Studies with recombinant U7 snRNP demonstrate that CPSF73 is both an endonuclease and a 5′–3′ exonuclease

XIAO-CUI YANG,1,4 YADONG SUN,2,4 WEI SHEN AIK,2,5 WILLIAM F. MARZLUFF,1,3 LIANG TONG,2 and ZBIGNIEW DOMINSKI1,3

1Integrative Program for Biological and Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA
2Department of Biological Sciences, Columbia University, New York, New York 10027, USA
3Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

ABSTRACT

Metazoan replication-dependent histone pre-mRNAs are cleaved at the 3′ end by U7 snRNP, an RNA-guided endonuclease that contains U7 snRNA, seven proteins of the Sm ring, FLASH, and four polyadenylation factors: symplekin, CPSF73, CPSF100, and CstF64. A fully recombinant U7 snRNP was recently reconstituted from all 13 components for functional and structural studies and shown to accurately cleave histone pre-mRNAs. Here, we analyzed the activity of recombinant U7 snRNP in more detail. We demonstrate that in addition to cleaving histone pre-mRNAs endonucleolytically, reconstituted U7 snRNP acts as a 5′–3′ exonuclease that degrades the downstream product generated from histone pre-mRNAs as a result of the endonucleolytic cleavage. Surprisingly, recombinant U7 snRNP also acts as an endonuclease on single-stranded DNA substrates. All these activities depend on the ability of U7 snRNA to base-pair with the substrate and on the presence of the amino-terminal domain (NTD) of symplekin in either cis or trans, and are abolished by mutations within the catalytic center of CPSF73, or by binding of the NTD to the SSU72 phosphatase of RNA polymerase II. Altogether, our results demonstrate that recombinant U7 snRNP functionally mimics its endogenous counterpart and provide evidence that CPSF73 is both an endonuclease and a 5′–3′ exonuclease, consistent with the activity of other members of the β-CASP family. Our results also raise the intriguing possibility that CPSF73 may be involved in some aspects of DNA metabolism in vivo.

Keywords: histone pre-mRNA; U7 snRNP; CPSF73; 3′ end processing; symplekin

INTRODUCTION

U7 snRNP is a multicomponent endonuclease responsible for 3′ end processing of metazoan replication-dependent histone pre-mRNAs (Dominski and Marzluff 2007; Dominski et al. 2013). The core of U7 snRNP contains two stably associated components: U7 snRNA of ∼60 nt and a Sm ring of seven proteins assembled around the centrally located Sm site in the U7 snRNA (Schumperli and Pillai 2004). Five of these subunits, SmB, SmD3, SmE, SmF, and SmG, are shared with the spliceosomal snRNPs. Lsm10 and Lsm11 are unique to U7 snRNP, replacing the SmD1 and SmD2 subunits found at the same positions in the spliceosomal snRNPs (Pillai et al. 2001, 2003). Lsm11 through its long amino-terminal region interacts with the amino-terminal region of the 200 kDa protein FLASH (Yang et al. 2009a; Aik et al. 2017) and the two proteins form a platform that recruits the Histone pre-mRNA Cleavage Complex (HCC), a subset of proteins shared with the canonical cleavage and polyadenylation machinery (Yang et al. 2009a, 2013). Analysis of purified Drosophila and mouse U7 snRNPs by mass spectrometry identified CPSF73, CPSF100, symplekin and CstF64 as the major components of the HCC, with the remaining CPSF subunits (CPSF160, WDR33, Fip1, and CPSF30) being present in much smaller amounts (Sabath et al. 2013; Skrajna et al. 2018). U7 snRNP containing the HCC is active in cleaving histone pre-mRNAs. Earlier studies demonstrated that CPSF73 contacts the pre-mRNA substrate, consistent with its role as the likely catalytic component (Dominski et al. 2005), and symplekin was shown to be essential for processing in vitro (Kolev and Steitz 2005). The role of the remaining HCC subunits in U7-dependent
processing and potential requirement for other components that might have escaped detection remained elusive.

U7 snRNP recognizes histone pre-mRNAs via the 5′ region of U7 snRNA, which base-pairs with a sequence in histone pre-mRNAs termed the histone downstream element (HDE) (Schaufele et al. 1986; Bond et al. 1991). The interaction of U7 snRNP with the HDE is facilitated by the stem–loop binding protein (SLBP), which tightly associates with a conserved stem–loop structure located upstream of the HDE (Tan et al. 2013; Zhang et al. 2014) and stably anchors U7 snRNP on the substrate by contacting FLASH and/or Lsm11 (Skrajna et al. 2017). Processing of histone pre-mRNAs by U7 snRNP involves a single endonucleolytic cleavage located between the stem–loop and the HDE, yielding mature histone mRNAs without a poly(A) tail, ending instead with the stem–loop followed by a single-stranded tail of 4–5 nt.

To investigate both structural and functional aspects of 3′ end processing of histone pre-mRNAs, we recently generated semirecombinant U7 snRNP, in which U7 snRNA, FLASH and the seven subunits of the Sm ring, including Lsm10 and Lsm11, were recombinant, whereas the HCC was endogenous (Bucholc et al. 2020). This semirecombinant U7 snRNP accurately cleaved histone pre-mRNA, functionally resembling endogenous U7 snRNP. More recently, we generated fully recombinant U7 snRNP by coexpressing CPSF73, CPSF100, symplekin, and CstF64 in the baculovirus system (Sun et al. 2020). Importantly, this U7 snRNP also accurately cleaved histone pre-mRNAs, demonstrating that other components, including the remaining polyadenylation factors identified in the endogenous HCC, are not required in vitro. The cleavage activity of the recombinant U7 snRNP was dependent on the presence of an intact catalytic site in CPSF73, confirming its role as the endonuclease, and on the amino-terminal HEAT domain (NTD) of symplekin. Strikingly, the NTD supported processing not only in cis, as part of the full-length protein and a stable component of the HCC, but also in trans, as a separate polypeptide that only transiently associates with recombinant U7 snRNP (Sun et al. 2020).

Cryoelectron microscopy studies of the recombinant U7 snRNP bound to SLBP and histone pre-mRNA revealed an unanticipated network of interactions within the reconstituted processing complex and demonstrated that the NTD of symplekin functions by recognizing the duplex of U7 snRNA and the HDE, triggering a conformational shift within the complex that results in opening of the catalytic site of CPSF73 for the cleavage reaction (Sun et al. 2020). The SSU72 phosphatase of RNA polymerase II binds to the same region of the NTD that recognizes the RNA duplex (Xiang et al. 2010), acting as a potential negative regulator of the U7-dependent processing.

Here, we took advantage of the fully recombinant U7 snRNP to identify some of the structural requirements essential for its function in vitro and to study its nucleolytic activities in more detail.

RESULTS

Recombinant U7 snRNP accurately cleaves histone pre-mRNAs

Fully recombinant U7 snRNP was reconstituted and analyzed functionally using a two-step approach, as previously described (Sun et al. 2020). In the first step, the three Sm subcomplexes of human U7 snRNP (Lsm10/Lsm11, SmE/SmF/SmG, and SmB/SmD3) were assembled on U7 snRNA and the resultant recombinant core U7 snRNP was purified by size exclusion chromatography (SEC) (Bucholc et al. 2020). U7 snRNA was synthesized by T7 transcription and had the same sequence as mouse U7 snRNA (WT), which only slightly differs from the sequence of human U7 snRNA (Bucholc et al. 2020). In addition to using wild-type mouse U7 snRNA (WT), we also generated various U7 snRNA mutants by altering the 5′ region that base-pairs with the HDE, the Sm binding site and the 3′ terminal stem–loop. In the second step, the purified core U7 snRNP was mixed with amino-terminal FLASH (amino acids 51–137) and the HCC (symplekin, CPSF73, CPSF100, and CstF64) to form recombinant U7 snRNP, which was tested in the presence of SLBP for the ability to cleave 5′-labeled histone pre-mRNA (Fig. 1A). Three HCC variants differing in the length of symplekin were used to assemble recombinant U7 snRNP: symplekin FL (amino acids 30–1101) encompassing most of the protein, including the essential NTD, symplekin 353 (amino acids 353–1110) that lacks the essential NTD, and symplekin 538 that lacks a larger amino-terminal fragment, including the region interacting with CstF64 (Fig. 1B). When indicated, the NTD (amino acids 30–360) was added to the processing reaction in trans as a separate polypeptide (Xiang et al. 2010). CPSF73 was either wild type or contained two point mutations, D75N and H76A (NA mutation), designed to disrupt its catalytic center. A complete processing reaction containing recombinant U7 snRNP, NTD of symplekin and SLBP is referred to as “All.”

