum of products of a specific hexamer and its expression}}{\text{sum of products of all hexamers and their expression}}$$

The data were further analyzed by a sequence logo algorithm to extract motif pattern information.[39]

PAS Hexamer Flanking Region Bioinformatics. Messenger RNA sequences of *H. sapiens* were obtained from the NCBI reference sequence database. The 3′ end sequences of transcripts were parsed by regular expression, and 5′ and 3′ end flanking regions of the canonical polyA signal (AAUAAA) were extracted.[40] Frequencies of each nucleotide at a position in the immediate 5′ and 3′ flanking sequences were calculated by counting its occurrences at each up- and downstream position in all transcripts carrying the conserved AAUAAA hexamer.

RESULTS AND DISCUSSION

Disease-Related PAS Variants Abrogate CPSF30-5F/RNA Binding. The region of pre-mRNA to which CPSF30 binds is called the polyadenylation signal (PAS) (Figure 1A). Within the PAS, there is a conserved sequence called the central hexamer. This sequence is most commonly AAUAAA, and our laboratory has shown that CPSF30 alone binds to this sequence with high affinity in a cooperative manner, when CPSF30 is loaded with both Zn and Fe (2Fe-2S cluster).[13] Similarly, the high affinity binding of CPSF30 for the AU hexanucleotide motif has been reported when the protein is in complex with other CPSF proteins (CPSF160-WDR33-FIP-1-CPSF30).[23,24] Binding of CPSF30 to the PAS is known to be a key step in the maturation of eukaryotic mRNAs in cells, and these studies of CPSF30 at the molecular level underscore the significance of the AU hexamer sequence. There have been several reports of variants within the AU-rich hexanucleotide sequence in diseases that involve improper 3′ end process-

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

Table 2. Sequences of the RNA Oligonucleotides Corresponding to Disease Variants Related to the PAS and the Binding Constants [$P_{1/2}$] and Hill Coefficients Determined from Titrations with CPSF30-5F Fit to a Cooperative Binding Model (±SD)

Name	Sequence $A_1A_2U_3A_4A_5A_6$ (5'→ 3')	Hill coefficient	$[P]_{1/2}$ (nM)
α-syn$_{30}$	UCUCACUUUAAU**AAUAAA**AAUCAUGCUUAU	1.8[a]	178 ± 24[a]
ATHA	UCUCACUUUAAU**AAUAAG**AAUCAUGCUUAU	2.0	135 ± 6.3
BTHA	UCUCACUUUAAU**AACAAA**AAUCAUGCUUAU	2.1	148 ± 9.0
IPEX	UCUCACUUUAAU**AAUGAA**AAUCAUGCUUAU	-	*n.b*
LUPUS	UCUCACUUUAAU**AAUAGA**AAUCAUGCUUAU	-	*n.b*
FABRY	UCUCACUUUAAU**AUUAAG**AAUCAUGCUUAU	-	*n.b*

[a]From ref 25.