As previously demonstrated (Sun et al. 2020), a processing reaction containing WT core U7 snRNP (assembled on wild-type U7 snRNA), FLASH, the HCC (symplekin 353 (Fig. 1C, lane 1) and the NTD (that together form recombinant U7 snRNP) cleaved in the presence of SLBP a modified mouse histone pre-mRNA (mH2a), yielding the upstream product of the expected 26 nt (Fig. 1D, lane 1). Cleavage was blocked by adding an antisense oligonucleotide complementary to the 5′ end of U7 snRNA (Fig. 1D, lane 2), by omitting the NTD (Fig. 1D, lane 3), or by incorporating the D75N/H76A mutation into CPSF73 (Fig. 1D, lane 4) that disrupts its catalytic site but does not affect the assembly of the HCC (HCC353,NA, Fig. 1C, lane 2). Recombinant U7 snRNP
CPSF73 is an endonuclease and a 5′-3′ exonuclease

containing the HCC\textsuperscript{353,'} (Fig. 1E, lane 2) cleaved mH2a\textsuperscript{*} pre-mRNA in the absence of the NTD polypeptide in trans (Fig. 1F, lane 5, and Fig. 1G, lane 3) but remained dependent on the presence of an intact catalytic site in CPSF73 (Fig. 1F, lanes 6,7). As expected, no cleavage occurred in the absence of the core U7 snRNP or the HCC (Fig. 1G, lanes 4,5). Altogether, these results are consistent with our recent report (Sun et al. 2020) and demonstrate that the reconstituted recombinant U7 snRNP accurately cleaves histone pre-mRNA, with the NTD of symplekin, the catalytic site of CPSF73 and the core U7 snRNP (U7 snRNA and the Sm ring) being indispensable components of the reaction.

Cleavage depends on the sequence of U7 snRNA

Wild-type (WT) mouse U7 snRNA forms an uninterrupted duplex of 16 bp with mH2a\textsuperscript{*} pre-mRNA and processing of this pre-mRNA in mouse nuclear extracts is independent of SLBP (Bucholc et al. 2020). We used recombinant U7 snRNP containing amino-terminally truncated symplekin and the NTD added in trans as a separate polypeptide to test the dependence of cleavage on the sequence of U7 snRNA. By varying the 5′-end region of mouse U7 snRNA, we created two variant snRNAs, Sup U7 and Hyb U7, as previously described (Bucholc et al. 2020). In Sup U7 snRNA, the CUCUUU sequence that interacts with the HDE was changed to GAGAAA. This U7 snRNA has a reduced ability to base-pair with mH2a\textsuperscript{*} pre-mRNA, forming a duplex of only 10 bp. In Hyb U7 snRNA, the entire 5′-end region of WT mouse U7 snRNA was replaced with a sequence complementary to Drosophila H3 histone pre-mRNA. As a result, Hyb U7 snRNA fails to base-pair with mH2a\textsuperscript{*} pre-mRNA but forms 21 bp with the Drosophila-specific dH3\textsuperscript{*} pre-mRNA (Fig. 2A).

In the absence of SLBP, only recombinant U7 snRNP containing WT U7 snRNA supported efficient processing of mH2a\textsuperscript{*} pre-mRNA, with Sup U7 snRNP being only weakly active in cleaving this substrate and Hyb U7 snRNP yielding no cleavage product (Fig. 2B, lanes 2–4, respectively). The efficiency of processing by Sup U7 snRNP was dramatically enhanced by SLBP (Fig. 2B, lanes 2,4). In the absence of SLBP, both WT and Sup core U7 snRNPs, in spite of the presence of the catalytic site of CPSF73 and the core U7 snRNP (U7 snRNA and the Sm ring) being indispensable components of the reaction.

FIGURE 1. Cleavage of histone pre-mRNAs by recombinant U7 snRNP depends on the catalytic site in CPSF73 and the amino-terminal domain of symplekin. (A) Diagram depicting the assembly of U7 snRNP from recombinant components: core U7 snRNP (U7 snRNA and seven subunits of the Sm ring), FLASH, and the HCC constituting recombinant U7 snRNP accurately cleaves histone pre-mRNA, with the NTD of symplekin, the catalytic site of CPSF73 and the core U7 snRNP (U7 snRNA and the Sm ring) being indispensable components of the reaction.
between the HDE of mH2a RNA (2020) Vol. 26, No. 10

The cleavage of mH2a

FIGURE 2. Cleavage of mH2a histone pre-mRNA by recombinant U7 snRNP depends on the sequence of U7 snRNA. (A) Base-pairing between the HDE of mH2a histone pre-mRNA (top sequence) and the 5′ end region of three different U7 snRNAs (bottom sequences). The mutated region of Sup U7 snRNA is indicated with lower case letters. Vertical lines and dots indicate Watson–Crick and GU base pairs, respectively. (B–F) Recombinant U7 snRNP containing indicated components was used to cleave the following histone pre-mRNAs: mH2a (panel B), Drosophila-specific dH3 (panel C), H2a WT (panel D), and H2a WT/Ys (panels E,F). The arrow in panel D indicates a shift in the length of the upstream cleavage product resulting from substituting 4 nt into mouse U7 snRNA. The arrow in panels E and F indicates a new upstream cleavage product generated as a result of mutating the U7-specific Sm site in Sup U7 snRNA with that present in the spliceosomal-type Sm site displayed reduced fidelity for fully recombinant U7 snRNP. To this end, we assembled fully recombinant U7 snRNP on Sup U75pl snRNA in which the U7-specific Sm site was modified to the spliceosomal type (AAUUUGUCUAG to AAUUUUGGAG, changed nucleotides are underlined). The substitution was made in the context of Sup U7 snRNA that contains the CUCUUU to GAGAAA mutation within the 5′ end region, and processing was tested using H2a WT/Ys pre-mRNA containing a compensatory mutation within the HDE (Ys mutation) (Bucholc et al. 2020). While fully recombinant Sup U7 snRNP containing the CUCUUU to GAGAAA mutation yielded only the correct processing product of 47 nt (Fig. 2E, lane 4), the Sup U75pl snRNP additionally containing the spliceosomal-type Sm site displayed reduced fidelity of cleavage, generating both the correct product and a longer product of ~50 nt (Fig. 2E, lane 3). As expected, recombinant U7 snRNP assembled on the wild-type mouse U7 snRNA (WT U7) failed to cleave H2a WT/Ys pre-mRNA due to the inability of these two RNAs to form a stable duplex (Fig. 2E, lane 5). Omitting the NTD of symplekin, HCC353 or core U7 snRNP resulted in no processing (Fig. 2F, lanes 4–6), while the absence of SLBP reduced processing activity without changing the cleavage pattern (Fig. 2F, lane 7). As a control, we tested processing of H2a WT/Ys pre-mRNA by semirecombinant Sup U75pl snRNP in which the HCC was endogenous (from the nuclear extract) and hence contained full-length symplekin. Again, the same two closely spaced products were observed, although the overall efficiency of processing was much lower (Fig. 2F, lanes 1,8), likely due to limiting the amount of the semirecombinant Sup U75pl snRNP used in the reaction (Bucholc et al. 2020). We conclude that fully recombinant U7 snRNP in which the HCC is expressed in the baculovirus system and the amino-terminal domain of symplekin is provided in trans functionally mimics both endogenous and semirecombinant U7 snRNPs, with the substrate and cleavage specificities being determined by the sequences of the 5′ end region and the Sm site of U7 snRNA.

formation of a strong U7 snRNA/HDE duplex or the presence of SLBP, if the formed duplex is too short or interrupted by mismatches.

We tested whether recombinant U7 snRNP determines the site of cleavage and is capable of length suppression, a phenomenon of cleavage site repositioning by insertions within U7 snRNA (Scharl and Steitz 1994, 1996). Recombinant U7 snRNP assembled on WT U7 snRNA cleaved mH2a, a derivative of mouse H2a pre-mRNA containing a 4-nt insertion between the stem–loop and HDE (Dominski et al. 1999), to generate a 51-nt product (Fig. 2D, lane 2). This product is 4 nt longer than the product generated from wild-type H2a pre-mRNA containing no insertion (Dominski et al. 1999), consistent with the observation that histone pre-mRNAs are cleaved at a fixed distance from the HDE, the site of U7 snRNP binding (Scharl and Steitz 1994). In the presence of recombinant U7 snRNP assembled on U7 snRNA containing a 4-nt insertion between the 5′ region and the Sm site, cleavage at the original site was restored generating a product of 47 nt (Fig. 2D, lane 4). Thus, inserting a sequence into the U7 snRNA moved the position of the cleavage site closer to the stem–loop, mirroring the efficient length suppression previously observed in Xenopus oocytes (Scharl and Steitz 1996) and in vitro using semirecombinant U7 snRNPs (Bucholc et al. 2020).

In vitro, U7 snRNA containing the spliceosomal Sm site incorporates Lsm10 and Lsm11 into the Sm ring, but endogenous HCC bound to this ring yields semirecombinant U7 snRNP that is functionally impaired, cleaving histone pre-mRNA both at the correct site and at a new site, located 4–5 nt further downstream (Bucholc et al. 2020). We tested whether this unusual behavior will also be observed for fully recombinant U7 snRNP. To this end, we assembled fully recombinant U7 snRNP on Sup U75pl snRNA in which the U7-specific Sm site was modified to the spliceosomal type (AAUUUGUCUAG to AAUUUUGGAG, changed nucleotides are underlined). The substitution was made in the context of Sup U7 snRNA that contains the CUCUUU to GAGAAA mutation within the 5′ end region, and processing was tested using H2a WT/Ys pre-mRNA containing a compensatory mutation within the HDE (Ys mutation) (Bucholc et al. 2020). While fully recombinant Sup U7 snRNP containing the CUCUUU to GAGAAA mutation yielded only the correct processing product of 47 nt (Fig. 2E, lane 4), the Sup U75pl snRNP additionally containing the spliceosomal-type Sm site displayed reduced fidelity of cleavage, generating both the correct product and a longer product of ~50 nt (Fig. 2E, lane 3). As expected, recombinant U7 snRNP assembled on the wild-type mouse U7 snRNA (WT U7) failed to cleave H2a WT/Ys pre-mRNA due to the inability of these two RNAs to form a stable duplex (Fig. 2E, lane 5). Omitting the NTD of symplekin, HCC353 or core U7 snRNP resulted in no processing (Fig. 2F, lanes 4–6), while the absence of SLBP reduced processing activity without changing the cleavage pattern (Fig. 2F, lane 7). As a control, we tested processing of H2a WT/Ys pre-mRNA by semirecombinant Sup U75pl snRNP in which the HCC was endogenous (from the nuclear extract) and hence contained full-length symplekin. Again, the same two closely spaced products were observed, although the overall efficiency of processing was much lower (Fig. 2F, lanes 1,8), likely due to limiting the amount of the semirecombinant Sup U75pl snRNP used in the reaction (Bucholc et al. 2020). We conclude that fully recombinant U7 snRNP in which the HCC is expressed in the baculovirus system and the amino-terminal domain of symplekin is provided in trans functionally mimics both endogenous and semirecombinant U7 snRNPs, with the substrate and cleavage specificities being determined by the sequences of the 5′ end region and the Sm site of U7 snRNA.
Structural requirements for the cleavage activity by recombinant U7 snRNP

Of the four major subunits identified in the endogenous HCC, the function of CstF64 was the most elusive. Cryo-EM studies of the recombinant U7 snRNP failed to identify the location of this subunit and its interacting partners within the complex. In addition, earlier in vivo studies yielded conflicting results. While RNAi-mediated depletion of CstF64 in Drosophila cells suggested that it has no role in 3′ end processing of histone pre-mRNAs (Wagner et al. 2007; Sabath et al. 2013), similar studies in mammalian cells resulted in the accumulation of polyadenylated histone mRNAs, an indication of a defect in the U7-dependent processing (Ruepp et al. 2011; Romeo et al. 2014).

We tested the importance of CstF64 in 3′ end processing of histone pre-mRNAs using HCC353–64, an HCC variant similar to HCC353 (full-length CPSF100, full-length CPSF73 and symplekin spanning amino acids 353–1110) but lacking CstF64. The NTD was added to the processing reaction in trans, as a separate polypeptide. In the purified HCC353–64, CPSF100 was readily detectable but symplekin and CPSF73 were substoichiometric (Fig. 3A), suggesting that in the absence of CstF64 these two subunits might be degraded during purification. Importantly, recombinant U7 snRNP reconstituted from the NTD, WT core U7 snRNP, FLASH, and the HCC353–64 purified on nickel beads was active in processing, yielding in the presence of SLBP nearly 40% of the cleavage product (Fig. 3B, lane 3). The activity was blocked by the removal of the NTD or the addition of the u7 oligonucleotide (Fig. 3B, lanes 2, 4), mirroring the behavior of recombinant U7 snRNP containing the HCC with all four subunits (Fig. 1D, lanes 1–3).

CstF64 binds a region in symplekin located between amino acids 390 and 465 (Ruepp et al. 2011). We also expressed HCC353–64, an HCC version containing CPSF100, CPSF73 and amino-terminally truncated symplekin (amino acid 538–1110) that lacked the CstF64 binding site. Coexpression of these three proteins in the absence of CstF64 yielded a high quality and purity complex (Fig. 3C, lane 2), in agreement with our previous results (Sun et al. 2020). We conclude that CstF64 is dispensable for 3′ end processing of histone pre-mRNAs in vitro.

Endogenous U7 snRNP purified from Drosophila and mouse cells contains detectable amounts of CPSF160, WDR33, Fip1, and CPSF30 (Sabath et al. 2013; Skrajna et al. 2018). These subunits are part of the mammalian polyadenylation specificity factor (mPSF), a subcomplex of CPSF that is essential for the recognition of the AAUAAA poly(A) signal during 3′ end processing of canonical pre-mRNAs (Chan et al. 2014; Schonemann et al. 2014; Schonemann et al. 2015).

FIGURE 3. Structural requirements for the cleavage activity. (A,B) The HCC lacking CstF64 (HCC353–64) was analyzed by SDS/polyacrylamide gel electrophoresis followed by Coomassie blue staining (panel A) and tested for the ability to support processing in a reaction additionally containing NTD, WT core U7 snRNP, FLASH, and SLBP (panel B). The input mH2a∗ pre-mRNA and processing in the absence of the NTD are shown in lanes 1, 2 of panel B, respectively. (C,D) As in panels A and B, with the exception that symplekin in HCC353–64 begins with amino acid 538. (E) Processing of mH2a∗ pre-mRNA in the presence of SLBP and recombinant U7 snRNP containing the NTD, WT core U7 snRNP, FLASH, and HCC353. mPSF was added at two-, four-, and eightfold molar excess relative to the HCC353. (F, G) As in panels C and D, with the exception that CPSF100 in HCC353, APIM–64 lacks amino acids 460–486 interacting with CPSF160 and WDR33. (H, I) Residues of the NTD that make contact with CPSF100 (panel H) were replaced with alanines in two overlapping blocks and the resultant mutant NTDs were tested for the ability to support processing of mH2a∗ pre-mRNA together with WT core U7 snRNP, FLASH, and HCC533. In lanes 4–6, mPSF was added at two-, four-, and eightfold molar excess relative to the HCC533. (J) As in panel I, with the exception that the mutated residues, Leu45 and Met116, were located at the interface of the NTD interacting with SmE, and each mutant NTD was added to the processing reaction to reach 5× and 1× molar ratio relative to the HCC533.
Casanal et al. 2017; Sun et al. 2017; Clerici et al. 2018). While fully recombinant U7 snRNP lacking the mPSF subunits is active in processing, indicating that they are not essential, we considered the possibility that they may play an auxiliary role by stimulating processing, or conversely, acting as an inhibitor by down-regulating the cleavage reaction. We tested these alternative possibilities using recombinant mPSF consisting of full-length CPSF160 and CPSF30 and a truncated version of WDR33 encompassing amino acids 1-572. These three proteins form a stable complex that interacts with the HCC (Zhang et al. 2019). We added increasing amounts of the recombinant mPSF to a processing reaction containing SLBP and recombinant U7 snRNP consisting of HCC\textsuperscript{353}, NTD, WT core U7 snRNP, and FLASH, and observed no effect on cleavage of mH2a\textsuperscript{*} pre-mRNA (Fig. 3E, lanes 3–6). Thus, subunits of the mPSF have neither positive nor negative role in cleaving histone pre-mRNAs in vitro.

The HCC binds the mPSF via a short loop in CPSF100 (amino acids 460–486) designated the PSF interaction motif (PIM) (Zhang et al. 2019). PIM contacts two different mPSF subunits: WDR33 and CPSF160. We additionally investigated potential roles of the mPSF module in 3' end processing of histone pre-mRNAs by expressing an HCC with the CPSF100 subunit that lacks the PIM and is unable to interact with mPSF (Zhang et al. 2019). The deletion was made in the context of the HCC\textsuperscript{353, 486}, generating HCC\textsuperscript{353, 486}APIM, 64. Recombinant U7 snRNP containing WT core U7 snRNP, the NTD, FLASH and this mutant HCC supported efficient processing of mH2a\textsuperscript{*} pre-mRNA (Fig. 3G, lane 3), further arguing that the entire mPSF module and its interacting motif in CPSF100 are specifically designated to function in 3' end processing by cleavage coupled to polyadenylation and are dispensable for the U7-dependent processing.

Cryo-EM reconstruction of active recombinant U7 snRNP in complex with histone pre-mRNA and SLBP revealed that a number of positively charged residues in the NTD of symplekin contact the duplex formed between the HDE and U7 snRNA. Mutations of these residues abolished the ability of the NTD to support processing, emphasizing the importance of this interaction in triggering the catalytic activity of CPSF73 (Sun et al. 2020). In addition to contacting the duplex RNA, the NTD also interacts with CPSF100, with most contacts being mediated by the KVQVELIINK motif (amino acids 56–65, interacting residues are underlined) of the first HEAT repeat of the NTD. Within the sequence, the first lysine (position 56) also interacts with the duplex RNA. We substituted the interacting residues with alanines in two overlapping blocks, creating two mutant NTDs, KAVOAlIAAA and AAVQAlIAAK (mutated residues are underlined), that were added to a processing reaction containing SLBP, HCC\textsuperscript{353}, WT core U7 snRNP, and FLASH. Note that KAVOAlIAAA contains substitutions of only those residues that specifically interact with CPSF100, with Lys56 remaining unaltered (Fig. 3H). In contrast to the wild-type NTD of symplekin, neither of these two mutant NTDs supported processing of mH2a\textsuperscript{*} (Fig. 3I, lanes 3–5), indicating that the interaction between the NTD of symplekin and CPSF100 is essential for the activity of U7 snRNP.