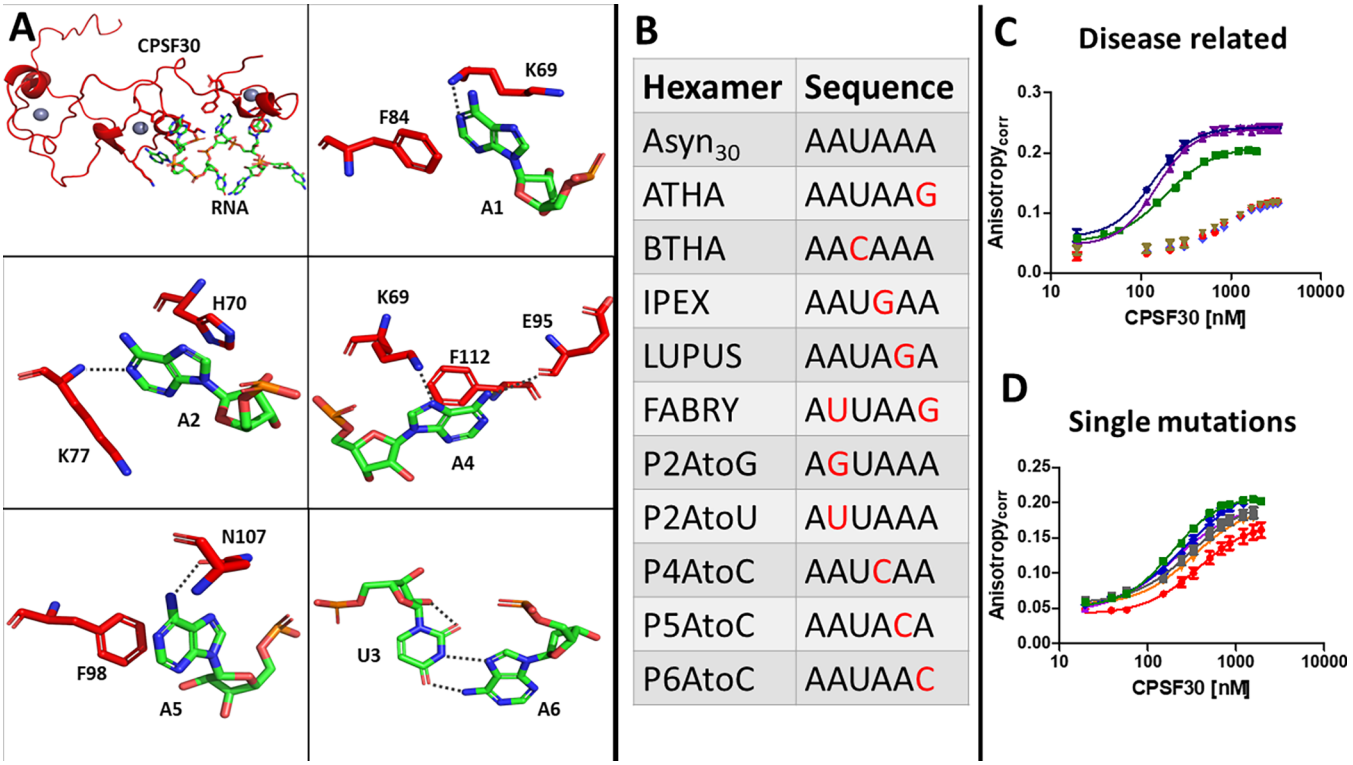


Figure 2. (A) Cryo-EM structures of CPSF160, WDR33, CPSF30, and RNA, focused on CPSF30 (ZF1−ZF3) bound to RNA (sequence of 5′-AACCUCCAAUAAACAAC-3′). Panels show the interactions of each nucleotide of the AAUAAA hexamer with amino acids from CPSF30. Hydrogen bonds between bases and amino acids are shown by yellow dotted lines. This figure was generated in Pymol (Protein Data Bank entry 6BLL). (B) Oligonucleotide sequences of the PAS hexamers investigated in this study with variants from the canonical AU hexamer colored red. (C) Fluorescence anisotropy titrations of PAS RNA variant sequences with CPSF30-5F (α-syn$_{30}$, green squares; ATHA, dark blue circles; BTHA, purple triangles; IPEX, light blue diamonds; LUPUS, red circles; FABRY, brown inverted triangles). (D) Fluorescence anisotropy titrations of single-nucleotide PAS RNA variant sequences with CPSF30-5F (α-syn$_{30}$, green squares; P2AtoG, red circles; P2AtoU, purple triangles; P4AtoC, inverted orange triangles; P5AtoC, dark blue diamonds; P6AtoC, gray squares). An average of three titrations is plotted, and the error is shown as the standard error of the mean (SEM). Data are fit to a cooperative binding model.

ing.[20,30−37] These include α- and β-thalassemia, blood disorders that result in absent or decreased concentrations of α or β globin chains;[31,32] IPEX (polyendocrinopathy, enteropathy, X-linked inheritance) syndrome, a disease that stems from a PAS variant in the FOXP3 gene leading to immunodeficiencies;[33] Fabry disease, which results from a disruption in glycosphingolipid catabolism due to a PAS variant in the α-galactosidase A gene;[34] and lupus, an autoimmune disease that involves a PAS variant in the GIMAP5 gene.[37] We sought to determine if these disease states are manifest by abrogation of binding between the CCCH domains of CPSF30 and the mutated PAS AU hexanucleotide RNA. Utilizing a construct of CPSF30 that contains just the five CCCH domains (CPSF30-5F), which we previously reported to selectively recognize the AU hexamer when loaded with Zn and a 2Fe-2S cluster,[13] we measured the affinity of CPSF30-5F for five disease-related RNA PAS variants [α-thalassemia (ATHA, sequence of AAUAA**G**), β-thalassemia (BTHA, sequence of AA**C**AAA), IPEX (AAU-**G**AA), Fabry (A**U**UAA**G**), and Lupus (AAUA**G**A)]. The experiments utilized fluorescence anisotropy (FA) and involved titrating CPSF30-5F with fluorescently labeled RNA molecules that correspond to the PAS variants (Table 2).[25] As Table 2 and Figure 2 show, CPSF30-5F/RNA binding was abrogated for three of the variants: IPEX, Fabry, and Lupus. In contrast, binding was retained for the α- and β-thalassemia variants and was of the same order of magnitude as that observed for the native PAS sequence (Table 2). All of the

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

Table 3. RNA Oligonucleotide Sequences of Variants within the PAS and Binding Constants [$P_{1/2}$] and Hill Coefficients Determined from Titrations with CPSF30-5F Fit to a Cooperative Binding Model (±SD)

Name	Sequence (5'→ 3')	Hill Coefficient	$[P]_{1/2}$ (nM)
P2AtoG	UCUCACUUUAAUA**G**UAAAAAUCAUGCUUAU	1.4	376 ± 12
P2AtoU	UCUCACUUUAAUA**U**UAAAAAUCAUGCUUAU	1.2	207 ± 0.60
P4AtoC	UCUCACUUUAAUAA**C**AAAAUCAUGCUUAU	1.4	335 ± 38
P5AtoC	UCUCACUUUAAUAAUA**C**AAAUCAUGCUUAU	1.2	278 ± 29
P6AtoC	UCUCACUUUAAUAAUAA**C**AAUCAUGCUUAU	1.4	309 ± 26

variants evaluated were single-base variants within the AU hexamer sequence, except for the Fabry variant that had two base modifications: positions 2 and 6 (P2AtoU and P6AtoG, respectively). Notably, these single variants in positions 2 and 6 alone do not affect CPSF30-5F/RNA binding, indicating that variants at both positions within the hexamer must be present to have an effect on binding (Tables 2 and 3). Taken together, we suggest that the loss of binding affinity of CPSF30s CCCH ZFs to the IPEX, Fabry, and Lupus disease-related RNA variants may contribute to the mechanisms of these diseases, whereas CPSF30 may not play a key role in ATHA or BTHA.