As revealed by the cryo-EM studies (Sun et al. 2020), the NTD makes a number of contacts with the Sm ring that likely help to stabilize the interaction of the NTD with the U7 snRNA/HDE duplex. These contacts include Leu45 and Met116 of the NTD that interact with the SmE subunit of the ring. We generated two mutations within the NTD by substituting Leu45, either alone and together with Met116, with alanine and tested whether the mutant proteins support processing at two different final concentrations, 1.5 µM (fivefold molar concentration relative to the concentration of the HCC) and 0.3 µM (stoichiometric with the HCC). While the L-A single mutant, when compared to the wild-type NTD, displayed a moderately reduced activity at lower concentration (Fig. 3J, compare lanes 4,6), the LM-AA double mutant was less active at both concentrations (Fig. 3J, lanes 7,8), suggesting that it is significantly impaired in the interaction with the SmE subunit of the U7 Sm ring and that this interaction plays an important role in forming a catalytically competent processing complex.

The release of the 5' terminal nucleotide from the DCP by recombinant U7 snRNP

Two fragments are generated as a result of endonucleolytic cleavage of histone pre-mRNAs by U7 snRNP. The upstream fragment ends with the stem–loop followed by a 5'-nt single-stranded tail and it corresponds to mature histone mRNA. The downstream cleavage product (DCP) contains a phosphate group at the 5' end. In nuclear extracts, this product is degraded in the 5'-3' direction by the bound U7 snRNP, liberating U7 snRNP from the HDE for another round of processing (Walther et al. 1998). Our previous UV-cross linking studies suggested that the degradation is catalyzed by a 5'-3' exonuclease activity of CPSF73 (Dominski et al. 2005; Yang et al. 2009b).

We carried out a set of experiments to determine whether the fully recombinant U7 snRNP can act as a 5'-3' exonuclease. As a diagnostic test for this activity, we initially monitored the release of the \( ^{32} \text{P}-\)labeled 5’ terminal nucleotide (Dominski et al. 2005; Yang et al. 2009b) from the DCP of mH2a\textsuperscript{*} pre-mRNA, a substrate that forms 16 continuous base pairs with mouse U7 snRNA (Fig. 4A). Recombinant U7 snRNP reconstituted from the HCC\textsuperscript{353}, NTD, WT core U7 snRNP, and FLASH released 5’ terminal CMP, mimicking the activity of endogenous U7 snRNP in the mouse nuclear extract (Fig. 4B, lanes 3,1, respectively). The presence of an antisense oligonucleotide complementary to U7 snRNA (Fig. 4B, lane 4) or the absence of the NTD, WT core U7 snRNP or HCC\textsuperscript{353} abolished the
activity (Fig. 4B, lanes 5–7), emphasizing the specificity of the reaction. In the absence of FLASH, recombinant U7 snRNP remained active in removing the 5′ terminal nucleotide (Fig. 4B, lanes 8), consistent with the observation that the dependence of the U7 snRNP activity on FLASH is alleviated by using high concentrations of recombinant HCC (Sun et al. 2020). In addition, we previously showed that in contrast to the endonuclease activity, the 5′–3′ exonuclease activity of endogenous and semirecombinant U7 snRNP is independent of FLASH (Burch et al. 2011; Yang et al. 2011; Bucholc et al. 2020). The DCP substrate lacks the stem–loop and, as expected, its degradation was also independent of SLBP (Fig. 4B, lanes 9). Only WT core U7 snRNP, whose U7 snRNA is capable of efficiently base-pairing with DCP RNA, supported the activity (Fig. 4B, lane 11). Recombinant U7 snRNPs containing Sup or Hyb U7 snRNAs were virtually inactive (Fig. 4B, lanes 12,13), demonstrating the importance of base-pairing in substrate recognition.

No activity was observed when recombinant U7 snRNP was reconstituted using HCC 353,NA, containing the D75N/H76A mutant of CPSF73 (Fig. 4C, lane 5). As expected, recombinant U7 snRNP containing full-length symplekin and wild-type CPSF73 (HCCFL) removed the 5′ terminal nt from DCP RNA even in the absence of the NTD, and excluding WT core U7 snRNP from the reaction abolished the activity (Fig. 4C, lanes 9,10, respectively). We previously showed that binding of SSU72 to the NTD of symplekin inhibits cleavage activity of recombinant U7 snRNP by interfering with the ability of the NTD to recognize the duplex of histone pre-mRNA and U7 snRNA (Sun et al. 2020). SSU72 also exerted a strong inhibitory effect on the release of the 5′ nt when the NTD was provided in trans. The reaction was severely reduced at 0.3 µM SSU72, a concentration equimolar with the NTD, and was virtually abolished at higher concentrations (Fig. 4D, lanes 4–7). Much weaker inhibition was observed when the NTD contained the K185A mutation that impairs its interaction with SSU72 but has no effect on processing (Fig. 4D, lanes 8–12) (Xiang et al. 2010; Sun et al. 2020). A significant decline in the efficiency of CMP release was only observed at 7.5 µM SSU72, a 25-fold molar excess of SSU72 over K185A NTD (Fig. 4D, lane 12). SSU72 also inhibited the removal of the 5′ terminal nucleotide in mouse nuclear extracts, although higher concentrations of SSU72 were required to achieve a substantial inhibition (Fig. 4D, lanes 2–5), consistent with the NTD being an integral part of endogenous symplekin and forming a strong competing complex with the duplex RNA during DCP degradation.

Recombinant U7 snRNP has a 5′-3′ exonuclease activity

The release of the 5′ terminal nucleotide from the DCP RNA, while providing a convenient assay for monitoring the degradation activity of U7 snRNP, does not prove that the responsible activity is a 5′–3′ exonuclease. The same effect can be achieved by an endonuclease activity that cleaves the substrate to completion, releasing individual mononucleotides, including 5′ labeled CMP.
distinguish between these two possibilities, we previously incorporated 2′O-methyl or 2′-deoxy nucleotides within DCP substrates and a site-specific internal label located downstream from the last modification (Yang et al. 2009b). Incubation of these substrates with a nuclear extract containing endogenous U7 snRNP resulted in the accumulation of specific intermediates whose length could be explained only by the degradation beginning from the 5′ end and stalling at the modified nucleotide(s), before reaching the internal label closer to the 3′ end, as expected for the 5′–3′ exonuclease activity.

To determine whether fully recombinant U7 snRNP displays the same 5′–3′ exonuclease activity, we used two substrates: DCP/1xM and DCP/3xD (Fig. 5A). In DCP/1xM, the nucleotide at position 6 was modified with a 2′O-methyl group, whereas in DCP/3xD three ribonucleotides at positions 13, 14, and 15 were replaced with deoxynucleotides. The two substrates were labeled by adding pCp to the 3′ end (Fig. 5A). Incubation of DCP/1xM and DCP/3xD in a mouse nuclear extract generated product patterns expected for endogenous U7 snRNP acting as a 5′–3′ exonuclease, yielding a single product and a group of shorter, closely migrating products, respectively (Fig. 5B, lanes 2,5), consistent with a previous analysis of similar substrates (Yang et al. 2009b). In both cases, the activity was blocked by αU7, an oligonucleotide complementary to endogenous U7 snRNA (Fig. 5B, lanes 3,6). Note that incubation of each substrate in the nuclear extract also generated a nonspecific product (labeled as X) that persisted in the presence of this oligonucleotide, emphasizing its selective inhibitory effect.

We tested degradation of each substrate by recombinant U7 snRNP containing WT core U7 snRNP, the NTD of symplekin, FLASH and one of three different HCCs: HCC353, HCC353,NA, and HCCFL. Since the DCP substrates lack the stem-loop, which is part of the upstream product, SLBP was omitted. Incubation of DCP/1xM with U7 snRNP containing HCC353 converted DCP/1xM RNA to the same length intermediate that was generated in the nuclear extract (Fig. 5C, lanes 4,5). As determined by a higher resolution gel electrophoresis, this product encompasses the last 26 nt of the substrate (not shown), consistent with the degradation of the DCP/1xM beginning at the 5′ end and stalling at the modified nucleotide before reaching the labeled 3′ end (Fig. 5A). Note that the DCP/1xM contains a hydroxyl group at the 5′ end, indicating that the activity of recombinant U7 snRNP generating the 26-nt product does not require a 5′ phosphate, contrasting with the activity of the Xm2 exonuclease and consistent with the 5′–3′ exonuclease activity of endogenous U7 snRNP (Yang et al. 2009b). No degradation of DCP/1xM was observed if the reaction lacked the HCC or if the HCC contained mutant CPSF73 (Fig. 5C, lanes 2,3), underscoring the specific nature of the activity that generates the 26-nt intermediate.

We next analyzed intermediates generated by recombinant U7 snRNP from DCP/3xD RNA, using a high resolution gel electrophoresis to separate closely migrating bands. No degradation was observed in the absence of the HCC or when the HCC contained the CPSF73 mutation.
D75N/H76A (HCC^{353,NA}) (Fig. 5D, lanes 2, 5). Importantly, incubation of DCP/3xD with recombinant U7 snRNP containing HCC^{353} or HCC^{FL} yielded a characteristic pattern of three closely spaced bands, consistent with the degradation beginning at the 5’ end and partially stalling at each of the deoxynucleotides (Fig. 5D, lanes 3, 4). As in the case of the endonuclease activity, generation of these intermediates by the recombinant U7 snRNP required the presence of core U7 snRNP and the NTD either in trans or cis (Fig. 5E, lanes 2–7). Sequencing-type gel electrophoresis determined the size of the three intermediates as 17, 18, and 19 nt long (Fig. 5F, lane 4), as expected (Fig. 5A).

Collectively, these experiments demonstrate that recombinant U7 snRNP degrades the DCP substrate using a 5’–3’ exonuclease activity of CPSF73, confirming our previous conclusions from studying endogenous (Yang et al. 2009b) and semirecombinant U7 snRNP (Bucholc et al. 2020). The activity requires the NTD of symplekin in cis or trans and an intact catalytic site of CPSF73. Deoxynucleotides only partially stall the progression of the enzyme toward the 3’ end. Note that the three deoxynucleotides were placed within the purine core of HDE that is engaged in formation of a duplex with U7 snRNA. Thus, the 5’–3’ exonuclease activity of CPSF73 can penetrate into the duplex structure, as suggested by previous studies on endogenous U7 snRNP (Yang et al. 2009b). We did not observe the release of the 3’ labeled nucleotide, suggesting that the three deoxynucleotides, and possibly the remaining part of the duplex, are sufficient to prevent further progression of the enzyme. As previously determined by studying endogenous U7 snRNP, the 2’O-methyl group is less tolerated by the 5’–3’ exonuclease activity of CPSF73, with a single 2’O-methyl-modified nucleotide being sufficient to prevent any further degradation of the RNA substrate (Yang et al. 2009b).

**Activity of recombinant U7 snRNP on DNA substrates**

The fact that the 5’–3’ exonuclease activity of recombinant U7 snRNP is capable of partially progressing through deoxynucleotides suggested that it might display an endonuclease activity on single-stranded DNA substrates.

We tested this possibility using DNA/50, a 50-nt single-stranded DNA substrate (Fig. 6A) with the sequence corresponding to the previously used ΔSL/H2a RNA and U bases being replaced with T bases. The ΔSL/H2a RNA lacks the stem–loop and instead contains an unstructured region of 20 nt upstream of the cleavage site, and the HDE is capable of forming 16 bp with U7 snRNA (Bucholc et al. 2020). Note that SLBP only binds RNA (Tan et al. 2013) and the stem–loop sequence, even if present in single-stranded DNA substrates, would not play any role in processing. Recombinant U7 snRNP was assembled using WT core U7 snRNP, FLASH, and one of two different HCCs: HCC^{353} or HCC^{FL}. Strikingly, while U7 snRNP containing HCC^{353} showed no activity with DNA/50 even in the presence of the NTD (Fig. 6B, lanes 2, 3), U7 snRNP containing HCC^{FL} (symplekin encompassing amino acids 30–1101) generated a single product with the expected length of ~20 nt (Fig. 6B, lane 6). When the HCC^{FL} or WT core U7 snRNP

**FIGURE 6.** Recombinant U7 snRNP cleaves single-stranded DNA substrates. (A) Nucleotide sequence of three single-stranded DNA substrates, DNA/50, DNA/50Ys, and DNA/55Ys, used in the assay. The arrows indicate expected length of products if the endonucleolytic cleavage occurs at the same site as in the RNA substrates. (B–F) Endonucleolytic cleavage of DNA/50 (panel B), DNA/50Ys (panels C–E), or DNA/55Ys (panel F) by recombinant U7 snRNP containing indicated components. The arrow in panel B indicates specific cleavage product of the expected size. (G) Cleavage of ΔSL/Ys pre-mRNA and DNA/50Ys substrates by recombinant U7 snRNPs containing either Sup U7 or Sup U7/D snRNAs. The input for each substrate is shown in lane 1 (panels B–F,G) and in lanes 2,3 (panel F). Note that the cleavage products generated from each substrate, in spite of having the same length and sequence, display slightly different electrophoretic mobility due to different chemical nature of DNA and RNA. In addition, ΔSL/Ys pre-mRNA was generated by T7 transcription, rather than by chemical synthesis, as in the case of DNA/50Ys, and it may contain an additional nucleotide at the 5’ end that would increase the length of the final product to 21 nt.
were omitted, no product was detected indicating that cleavage was specific (Fig. 6B, lanes 4,5).

To further confirm the specificity of DNA cleavage by recombinant U7 snRNP, we used DNA/50Ys, a substrate in which six purines in the region corresponding to the HDE in histone pre-mRNAs were replaced with complementary pyrimidines (Fig. 6A). Recombinant U7 containing WT core U7 snRNP unable to efficiently bind this substrate yielded no product (Fig. 6C, lane 2). Importantly, Sup core U7 snRNP containing a compensatory mutation in its RNA component rescued processing (Fig. 6C, lane 3). The generated product comigrated with a size marker encompassing the first 20 nt of DNA/50Ys, demonstrating that cleavage of DNA occurs precisely after the ACCCA sequence, the same site that is used in RNA substrates (Bucholc et al. 2020). Again, cleavage of DNA/50Ys required the presence of Sup core U7 snRNP and HCCFL (Fig. 6D, lanes 2,3). As expected given that symplekin in HCCFL contains covalently attached NTD, the exclusion of the NTD as a separate protein did not have any effect (Fig. 6D, lane 4). Cleavage was inhibited by an antisense oligonucleotide complementary to the 5′ end of U7 snRNA (Fig. 6D, lane 8) or by wild-type SSU72, but not by its triple mutant unable to bind the HEAT domain (Fig. 6D, lanes 6,7, respectively) (Xiang et al. 2010; Sun et al. 2020). Finally, no cleavage was detected in the absence of FLASH or when CPSF73 was mutated by two amino acid substitutions to disrupt its catalytic site (Fig. 6E, lanes 2–5).

To assure that cleavage of DNA substrates occurs at a fixed distance from the HDE rather than from the 5′ end, we extended DNA/50Ys by adding 5 nt to its 5′ end (Fig. 6A). Cleavage of the resultant DNA/55Ys substrate by recombinant U7 snRNP consisting of HCCFL, Sup core U7 snRNP, and FLASH yielded a labeled product longer than 20 nt (Fig. 6F, lane 1), indicating that the cleavage of DNA/50 Ys and DNA/55Ys occurred at the same site, that is, at a fixed distance from the site of U7 snRNP binding. We conclude that recombinant U7 snRNP can act on single-stranded DNA substrates as an RNA-guided endonuclease, with the reaction requiring the presence of the catalytic site in CPSF73, FLASH, and the NTD of symplekin covalently attached to the rest of the protein. Unlike cleavage of RNA substrates, using the NTD as a separate polypeptide did not support cleavage of DNA substrates.

Cleavage of DNA substrates by recombinant U7 snRNP indicates that the NTD of symplekin readily recognizes the heteroduplex formed between the 5′ end of U7 snRNA and the DNA substrate. To test whether the NTD of symplekin also recognizes double-stranded DNA that typically adopts B- rather than A-form helical conformation we used Sup U7/D snRNP. This RNA/DNA chimera has the same sequence as Sup U7 snRNA (Bucholc et al. 2020), with the exception that its 5′ end region of 16 nt that base-pairs with the HDE contains deoxynucleotides and T rather than U bases. In this chimeric DNA/RNA snRNA, the Sm site that is critical for the assembly of the Sm ring and the 3′ terminal stem–loop consist of ribonucleotides. Core snRNPs assembled on Sup U7 and Sup U7/D snRNPs were purified by size exclusion chromatography and tested in the presence of the HCCFL and FLASH for the ability to cleave ΔSL/Ys RNA (Fig. 6G, lane 1) or its deoxy equivalent, DNA/50Ys (Fig. 6G, lane 4). Importantly, while the RNA substrate was efficiently cleaved in the presence of both Sup U7 and Sup U7/D snRNPs (Fig. 6G, lanes 2,5), the DNA substrate yielded cleavage product only in the presence of Sup U7 snRNP (Fig. 6G, lane 5), with Sup U7/D snRNP remaining inactive (Fig. 6G, lane 6). Thus, formation of a double-stranded DNA between the substrate and the chimeric U7 snRNA does not support processing, likely due to the inability of the NTD of symplekin to recognize the B-form helix. Note that ΔSL/Ys (80 nt) is longer than DNA/50Ys (50 nt) due to the presence of additional nucleotides at the 3′ end.