To better understand why the α- and β-thalassemia variants (ATHA and BTHA, respectively) did not exhibit loss of binding to CPSF30-5F, we mapped the mutated ATHA and BTHA RNA sequences onto a recent cryo-EM structure of the CPSF complex. In these structures, CPSF30 is partially visible with ZF1−ZF3 present. CPSF30s CCCH ZF domains 2 and 3 interact directly with adenosines 1, 2, 4, and 5 of the hexamer (AAUAAA).[28,29] The ATHA and BTHA variants are in positions 3 and 6 of the hexamer, which are not involved in direct binding with CPSF30 (Figure 2A). Positions 3 and 6 are recognized by WDR33 in the CPSF complex; therefore, variants in this region likely affect other partners in the CPSF complex.[28,29]

Effects of Single-PAS Variants on CPSF30-5F/RNA Binding. The finding that single-nucleotide variations within the AU hexamer RNA sequence can abrogate binding between CPSF30-5F and RNA led to the question of the importance of sequence conservation within the AU hexamer for CPSF30 binding. To address this question, we examined variants in positions 4 and 5 of the AU hexamer (Table 2). These positions were chosen because the disease variants revealed that A to G transitions (purine to purine) in those positions abrogated binding. The cryo-EM structure indicates that the amine groups on A4 and A5 hydrogen bond to E95 and N107 amino acids in CPSF30, respectively (Figure 2). The variant chosen was an A to C variant, or a purine to pyrimidine change within the hexamer. This variant is predicted to retain hydrogen bonding via an amine group, and we hypothesized that CPSF30-5F/RNA binding would be retained. This prediction was confirmed by the FA data: CPSF30-5F bound to the PAS RNA sequences with the A to C variations in positions 4 and 5 of the AU hexamer, albeit with slightly weaker affinities compared to those measured for the native RNA target sequence (Table 3). These results support a role for hydrogen bonding in high-affinity binding of CPSF30 to the PAS hexamer.

We also examined a P2AtoU variation. This variant was chosen because AUUAAA is the second most common PAS hexamer at ∼10% of human polyA selection sites, and we predicted that the variant would not affect CPSF30/PAS RNA binding. The FA data supported this prediction: the AUUAAA sequence exhibited CPSF30 binding comparable to that of the AAUAAA sequence ($[P]_{1/2}$ = 207 ± 0.60 nM vs 178 ± 24 nM) (Table 3). We note that in nuclear extracts the AUUAAA polyadenylation site exhibits 77% of the processing efficiency, which is defined by Sheets et al. as the percentage of the polyadenylated product of a mutated PAS as compared to the canonical site AAUAAA, suggesting a potential connection between binding affinity and processing efficiency (Table S1).[22] The affinity of CPSF30 for a P2AtoG variant was also examined. Here, the prediction was that the variants would abrogate binding because this same A to G variant in positions 4 and 5 of the hexamer abrogates binding; however, the P2AtoG variants retained RNA binding, although it was weaker than that measured for the native RNA target (Table 3). This suggests that the identity of the base in position 2 can be variable. On the cellular level, there is a body of literature that has measured polyadenylation efficiency mediated by CPSF30 (as part of a nuclear extract that contains the other CSPF proteins) as a function of RNA sequence.[22] In these studies, it has been reported that variants of the AAUAAA hexamer to guanine in positions 3−5 significantly attenuate polyadenylation efficiency (<6% of the polyadenylation efficiency of AAUAAA), whereas, this same guanine variant in position 2 (P2AtoG transition) decreases the polyadenylation efficiency to only 29% of the efficiency compared to that of the canonical AAUAAA hexamer. These data suggest that the identity of the base in position 2 is not as important for function.[22] Thus, the finding that the P2AtoG transition does not severely affect CPSF30/RNA binding may be connected to polyadenylation efficiency.

Bioinformatic Analysis of PAS Hexamers to Identify Common Sequence Variations. The RNA binding studies revealed that CPSF30 will bind to sequences that are modified from the canonical AAUAAA hexamer PAS, and we sought to define the common sequence variations that are present within the PAS hexamer sequence. The PAS is known to have some sequence variability, and there have been several studies that utilized bioinformatics to identify this variability.[41−44] Building upon these studies, we applied a bioinformatics approach to identify the PAS site variability in pre-mRNAs from *H. sapiens*. We utilized the PolyASite 2.0 database, developed by Herrmann et al.,[38] to identify the most abundant AU hexamer type sequences. The pre-mRNA sequences that were parsed in our bioinformatics analysis were defined as those between nucleotides −60 to +10 relative to the polyA site.[38] In some cases, more than one PAS was found in this region. In this case, the most defined site was called the "representative" site, and the remaining signals were called "clustered" as described by Herrmann et al.[38] The results were grouped in three manners: (i) site-based frequency, (ii) cluster-based frequency, and (iii)

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

Table 4. Heat Maps Generated from Bioinformatic Analyses of the Most Frequent PAS Hexamer Sequences in *H. sapiens* Based upon Site-Based, Cluster-Based, and mRNA Expression-Weighted Frequencies[a]

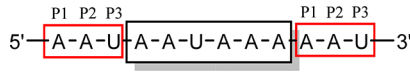
PAS Signal	Site based frequency	Cluster based frequency	mRNA expression weighted frequency
Sequence logo			
AAUAAA	0.2977	0.5885	0.4790
AUUAAA	0.1031	0.2039	0.1096
AAGAAA	0.0795	0.1572	0.0433
AAUAAU	0.0591	0.1168	0.0420
AACAAA	0.0587	0.1160	0.0375
UAUAAA	0.0581	0.1148	0.0396
AAUAUA	0.0511	0.1010	0.0284
AAUGAA	0.0429	0.0848	0.0412
AGUAAA	0.0360	0.0712	0.0282
AAUACA	0.0320	0.0632	0.0121
AUUAUA	0.0309	0.0611	0.0185
CAUAAA	0.0296	0.0586	0.0215
GAUAAA	0.0245	0.0483	0.0143
ACUAAA	0.0234	0.0462	0.0229
AAUAGA	0.0204	0.0403	0.0114
AAUAAG	0.0200	0.0396	0.0127
AUUACA	0.0176	0.0348	0.0121
AACAAG	0.0153	0.0302	0.0159

[a]A heat map is shown going from most frequent to least frequent (red to yellow to green, respectively).

mRNA expression-weighted frequency (Table 4 and PAS hexamer analysis XLSX file). The site-based frequency is the ratio of each hexamer's occurrence in all signals found; the cluster-based frequency is the fraction of clusters containing the hexamer, and the mRNA expression-weighted frequency is the fraction of the sequence present in the total number of mRNAs that are expressed. The sequences identified from our analysis are listed in Table 4. Each PAS hexamer sequence is presented with a value assigned on the basis of the type of grouping. A value of 1 would mean that the sequence is present 100% of the time. The most abundant sequences are colored red, and the least abundant sequences are colored green, with gradations between the two colors reflective of the frequency of each sequence. Notably, in all three groupings, the most abundant sequence is the canonical hexamer AAUAAA as predicted. Sequence logos were generated for each grouping, shown at the top of Table 4. In these logos, the base at position 2 is the least conserved, which is consistent with the CPSF30/RNA binding data whereby variants at this position have little effect on binding affinity (*vide supra*).

Bioinformatics of AU Hexamer Flanking Regions. The α-synuclein pre-mRNA sequence utilized in these studies of CPSF30/RNA binding contains AU sequences at the immediate 3′ and 5′ flanking regions (AAU). Inspection of other pre-mRNA sequences revealed that AU-rich sequences often extend beyond the AU hexamer. This suggests that pre-mRNA sequences have a built-in redundancy. This sequence redundancy could be important for CPSF30/RNA binding or for mitigating any effects of variants to the AU hexamer motif on CPSF30/RNA binding. To determine whether our observation that AU sequences often extend beyond the AU hexamer is a general feature of all pre-mRNAs, we performed a bioinformatic analysis of the 5′ and 3′ flanking regions of the AAUAAA PAS hexamer. A total of 62391 human gene transcript variants from the NCBI nonredundant database were examined, and the frequency with which each of the four bases occurred in each flanking region was determined. As shown in Table 5 and Figure S1, adenosines and uracils

Table 5. (Top) Sequences of the PAS with the Nucleotide Positions of the 5′ and 3′ Flanking Regions Indicated and (Bottom) Comparison of the Frequency with Which Each Nucleotide in the Immediate 5′ and 3′ Flanking Regions of the PAS Occurs[a]



	P1	P2	P3
5′ Flanking Region			
A	0.30	0.31	0.43
U	0.35	0.36	0.24
C	0.17	0.15	0.21
G	0.17	0.17	0.12
3′ Flanking Region			
A	0.39	0.32	0.27
U	0.23	0.32	0.37
C	0.15	0.16	0.18
G	0.23	0.19	0.17

[a]Data from a bioinformatic analysis utilizing the NCBI nonredundant database.

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

Table 6. RNA Oligonucleotide Sequences Derived from a Bioinformatic Analysis of the 5′ and 3′ Flanking Regions of the PAS and Their Apparent Dissociation Constants (±SD) with CPSF30-5F

Name	Sequence (5'→ 3')	Hill coefficient	CPSF30-5F $[P]_{1/2}$ (nM)
Most freq fl	UCUCACUUUAA**AAUAAA**AAACAUGCUUAU	2.0	226 ± 1.6
Least freq fl	UCUCACUUU**UCG****AAUAAA****UCG**CAUGCUUAU	1.5	309 ± 25

Table 7. RNA Oligonucleotide Sequences from the ARE Region of TNF-α and Their Relative Dissociation Constants (±SD) with CPSF30-5F

Name	Sequence (5'→ 3')	Hill coefficient	$[P]_{1/2}$ (nM)
ARE-16	AUUAUUUAUUUAUUUA	3.6	1000 ± 82
ARE-18	UUAUUAUUUAUUUAUUUA	2.1	671 ± 49
ARE-20	UUUAUUAUUUAUUUAUUUAG	2.1	499 ± 37
ARE-24	AUUUAUUUAUUAUUUAUUUAUUUA	1.8	265 ± 31
ARE-30	UUUAUUAUUUAUUUAUUAUUUAUUUAUUUA	3.0	163 ± 10
ARE-38	GUGAUUAUUUAUUAUUUAUUUAUUAUUUAUUUAUUUAG	2.5	141 ± 5.0

occurred on average twice as frequently as cytosines and guanines in the flanking regions. For the 5′ and 3′ flanking regions, AAA was the most frequent flanking region and UCG was the least frequent (PAS flanking region bioinformatics XLSX file).

The higher frequency with which the adenine and uracil residues are present in the AU hexamer flanking regions compared to guanine and cytosine suggests that they may contribute to CPSF30/RNA binding. To test this hypothesis, fluorescence anisotropy (FA) binding assays were conducted for CPSF30-5F with α-syn_{30} variants in which the flanking regions of α-syn_{30} were mutated to AAA and UCG. CPSF30-5F bound to the α-syn_{30} AAA flanking region variant with an affinity of 226 ± 1.6 nM (compared to 178 ± 24 nM for native α-syn_{30}), while the affinity of CPSF30-5F α-syn_{30} for the UCG flanking region variant was weaker, 309 ± 25 nM (Table 6). These data support a role for flanking region AU sequence in CPSF30/pre-mRNA binding. We propose that the AU-rich flanking sequences may serve as redundant sequences either to ensure PAS recognition or to fine-tune the efficiency of one PAS versus another.