**DISCUSSION**

U7 snRNP is an RNA-guided endonuclease that acts in conjunction with SLBP to cleave replication-dependent histone pre-mRNAs at the 3′ end, yielding mature histone mRNAs terminated with a conserved stem–loop and the ACCCA single stranded tail. Of 13 components of the U7 snRNP, four are shared with the canonical cleavage/polyadenylation machinery that generates mRNAs ending with a poly(A) tail: symplekin, CPSF100, CPSF73 and CstF64. We recently reconstituted U7 snRNP entirely from recombinant components to analyze its endonucleolytic activity and to determine its structure in complex with histone pre-mRNA and SLBP by cryo-EM (Sun et al. 2020). Our studies demonstrated that the recombinant U7 snRNP cleaves histone pre-mRNA at the correct site, with the activity depending on the presence of the NTD of symplekin and an intact catalytic site in CPSF73. Strikingly, the NTD is functional both in cis (as part of full-length symplekin) and in trans (as a separate polypeptide) and acts by recognizing the duplex RNA formed between the U7 snRNA and the substrate. This interaction likely initiates a set of conformational changes within U7 snRNP that result in the opening of the catalytic site of CPSF73 for substrate binding and cleavage (Sun et al. 2020). Interaction of the SSU72 phosphatase with the NTD interferes with its ability to recognize the duplex, inhibiting the activity of U7 snRNP. In this report, we analyzed various functional and structural aspects of the fully recombinant U7 snRNP in more detail and investigated whether its catalytic component, CPSF73, is solely an RNA-specific endonuclease, or it can also act as a 5′–3′ exonuclease, as suggested by previous studies with nuclear extracts on endogenous U7 snRNP (Yang et al. 2009b; Dominski 2010; Dominski et al. 2013).
Minimal composition of recombinant U7 snRNP that supports cleavage activity

In agreement with our previous report, fully recombinant U7 snRNP accurately cleaved histone pre-mRNA, selecting the same site that is used by endogenous U7 snRNP in vivo or in nuclear extracts. No cleavage was observed if CPSF73 was mutated within its catalytic site, or when two essential components of U7 snRNP, the core consisting of U7 snRNA and the Sm ring, or the NTD of symplekin, were omitted. Consistent with U7 snRNPs acting as an RNA-guided endonuclease, mutations within U7 snRNA that reduced base-pairing to the substrate impaired processing and increased its dependence on SLBP. These results highlight the importance of efficient recruitment of the U7 snRNP to histone pre-mRNA for the cleavage reaction and demonstrate that CPSF73 is unable to act as an endonuclease by itself, requiring assembly into a larger complex for proper substrate recognition and catalytic activation. The position of the cleavage site selected by fully recombinant U7 snRNP could be shifted in a predictable manner by increasing the spacing between the 5′ end region and the Sm binding site, as explained by the length suppression effect (Scharl and Steitz 1994, 1996). Moreover, nucleotide substitutions that convert the U7-specific Sm binding site of recombinant U7 snRNP into a spliceosomal-type Sm site result in generation of an additional cleavage site, as previously observed for semirecombinant U7 snRNP (Bucholc et al. 2020). A plausible explanation for this reduced fidelity of processing is provided by the structure of the recombinant U7 snRNP, which revealed a number of specific contacts between the U7-specific Sm binding site in U7 snRNA and both histone pre-mRNA and proteins of the Sm ring, including Lsm10 and Lsm11 (Sun et al. 2020). All these interactions are incompatible with the spliceosomal Sm site, likely resulting in changes of the overall geometry of the Sm ring in U7 snRNP and misalignment of CPSF73 with the substrate.

CstF64 was identified as one of four major subunits of endogenous U7 snRNP shared with the cleavage and polyadenylation machinery but its role in 3′ end processing of histone pre-mRNAs remained elusive. Our studies with fully recombinant U7 snRNP demonstrate that this subunit is, at least in vitro, dispensable for 3′ end processing of histone pre-mRNAs. This conclusion is consistent with previous observations that Drosophila cells depleted of CstF64 show no detectable defect in U7-dependent processing (Wagner et al. 2007; Sabath et al. 2013). It is possible that in vivo CstF64 and U7 snRNPs are separate entities and only associate together as a result of cell lysis during nuclear extract preparation (Mili and Steitz 2004). Alternatively, CstF64 may have a role in vivo, but not in vitro, coordinating transcription of histone genes with processing of the resultant histone pre-mRNAs. This interpretation is consistent with RNAi experiments in mammalian cells showing partial defect in 3′ end processing of histone pre-mRNAs following depletion of CstF64 (Ruepp et al. 2011; Romeo et al. 2014).

The NTD of mammalian symplekin consists of seven pairs of antiparallel α-helices and folds into an arc, with the first (αA) and second (αB) helix in each pair forming the convex and the concave face, respectively (Xiang et al. 2010). SSU72 binds the NTD by making contacts with the concave face, specifically with αB helices of repeats 3–6 (Xiang et al. 2010). The cryo-EM structure of recombinant U7 snRNP in complex with histone pre-mRNA identified multiple residues within an overlapping region of the NTD, including repeats 1–4, that contact the RNA duplex, CPSF100 and subunits of the Sm ring (Sun et al. 2020). Mutating these residues significantly compromised the ability of the NTD to support cleavage of histone pre-mRNAs by recombinant U7 snRNP. Thus, the NTD of symplekin is engaged in a dense network of cooperative interactions that primarily serve to position its repeats 1–4 and its key positively charged residues near the RNA duplex, ultimately resulting in the catalytic activation by CPSF73. We note that the detection of the inhibitory effect of these mutations was facilitated by using the NTD as a separate polypeptide, which only transiently binds U7 snRNP (Sun et al. 2020), with the mutations further weakening this interaction.

Purified preparations of endogenous U7 snRNP contain substoichiometric amounts of CPSF160, WDR33, Fip1, and CPSF30 (Sabath et al. 2013; Yang et al. 2013; Skrajna et al. 2018). Recombinant U7 snRNP without these subunits is catalytically active, demonstrating that they are dispensable for processing of histone pre-mRNAs in vitro. They also do not have any stimulatory or inhibitory effect on the activity of recombinant U7 snRNP. Thus, CPSF160, WDR33, Fip1 and CPSF30 may bind to U7 snRNP during nuclear extract preparation due to reassociation of normally separated factors (Mili and Steitz 2004), or they may be legitimate components of the U7 snRNP, playing a role in integrating transcription and 3′ end processing during histone gene expression in vivo, as suggested for CstF64.

Recombinant U7 snRNP active in cleaving histone pre-mRNAs lacked other nuclear components previously implicated in the activity of U7 snRNP in vitro or in vivo, including ZFP100 (Dominski et al. 2002a), CFI, 68 kDa (Ruepp et al. 2010), CDC73/parafibromin (Farber et al. 2010; Ruepp et al. 2010), and the Y3′ ncRNA (Kohn et al. 2015). These components were not detected in stable processing complexes containing endogenous U7 snRNP (Skrajna et al. 2018), and our results with fully recombinant U7 snRNP demonstrated they are not essential for the cleavage reaction in vitro. Further studies are required to determine their relationship with histone gene expression in vivo.
5′–3′ Exonuclease activity of recombinant U7 snRNP

Previous studies with nuclear extracts demonstrated that endogenous U7 snRNP in addition to endonucleolytically cleaving histone pre-mRNAs displays an exonuclease activity that degrades the downstream cleavage product (DCP) in the 5′–3′ direction (Walther et al. 1998). UV-crosslinking experiments suggested that both activities are provided by CPSF73 (Dominski et al. 2005; Yang et al. 2009b; Dominski et al. 2013). The 5′–3′ exonuclease activity degraded the DCP also within the region that extensively base-pairs with U7 snRNA, raising the possibility that the endogenous U7 snRNP may function in conjunction with an unidentified helicase that unwinds the duplex facilitating the progression of CPSF73.

In this report, we took advantage of fully recombinant U7 snRNP consisting of strictly defined components to re-investigate this presumptive 5′–3′ exonuclease activity of CPSF73 in more detail. Our results with substrates containing site-specifically modified nucleotides and a single label at the 3′ end clearly demonstrated that recombinant U7 snRNP degrades DCP substrates in the 5′–3′ direction, mimicking the behavior of endogenous U7 snRNP. The 5′–3′ exonuclease activity of recombinant U7 snRNP, similarly to the endonuclease activity, required the NTD of symplekin in either cis or trans, and the recognition of the substrate by base-pairing interaction between the HDE and U7 snRNA. Importantly, the D75N/H76A mutation within the catalytic site of CPSF73 abolished the 5′–3′ exonuclease activity of U7 snRNP, supporting the notion that CPSF73 is a dual activity enzyme, with the endonuclease and 5′–3′ exonuclease activities depending on a single catalytic center and the same activation mechanism involving the NTD of symplekin. The degradation of the DCP by recombinant U7 snRNP extended into the duplex RNA without the requirement for an accessory helicase, consistent with the processive nature of the 5′–3′ exonuclease activity of CPSF73 and RNase J, a member of the same β-CASP family of nucleases (Mathy et al. 2007). Further studies are required to determine how CPSF73 discriminates between various substrates at the structural level, using either endonuclease or 5′–3′ exonuclease mode, as previously determined for RNase J (Richards and Belasco 2011), and how various nucleotide modifications in the substrate affect either of these two CPSF73 activities.