CPSF30-5F Binds to AU-Rich Element (ARE) RNA Sequences. CPSF30 belongs to the CCCH class of ZF proteins. The best studied protein in this class is tristetraprolin (TTP), which regulates mRNA transcripts by selectively recognizing and binding to the mRNA sequence UUUA-UUUAUUU (also called the ARE), which is present in the 3′ untranslated region of mRNA.[15,45] This ZF/RNA binding event is important for regulating the stability of the mRNA and its ability to be marked for rapid degradation during stress responses.[10,12,26,46] Given the high degree of sequence homology between the CCCH ZF domains of CPSF30 and TTP, we predicted that CPSF30 would also bind to the ARE sequence (Figure S2). To test this hypothesis, the affinities of CPSF30-5F for a series of oligonucleotides with the ARE sequence (from TNF-α mRNA) of various lengths (16−38 nucleotides) were measured. Tight binding, on par with that reported for the AU hexamer, was observed for oligonucleotides with a length of ≥24 nucleotides (Table 7 and Figure S3). These data contrast with our previously obtained data for non-ARE sequences. CPSF30-5F does not bind to control RNAs, including GU-rich, polyU, polyC, Rβ_{31}, and others.[13,25] The finding that both CPSF30 and TTP selectively recognize AU-rich RNA sequences suggests that other ZFs that contain CCCH domains may also preferentially bind AU-rich RNA sequences. Future experiments will test this hypothesis.

Addition of the Zinc Knuckle Domain to CPSF30 Restores RNA Binding. Full length CPSF30 contains a "CCHC" zinc knuckle domain at the C-terminus, in addition to the five CCCH domains that make up CPSF30-5F. Zinc knuckle domains often recognize U-rich sequences, and we have reported that full length CPSF30 binds to a polyU RNA sequence, in addition to the AU hexamer.[25] We have also found that when four of six residues within the hexamer (CCUCCA) are modified in the context of α-syn, full length CPSF30 still binds to PAS RNA, albeit with weaker affinity. This was attributed to the presence of a polyU sequence outside of the PAS hexamer of α-syn RNA.[25] On the basis of these results, we proposed that the loss of RNA binding observed here for CPSF30-5F with the PAS variants would be recovered upon addition of the zinc knuckle domain to the protein construct (CPSF30-FL). FA experiments confirmed this hypothesis; as shown in Table S2, CPSF30-FL bound to all of the RNA variants. Pre-mRNA sequences contain U-rich motifs at the 3′ end, in addition to the AU hexamer sequence, and the RNA binding observed for full length CPSF30 is likely due to the zinc knuckle domain binding to the U-rich sequence present in α-syn RNA.

The finding that full length CPSF30 binds to both an AU hexamer sequence and a polyU sequence raises the question of whether one sequence is preferentially recognized over the other. FA-monitored competition experiments were performed to determine whether there is a preference. In the first experiment, the CPSF30-FL/α-syn_{24} RNA complex was formed (with fluorescently labeled α-syn_{24}) after which nonfluorescently labeled polyU RNA was titrated and the effect on fluorescence anisotropy was measured. In the second experiment, the CPSF30-FL/polyU RNA complex was formed, nonfluorescently labeled α-syn_{24} RNA was titrated, and the effect on fluorescence anisotropy was measured. As shown in Figure 3 and Figure S4, significantly less polyU RNA has to be titrated with CPSF30-FL/α-syn_{24} (20 nM) than AU hexamer

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

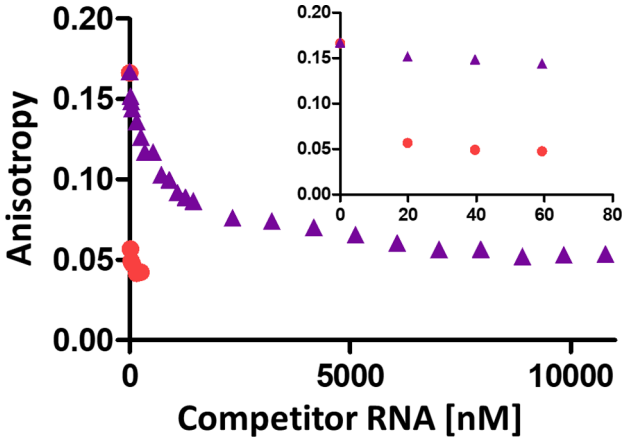


Figure 3. Competitive titration of unlabeled α-syn$_{24}$ with 5 nM polyU-F and 198 nM CPSF30-FL (purple) and competitive titration of unlabeled polyU with 5 nM α-syn$_{24}$-F and 301 nM CPSF30-FL (red). Experiments were performed in 20 mM Tris, 100 mM NaCl, 0.3 mg/mL polyC, and 0.1 mg/mL BSA (pH 8). The inset shows a close-up of titration data between 0 and 80 nM titrant RNA.

(α-syn$_{24}$) with CPSF30-FL/polyU (7015 nM) to displace the starting RNA sequence. These results reveal that CPSF30-FL preferentially binds to the polyU RNA.

CONCLUSIONS

CPSF30 is a multidomain ZF protein that contains both CCCH and CCHC domains, along with an Fe-S cofactor. CPSF30 binds to two highly conserved pre-mRNA sequences: AAUAAA (the PAS hexamer) and U-rich (polyuracil). We have identified single-nucleotide variants within the PAS hexamer that severely attenuate and, in some cases, abrogate CPSF30-5F/RNA binding. Some of the variants are associated with disease states that involve altered PASs. These findings suggest a potential link between CPSF30 function and these diseases. We have also demonstrated that CPSF30 binds to RNA sequences that are rich in adenines and uracils, with high affinity and selectivity. These results support a growing body of evidence that CCCH type ZFs preferentially bind AU-rich RNA sequences. As such, we predict that newly identified CCCH type ZFs will be found to target AU-rich RNA sequences when they are isolated and studied experimentally.

The addition of the C-terminal zinc knuckle domain to CPSF30-5F (forming CPSF30-FL) restores binding to the pre-mRNA sequences for which the PAS hexamer has been mutated. The zinc knuckle domain recognizes a polyU sequence on pre-mRNA, and these results suggest a hierarchy in RNA binding by CPSF30, whereby polyU binding is preferred over PAS binding. We propose that the hierarchy of

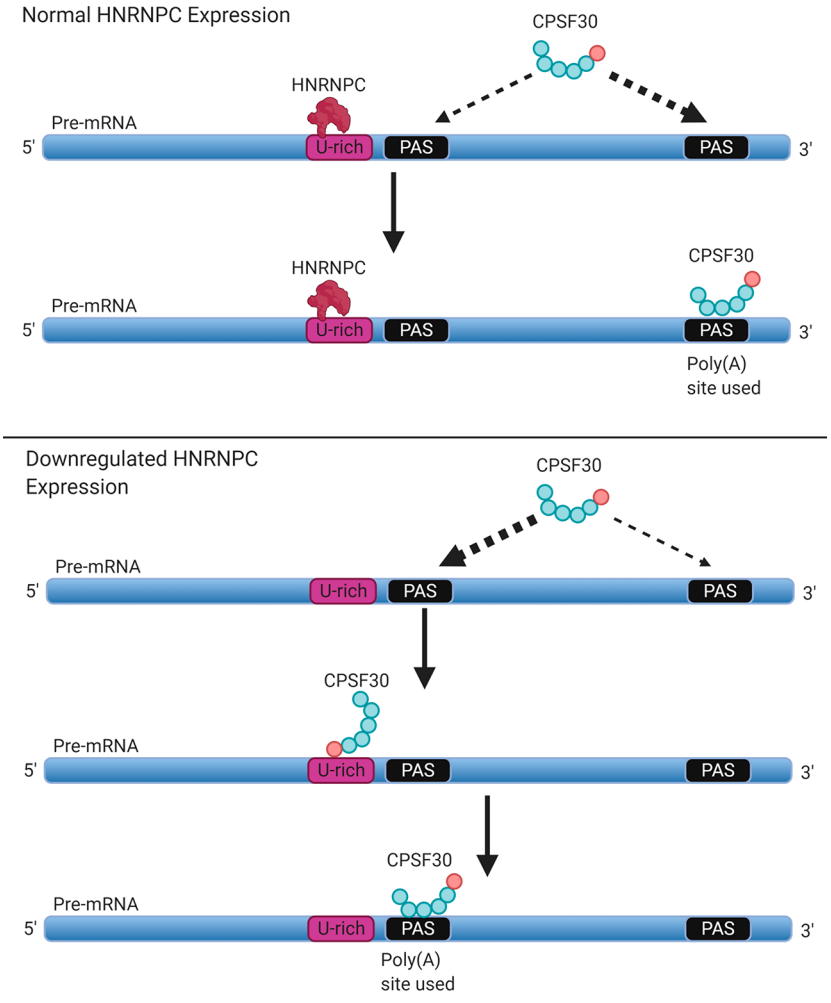


Figure 4. Possible mechanism by which U-rich binding protein HNRNPC downregulation leads to increased PAS site usage with U-rich regions in the proximity of the PAS.

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790

RNA binding observed *in vitro* is connected to the accessibility of specific sequence motifs within the pre-mRNA targets of CPSF30. During pre-mRNA processing, U-rich RNA binding proteins are often expressed and regulate pre-mRNA via direct protein−RNA binding interactions. When these U-rich RNA binding proteins are present and bound to pre-mRNA, U-rich sequences will be inaccessible for CPSF30 binding. Consequently, CPSF30 will bind to the AU hexamer of the PAS. This hypothesis has support in the biological literature. For example, when the polyU binding protein HNRNPC is absent (downregulated), nearby PAS sites with U-rich motifs are used more often in pre-mRNA processing, suggesting that CPSF30 is involved in binding to these motifs.[42] Taken together, we propose a model whereby U-rich binding proteins block binding of CPSF30 to the polyU sequence by rendering the polyU sequence inaccessible, thereby promoting PAS binding. When the U-rich binding proteins are downregulated, the polyU sequence is made accessible to CPSF30 and direct binding to polyU occurs (Figure 4). In addition, during oogenesis, CPSF30 has been found to be involved in cytoplasmic polyadenylation. For this process to occur efficiently, both AAUAAA and a U-rich motif are needed, suggesting an additional role of multi-RNA binding by CPSF30.[47−50] Future studies in cells will test these predictions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.0c00940.

Bar graph showing the flanking region FA data, comparison of CPSF30 to TTP sequences, bar graph describing the ARE FA data, and plot describing further polyU and α-syn$_{24}$ competition studies (Figures S1−S4, respectively) and comparison of this paper to the work on polyadenylation efficiency by Sheets et al. and data from FA studies with PAS variants and full length CPSF30 investigated in this work (Tables S1 and S2, respectively) (PDF)

Tables providing additional details of the PAS hexamer bioinformatics (XLSX)

Tables providing additional details of the PAS flanking region bioinformatics (XLSX)

Accession Codes

CPSF30, UniProt O19137.