During 3′ end processing of canonical pre-mRNAs by cleavage and polyadenylation, CPSF73 remains bound to the upstream cleavage product that is subsequently polyadenylated, with the downstream cleavage product being degraded by ubiquitous 5′–3′ exonuclease, Xrn2. In vivo, this activity promotes transcriptional termination via the torpedo mechanism (Tollervey 2004). Our results support the previously proposed model suggesting that in 3′ end processing of histone pre-mRNAs the two activities are provided by CPSF73 (Yang et al. 2009b; Dominski 2010). In this model, CPSF73, as part of U7 snRNP, first uses its endonucleolytic activity to cleave histone pre-mRNA downstream from the stem–loop and remains associated with the DCP to catalyze its 5′–3′ degradation, ultimately displacing U7 snRNP from the base-pair interaction for another round of processing. In vivo, the 5′–3′ exonuclease activity of CPSF73 may promote transcriptional termination on histone genes, eliminating the requirement for Xrn2 (Yang et al. 2009b). In support of this hypothesis, recent in vivo studies demonstrated that depletion of Xrn2 impaired transcription termination on genes encoding canonical mRNAs but had no effect on transcription termination of replication-dependent histone genes (Eaton et al. 2018).

The ability of CPSF73 to act as both endonuclease and 5′–3′ exonuclease is consistent with in vivo and in vitro activities of other β-CASP members of the metallo-β-lactamase family, including Snm1A, Snm1B, Snm1C, and RNase J1/2 (Dominski 2007; Mathy et al. 2007; Clouet-d’Orval et al. 2010; Dorleans et al. 2011; Newman et al. 2011; Richards and Belasco 2011; Dominski et al. 2013; Phung et al. 2013). In contrast to the 5′–3′ exonuclease activity of Xrn2 (Tollervey 2004), the 5′–3′ exonuclease activity of CPSF73 in recombinant U7 snRNP does not require a phosphate group at the 5′ end to degrade the substrate, acting with the same efficiency on substrates containing a 5′ phosphate or 5′ hydroxyl group, consistent with the behavior of endogenous U7 snRNP (Yang et al. 2009b). The same relaxed requirement for the presence of either a hydroxyl or a phosphate group at the 5′ end of the substrate was previously shown for homologs of RNase J1 from various bacterial and archaeal species and a homolog of CPSF73 from the archaeon Pyrococcus abyssi (Mathy et al. 2007; Clouet-d’Orval et al. 2010; Dorleans et al. 2011; Newman et al. 2011; Richards and Belasco 2011; Phung et al. 2013).

CPSF73 acts as a site-specific endonuclease on DNA substrates

Intriguingly, recombinant U7 snRNP can also endonucleolytically cleave single stranded DNA substrates, with the activity depending on the same catalytic site in CPSF73 that acts on RNA substrates. These data are consistent with the structure of U7 snRNP in a complex with histone pre-mRNA poised for cleavage, which shows no major contacts of CPSF73 with 2′ hydroxyl groups in the RNA chain (Sun et al. 2020). The dual specificity toward both RNA and DNA substrates is also consistent with known in vitro activities of other β-CASP nucleases (Dominski 2007). However, at least some of these nucleases may have a preference for one or the other substrate type. Compared to single-stranded RNAs, single-stranded DNAs are clearly less favorable substrates for CPSF73, as demonstrated by partial stalling of the enzyme on deoxy-nucleotides during degradation of the DCP.
As in the case of RNA substrates, cleavage of single-stranded DNAs depends on substrate recognition by formation of a duplex with the 5′ region of U7 snRNA. Catalytic activation of CPSF73 also requires the NTD of symplekin, but in contrast to RNA substrates, it only functions in cis, as the integral part of symplekin. This difference likely reflects the fact that single-stranded DNAs are less suitable substrates for CPSF73, with their cleavage taking longer time and requiring a more stable interaction between the NTD and the processing complex.

Cleavage of single-stranded DNAs by recombinant U7 snRNP indicates that the NTD of symplekin is capable of recognizing the heteroduplex formed between the DNA substrate and the 5′ end of U7 snRNA. This is consistent with both the DNA/RNA heteroduplex and double-stranded RNA adopting the same A-type helix. No cleavage of DNA substrates was observed by recombinant U7 snRNP containing U7 snRNA whose 5′ end was composed of deoxynucleotides, although the same recombinant U7 snRNP efficiently cleaved RNA substrates. Thus, formation of a double-stranded DNA is incompatible with processing by U7 snRNP, likely due to the inability of the symplekin NTD to recognize the conformation of the structurally distinct B-helix that is typically adopted by DNA duplexes. This result highlights the versatile role of base-pairing interaction between U7 snRNA and the HDE in 3′ end processing by U7 snRNP, which is not limited to recruiting the substrate but also serves as a critical structural element required for the activation of CPSF73 (Sun et al. 2020).

Our finding that recombinant U7 snRNP is capable of cleaving single-stranded DNA substrates suggests that CPSF73 may play a role in some aspects of DNA metabolism in vivo. CPSF73 is unable to function as an enzyme without proper mechanisms of substrate recognition and catalytic activation, and one possibility is that it may act on DNA substrates in conjunction with a novel DNA-specific complex or as part of appropriately modified 3′ end processing machineries. Clearly, further studies are required to determine whether the enzymatic potential offered by CPSF73 has been used during evolution to perform functions in DNA metabolism.

MATERIALS AND METHODS

RNAs and DNAs

Sup U7/D snRNA (50 nt) was synthetized by GE Dharmacon and had the following sequence (written in 5′–3′ orientation): dGdGdTdTdTdAdAdGdGdGdAdAdAUAGAAAUUGU UAGCAGGCUGACUUUCGGCGGCC. Compared to the sequence of mouse U7 snRNA, the 3′ terminal stem-loop lacks 4 bp. DNA/50, DNA/50Ys, DNA/55Ys and DNA size markers (19-mer, 20-mer and 21-mer) were synthesized by Integrated DNA Technologies. All other RNAs, including mH2a* and dH3* pre-mRNAs, previously referred to as mH2a/2m-3′B and dH3/2m-3′B, respectively, were as described (Skrajna et al. 2018; Bucholc et al. 2020; Sun et al. 2020).

Expression of recombinant components

Subunits of the U7-specific Sm ring were expressed either in bacteria (SmB, SmD3, SmE, SmF, and SmG) or in insect cells using the baculovirus system (Lsm10 and Lsm11 with an internal deletion of amino acids 211–332), and assembled on U7 snRNA yielding core U7 snRNP, as described (Bucholc et al. 2020). The core U7 snRNP was either directly purified by size exclusion column (SEC), or pre-bound to bacterially expressed FLASH encompassing amino acids 51–137 (Aik et al. 2017; Sun et al. 2020), and the resultant complex was stored at −80°C, as described (Bucholc et al. 2020; Sun et al. 2020). SLBP (amino acids 125–270) with an amino-terminal 6xHis tag was expressed in bacteria (Bucholc et al. 2020). The HCC was expressed in insect cells using Multibac technology (Sari et al. 2016), as described (Sun et al. 2020), and stored at −80°C.

RNA labeling and in vitro processing

RNA and DNA substrates were labeled at the 5′ end with 32P using T4 polynucleotide kinase (New England Biolabs), as described (Dominski et al. 1999). T7-generated histone pre-mRNAs were treated with calf intestinal phosphatase (New England Biolabs) prior to the labeling to remove the 5′ triphosphate. Synthetic RNA and DNA substrates were labeled at the 5′ end without prior phosphatase treatment. DCP/1xM and DCP/3xD RNAs were labeled at the 3′ end with [5′-32P]pCp using high concentration T4 RNA ligase (New England Biolabs), as suggested by the manufacturer with the exception that only 50 ng (~ 5 pmol) of each substrate was used in the labeling reaction.

In vitro processing

Recombinant U7 snRNP was reconstituted in a final volume of 10 µL by mixing the core U7 snRNP, FLASH, the HCC and the symplekin NTD and tested for the ability to process radiolabeled pre-mRNA substrates in the presence of SLBP, as described (Sun et al. 2020). When indicated, individual components of the reaction were omitted, or the reaction was supplemented with SSU72 or an oligonucleotide complementary to the 5′ end of U7 snRNA. Processing in a mouse nuclear extract was carried out as described (Dominski et al. 2002b; Yang et al. 2011).

ACKNOWLEDGMENTS

This research was funded by the National Institutes of Health (NIH) grants R01GM029832 to Z.D. and W.F.M., and R35GM118093 to L.T. W.S.A. was supported by a fellowship from the Raymond and Beverley Sackler Center for Research at Convergence of Disciplines at Columbia University Medical Center.