■ AUTHOR INFORMATION

Corresponding Author

Sarah L. J. Michel − *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201-1180, United States;* orcid.org/0000-0002-6366-2453; Phone: (410) 706-7038; Email: smichel@rx.umaryland.edu; Fax: (410) 706-5017

Authors

Jordan D. Pritts − *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201-1180, United States;* orcid.org/0000-0002-2806-9849

Abdulafeez A. Oluyadi − *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201-1180, United States*

Weiliang Huang − *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201-1180, United States*

Geoffrey D. Shimberg − *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201-1180, United States*

Maureen A. Kane − *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201-1180, United States;* orcid.org/0000-0002-5525-9170

Angela Wilks − *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201-1180, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.0c00940

Funding

S.L.J.M. is grateful for the financial support from the National Science Foundation (CHE-1708732), and J.D.P. acknowledges a National Institutes of Health-sponsored CBI training grant (GM066706) and the AFPE for support. Additional support was provided by the University of Maryland School of Pharmacy Mass Spectrometry Center (SOP1841-IQB2014).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Biorender.com was used to design the TOC graphic, Figure 1, and Figure 4.

■ REFERENCES

(1) Decaria, L., Bertini, I., and Williams, R. J. (2010) Zinc proteomes, phylogenetics and evolution. *Metallomics 2*, 706−709.

(2) Kluska, K., Adamczyk, J., and Krężel, A. (2018) Metal binding properties, stability and reactivity of zinc fingers. *Coord. Chem. Rev. 367*, 18−64.

(3) Lee, S. J., and Michel, S. L. J. (2014) Structural Metal Sites in Nonclassical Zinc Finger Proteins Involved in Transcriptional and Translational Regulation. *Acc. Chem. Res. 47*, 2643−2650.

(4) Searles, M. A., Lu, D., and Klug, A. (2000) The role of the central zinc fingers of transcription factor IIIA in binding to 5 S RNA. *J. Mol. Biol. 301*, 47−60.

(5) Klug, A. (2010) The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annu. Rev. Biochem. 79*, 213−231.

(6) Krishna, S. S., Majumdar, I., and Grishin, N. V. (2003) Structural classification of zinc fingers. *Nucleic Acids Res. 31*, 532−550.

(7) Matthews, J. M., and Sunde, M. (2002) Zinc Fingers−Folds for Many Occasions. *IUBMB Life 54*, 351−355.

(8) Maret, W. (2012) New perspectives of zinc coordination environments in proteins. *J. Inorg. Biochem. 111*, 110−116.

(9) Maret, W. (2004) Zinc and sulfur: a critical biological partnership. *Biochemistry 43*, 3301−3309.

(10) Fu, M., and Blackshear, P. J. (2017) RNA-binding proteins in immune regulation: a focus on CCCH zinc finger proteins. *Nat. Rev. Immunol. 17*, 130−143.

(11) Maeda, K., and Akira, S. (2017) Regulation of mRNA stability by CCCH-type zinc-finger proteins in immune cells. *Int. Immunol. 29*, 149−155.

(12) Lai, W. S., Carballo, E., Strum, J. R., Kennington, E. A., Phillips, R. S., and Blackshear, P. J. (1999) Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA. *Mol. Cell. Biol. 19*, 4311−4323.