Received May 5, 2020; accepted May 26, 2020.
REFERENCES

Aik WS, Lin MH, Tan D, Tripathy A, Marzullo WF, Dominski Z, Chou CY, Tong L. 2017. The N-terminal domains of FLASH and Lsm11 form a 2:1 heterotrimer for histone pre-mRNA 3′-end processing. PLoS One 12: e0186034. doi:10.1371/journal.pone.0186034

Bond UM, Yario TA, Steitz JA. 1991. Multiple processing-defective mutations in a mammalian histone premessenger RNA are suppressed by compensatory changes in U7 RNA both in vivo and in vitro. Genes Dev 5: 1709–1722. doi:10.1101/gad.5.9.1709

Bucholc K, Aik WS, Yang XC, Wang K, Zhou ZH, Ladleiz M, Marzullo WF, Tong L, Dominski Z. 2020. Composition and processing activity of a semi-recombinant holu U7 snRNP. Nucleic Acids Res 48: 1508–1530. doi:10.1093/nar/gkz1148

Burch BD, Godfrey AC, Gasdaska PY, Salzler HR, Duronio RJ, Marzullo WF, Dominski Z. 2011. Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila. RNA 17: 1132–1147. doi:10.1261/ma.2566811

Casanal A, Kumar A, Hill CH, Easter AD, Emsley P, Degliesposti G, Gordienko Y, Santhanam B, Wolf J, Wiederhold K, et al. 2017. Lsm11 is essential for histone pre-mRNA processing in vivo in Euryarchaeal. Mol Biol Cell 28: 2370–2380. doi:10.1093/mc/22.18.6648-6660.2002

Clouet-d’Orval B, Rinaldi D, Quentin Y, Carpousis AJ. 2010. Archaeal β-CASP proteins with homology to bacterial RNase J have 5′- to 3′-exonuclease activity. J Biol Chem 285: 17574–17583. doi:10.1074/jbc.M109.095117

Dominski Z. 2007. Nucleases of the metallo-β-lactamase family and their role in DNA and RNA metabolism. Crit Rev Biochem Mol Biol 42: 67–93. doi:10.1080/10409230701279118

Dominski Z. 2010. The hunt for the 3′ endonuclease. Wiley Interdiscip Rev RNA 1: 325–340. doi:10.1002/wrna.33

Dominski Z, Marzullo WF. 2007. Formation of the 3′ end of histone mRNA: getting closer to the end. Gene 396: 373–390. doi:10.1016/j.gene.2007.04.021

Dominski Z, Zheng LX, Sanchez R, Marzullo WF. 1999. Stem-loop binding protein facilitates 3′-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. Mol Cell Biol 19: 3561–3570. doi:10.1128/MCB.19.5.3561

Dominski Z, Erkmann JA, Yang X, Sanchez R, Marzullo WF. 2002a. A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3′-end processing. Genes Dev 16: 58–71. doi:10.1101/gad.932302

Dominski Z, Yang X, Raska CS, Santiago CS, Borchers CH, Duronio RJ, Marzullo WF. 2002b. 3′ end processing of Drosophila histone pre-mRNAs: requirement for phosphorylated dsSLBP and co-evolution of the histone pre-mRNA processing system. Mol Cell Biol 22: 6648–6660. doi:10.1128/MCB.22.18.6648-6660.2002

Dominski Z, Yang XC, Marzullo WF. 2005. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell 123: 37–48. doi:10.1016/j.cell.2005.08.002

Dominski Z, Carpousis AJ, Clouet-d’Orval B. 2013. Emergence of the β-CASP ribonucleases: highly conserved and ubiquitous metalloenzymes involved in messenger RNA maturation and degradation. Biochim Biophys Acta 1829: 532–551. doi:10.1016/j.bbapap.2013.01.010

Dorleans A, Sienna-Gallay L, Piton J, Zig L, Gilet L, Putzer H, Condon C. 2011. Molecular basis for the recognition and cleavage of RNA by the bifunctional 5′-3′ exo/endoribonuclease RNase J. Structure 19: 1252–1261. doi:10.1016/j.str.2011.06.018

Eaton JD, Davidson L, Bauer DLV, Natsume T, Kanemaki MT, West S. 2018. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev 32: 127–139. doi:10.1101/gad.308522.117

Farber LJ, Kort DJ, Wang P, Chen J, Teh BT. 2010. The tumor suppressor parafibromin is required for posttranscriptional splicing of histone mRNA. Mol Carcinog 49: 215–223. doi:10.1002/mc.20591

Kohn M, Ilbing C, Sinz A, Krohn K, Huttelmaier S. 2015. The Y3′ ncRNA promotes the 3′ end processing of histone mRNAs. Genes Dev 29: 1998–2003. doi:10.1101/gad.266486.115

Kolev NG, Steitz JA. 2005. Symplekin and multiple other polyadenylation factors participate in 3′-end maturation of histone mRNAs. Genes Dev 19: 2583–2592. doi:10.1101/gad.1371105

Mathy N, Benard L, Pellegrini O, Daou R, Wen T, Condon C. 2007. 5′-to-3′ exonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5′ stability of mRNA. Cell 129: 681–692. doi:10.1016/j.cell.2007.02.051

Mili S, Steitz JA. 2004. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA 10: 1692–1694. doi:10.1261/ma.7151404

Newman JA, Hewitt L, Rodrigues C, Solovyova A, Harwood CR, Lewis RJ. 2011. Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. Structure 19: 1241–1251. doi:10.1016/j.str.2011.06.017

Phung DK, Rinaldi D, Langendijk-Genevaux PS, Quentin Y, Carpousis AJ, Clouet-d’Orval B. 2013. Archaeal β-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. Nucleic Acids Res 41: 1091–1103. doi:10.1093/nar/gks1237

Pillai RS, Will CL, Luhrmann R, Schumperli D, Muller B. 2001. Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. EMBO J 20: 5470–5479. doi:10.1093/emboj/20.19.5470

Pillai RS, Grimmier M, Meister G, Will CL, Luhrmann R, Fischer U, Schumperli D. 2003. Unique Sm core structure of U7 snRNPs: assembled by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev 17: 2321–2333. doi:10.1101/gad.274403

Richards J, Belasco JG. 2011. Ribonuclease J: how to lead a double life. Structure 19: 1201–1203. doi:10.1016/j.str.2011.08.004

Romeo V, Griesbach E, Schumperli D. 2014. CstF-64: cell cycle regulation and functional role in 3′ end processing of replication-dependent histone mRNAs. Mol Cell Biol 34: 4272–4284. doi:10.1128/MCB.00791-14

Ruepp MD, Vivarelli S, Pillai RS, Kleinschmidt N, Azzou TN, Barabino SM, Schumperli D. 2010. The 68 kDa subunit of mammalian cleavage factor I interacts with the U7 small nuclear ribonucleoprotein and participates in 3′-end processing of animal histone mRNAs. Nucleic Acids Res 38: 7637–7650. doi:10.1093/nar/gkq613

Ruepp MD, Schweinruber C, Kleinschmidt N, Schumperli D. 2011. Interactions of CstF-64, CstF-77, and symplekin: implications on localisation and function. Mol Biol Cell 22: 91–104. doi:10.1091/mc.e10-06-0543

Sabath I, Skrajna A, Yang XC, Ladleiz M, Marzullo WF, Dominski Z. 2013. 3′-end processing of histone pre-mRNAs in Drosophila: U7 snRNP is associated with FLASH and polyadenylation factors. RNA 19: 1726–1744. doi:10.1261/rna.040360.113

Sari D, Gupta K, Thimiri Govinda Raj DB, Aubert A, Dmccova P, Garzoni F, Fitzgerald D, Berger I. 2016. The multiBac
baculovirus/insect cell expression vector system for producing complex protein biologics. Adv Exp Med Biol 896: 199–215. doi:10.1007/978-3-319-27216-0_13

Scharl EC, Steitz JA. 1994. The site of 3′ end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. EMBO J 13: 2432–2440. doi:10.1002/j.1460-2075.1994.tb06528.x

Scharl EC, Steitz JA. 1996. Length suppression in histone messenger RNA 3′-end maturation: processing defects of insertion mutant premessenger RNAs can be compensated by insertions into the U7 small nuclear RNA. Proc Natl Acad Sci U S A 93: 14659–14664. doi:10.1073/pnas.93.25.14659

Schafle F, Gilmartin GM, Bannwarth W, Birnstiel ML. 1986. Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3′ end of H3 messenger RNA. Nature 323: 777–781. doi:10.1038/323777a0

Schonermann L, Kuhn U, Martin G, Schafer P, Gruber AR, Keller W, Zavolan M, Wahle E. 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes Dev 28: 2381–2393. doi:10.1101/gad.250985.114

Schumperli D, Pillai RS. 2004. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. Cell Mol Life Sci 61: 2560–2570. doi:10.1007/s00018-004-4190-0

Skrajna A, Yang XC, Burch BD, Marzluff WF, Dominski Z. 2010. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. Nature 467: 729–733. doi:10.1038/nature09391

Yang XC, Sullivan KD, Marzluff WF, Dominski Z. 2009b. Studies of the 5′ exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. Mol Cell Biol 29: 31–42. doi:10.1128/MCB.00776-08

Yang XC, Xu B, Sabath I, Kunduru L, Burch BD, Marzluff WF, Dominski Z. 2011. FLASH is required for the endonucleolytic cleavage of histone pre-mRNAs but is dispensable for the 5′ exonucleolytic degradation of the downstream cleavage product. Mol Cell Biol 31: 1492–1502. doi:10.1128/MCB.00979-10

Zhang J, Tan D, DeRose EF, Perera L, Dominski Z, Marzluff WF, Tong L. 2019. Structural insights into the human pre-mRNA 3′-end processing machinery. Mol Cell 77: 800–809.e6. doi:10.1016/j.molcel.2019.11.005

CPSF73 is an endonuclease and a 5′-3′ exonuclease