(13) Shimberg, G. D., Michalek, J. L., Oluyadi, A. A., Rodrigues, A. V., Zucconi, B. E., Neu, H. M., Ghosh, S., Sureschandra, K., Wilson, G. M., Stemmler, T. L., and Michel, S. L. J. (2016) Cleavage and polyadenylation specificity factor 30: An RNA-binding zinc-finger protein with an unexpected 2Fe−2S cluster. *Proc. Natl. Acad. Sci. U. S. A. 113*, 4700−4705.
(14) Kafasla, P., Skliris, A., and Kontoyiannis, D. L. (2014) Post-transcriptional coordination of immunological responses by RNA-binding proteins. *Nat. Immunol. 15*, 492−502.
(15) diTargiani, R. C., Lee, S. J., Wassink, S., and Michel, S. L. (2006) Functional characterization of iron-substituted tristetraprolin-2D (TTP-2D, NUP475−2D): RNA binding affinity and selectivity. *Biochemistry 45*, 13641−13649.
(16) Barabino, S. M., Hubner, W., Jenny, A., Minvielle-Sebastia, L., and Keller, W. (1997) The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev. 11*, 1703−1716.
(17) Yang, Q., and Doublié, S. (2011) Structural biology of poly(A) site definition. *Wiley Interdiscip. Rev.: RNA 2*, 732−747.
(18) Chan, S. L., Huppertz, I., Yao, C., Weng, L., Moresco, J. J., Yates, J. R., 3rd, Ule, J., Manley, J. L., and Shi, Y. (2014) CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. *Genes Dev. 28*, 2370−2380.
(19) Proudfoot, N. J. (2011) Ending the message: poly(A) signals then and now. *Genes Dev. 25*, 1770−1782.
(20) Chang, J. W., Yeh, H. S., and Yong, J. (2017) Alternative Polyadenylation in Human Diseases. *Endocrinol. Metab. 32*, 413−421.
(21) Thore, S., and Fribourg, S. (2019) Structural insights into the 3′-end mRNA maturation machinery: Snapshot on polyadenylation signal recognition. *Biochimie 164*, 105−110.
(22) Sheets, M. D., Ogg, S. C., and Wickens, M. P. (1990) Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res. 18*, 5799−5805.
(23) Hamilton, K., Sun, Y., and Tong, L. (2019) Biophysical characterizations of the recognition of the AAUAAA polyadenylation signal. *RNA 25*, 1673−1680.
(24) 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.
(25) Pritts, J. D., Hursey, M. S., Michalek, J. L., Batelu, S., Stemmler, T. L., and Michel, S. L. J. (2020) Unraveling the RNA Binding Properties of the Iron−Sulfur Zinc Finger Protein CPSF30. *Biochemistry 59*, 970−982.
(26) Lai, W. S., Wells, M. L., Perera, L., and Blackshear, P. J. (2019) The tandem zinc finger RNA binding domain of members of the tristetraprolin protein family. *Wiley Interdiscip. Rev.: RNA 10*, e1531.
(27) Shimberg, G. D., Pritts, J. D., and Michel, S. L. J. (2018) Iron-Sulfur Clusters in Zinc Finger Proteins. *Methods Enzymol. 599*, 101−137.
(28) 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.
(29) 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. U. S. A. 115*, E1419−E1428.
(30) Danckwardt, S., Hentze, M. W., and Kulozik, A. E. (2008) 3′ end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J. 27*, 482−498.
(31) Higgs, D. R., Goodbourn, S. E., Lamb, J., Clegg, J. B., Weatherall, D. J., and Proudfoot, N. J. (1983) Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature 306*, 398−400.
(32) Orkin, S. H., Cheng, T. C., Antonarakis, S. E., and Kazazian, H. H., Jr. (1985) Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *EMBO J. 4*, 453−456.
(33) Bennett, C. L., Brunkow, M. E., Ramsdell, F., O'Briant, K. C., Zhu, Q., Fuleihan, R. L., Shigeoka, A. O., Ochs, H. D., and Chance, P. F. (2001) A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA−>AAUGAA) leads to the IPEX syndrome. *Immunogenetics 53*, 435−439.
(34) Yasuda, M., Shabbeer, J., Osawa, M., and Desnick, R. J. (2003) Fabry disease: novel alpha-galactosidase A 3′-terminal mutations result in multiple transcripts due to aberrant 3′-end formation. *Am. J. Hum. Genet. 73*, 162−173.
(35) Turner, R. E., Pattison, A. D., and Beilharz, T. H. (2018) Alternative polyadenylation in the regulation and dysregulation of gene expression. *Semin. Cell Dev. Biol. 75*, 61−69.
(36) Tian, B., and Manley, J. L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol. 18*, 18−30.
(37) Hellquist, A., Zucchelli, M., Kivinen, K., Saarialho-Kere, U., Koskenmies, S., Widen, E., Julkunen, H., Wong, A., Karjalainen-Lindsberg, M. L., Skoog, T., Vendelin, J., Cunninghame-Graham, D. S., Vyse, T. J., Kere, J., and Lindgren, C. M. (2007) The human GIMAP5 gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus. *J. Med. Genet. 44*, 314−321.
(38) Herrmann, C. J., Schmidt, R., Kanitz, A., Artimo, P., Gruber, A. J., and Zavolan, M. (2020) PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3′ end sequencing. *Nucleic Acids Res. 48*, D174−D179.
(39) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res. 14*, 1188−1190.
(40) Stephens, S. M., Chen, J. Y., Davidson, M. G., Thomas, S., and Trute, B. M. (2005) Oracle Database 10g: a platform for BLAST search and Regular Expression pattern matching in life sciences. *Nucleic Acids Res. 33*, D675−D679.
(41) 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.
(42) Gruber, A. J., Schmidt, R., Gruber, A. R., Martin, G., Ghosh, S., Belmadani, M., Keller, W., and Zavolan, M. (2016) 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*, 1145−1159.
(43) Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J. Y., Yehia, G., and Tian, B. (2013) Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nat. Methods 10*, 133−139.
(44) Tian, B., and Graber, J. H. (2012) Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev.: RNA 3*, 385−396.
(45) Carballo, E., Lai, W. S., and Blackshear, P. J. (1998) Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Science 281*, 1001−1005.
(46) Brewer, B. Y., Malicka, J., Blackshear, P. J., and Wilson, G. M. (2004) RNA sequence elements required for high affinity binding by the zinc finger domain of tristetraprolin: conformational changes coupled to the bipartite nature of Au-rich MRNA-destabilizing motifs. *J. Biol. Chem. 279*, 27870−27877.
(47) Dickson, K. S., Thompson, S. R., Gray, N. K., and Wickens, M. (2001) Poly(A) polymerase and the regulation of cytoplasmic polyadenylation. *J. Biol. Chem. 276*, 41810−41816.
(48) Dickson, K. S., Bilger, A., Ballantyne, S., and Wickens, M. P. (1999) The cleavage and polyadenylation specificity factor in Xenopus laevis oocytes is a cytoplasmic factor involved in regulated polyadenylation. *Mol. Cell. Biol. 19*, 5707−5717.
(49) Bilger, A., Fox, C. A., Wahle, E., and Wickens, M. (1994) Nuclear polyadenylation factors recognize cytoplasmic polyadenylation elements. *Genes Dev. 8*, 1106−1116.
(50) de Moor, C. H., and Richter, J. D. (1999) Cytoplasmic polyadenylation elements mediate masking and unmasking of cyclin B1 mRNA. *EMBO J. 18*, 2294−2303.

https://dx.doi.org/10.1021/acs.biochem.0c00940
Biochemistry 2021, 60, 780−790