CHARACTERIZATION OF FACTORS REQUIRED FOR 3’ END PROCESSING OF HISTONE PRE-MRNAs

Ivan Gwilym Sabath

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics in the School of Medicine.

Chapel Hill
2014

Approved by:

Zbigniew Dominski

William Marzluff

Howard Fried

Greg Matera

Brian Strahl

Yue Xiong

ABSTRACT

Ivan Gwilym Sabath: Characterization of Factors Required for 3' End Processing of Histone pre-mRNAs
(Under the direction of Dr. Zbigniew Dominski)

Metazoan replication dependent histone mRNAs form their 3' end by endonucleolytic cleavage of the pre-mRNA transcript without the subsequent addition of a poly(A) tail. These are the only cellular mRNAs that are not polyadenylated but instead end in a highly conserved stem-loop. Analogous to the poly(A) tail of canonical mRNAs, the terminal 3' stem-loop of histone mRNAs is a crucial regulatory feature that provides an interface for mechanisms coupling the mRNA to cell-cycle regulation, stability, and translation. Histone pre-mRNA 3' end processing requires the U7 snRNP, which interacts with a purine rich sequence called the histone downstream element (HDE) located about 10 nts 3' of the cleavage site. The U7 snRNP is composed of a small 60 nucleotide snRNA bound with a U7 specific heptameric ring resembling that of splicesomal snRNPs except that SmD1 and SmD2 are replaced by Lsm10 and Lsm11, respectively. Unlike other Lsm proteins, Lsm11 has an extended N-terminal domain, which interacts with FLASH. The interaction between the N-terminus of Lsm11 and FLASH is essential for processing. My research shows that the native U7 snRNP has a larger, more complex, composite structure than previously described. The Lsm11/FLASH complex in the U7 snRNP stably associates with a subset of polyadenylation factors termed the histone pre-mRNA cleavage complex (HCC). This composite U7 snRNP can be isolated from mammalian nuclear

extracts in a single step biochemical purification. Mammalian HCC consists of all six subunits of CPSF in addition to symplekin and CstF64. Association of the U7 snRNP with the HCC is absolutely dependent on a highly conserved LDLY motif in FLASH. I define the region of Lsm11 required for binding FLASH and identify the region of Lsm11 in the context of an Lsm11/FLASH complex necessary to recruit the HCC. Additionally, my work identified the *Drosophila* HCC, which has a similar composition to the mammalian HCC but lacks two CPSF subunits, Fip1 and CPSF30. Using extracts prepared from *Drosophila* cultured cells with FLASH knocked down by RNA interference, I also determined the critical elements in FLASH necessary to support histone pre-mRNA processing *in vitro*.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

CB	cajal body
CPSF	cleavage and polyadenylation specificity factor
CStF	cleavage stimulation factor
FLASH	FLICE associated huge protein
HCC	histone pre-mRNA cleavage complex
HDE	histone downstream element
HLB	histone locus body
mRNA	messenger RNA
PAS	polyadenylation signal
pre-mRNA	precursor messenger RNA
RBD	RNA binding domain
RD	replication dependent
RNAi	RNA interference
SL	stem-loop
SLBP	stem loop binding protein

snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
snoRNA	small nucleolar RNA

CHAPTER 1: INTRODUCTION

ROLE OF HISTONES

The eukaryotic genome is packaged in a tightly and highly ordered arrangement of nucleo-protein units called nucleosomes. These repeating units form the basis of chromatin. At the center of the nucleosome is an octameric core formed from two copies each of the four histone proteins H2A, H2B, H3 and H4. At the center of the core, H3 and H4 exist as dual heterodimers with two H2A/H2B heterodimers on either side of the H3/H4 tetramer. There are 147 base pairs of DNA wrapped 1.67 helical turns around each octamer with a stretch of linker DNA approximately 10-80 base pairs between nucleosomes. While the ordered core of the histone proteins in the nucleosome make direct contact with the DNA, a significant portion of the histone proteins protrude out beyond the nucleosome as less ordered N-terminal tails available for a variety of regulatory post-translational modifications collectively termed the histone code (Strahl and Allis, 2000). The core nucleosome particle also interacts with the histone H1 linker protein. H1 binds the linker DNA to the octamer at the point at which 5' and 3' ends of the wrapped DNA make contact with the core nucleosome. H1 promotes compaction of nucleosomes to form a dense chromatin structure. This chromatin environment serves two critical functions: protecting the integrity of the genome and providing complex architecture for regulation of gene expression.

There are two classes of histones in multicellular eukaryotes: those that are replication-dependent (RD) and those that are replication-independent. RD histone genes encode the major

core histone proteins H2A, H2B, H3, H4 and H1 that form the nucleosome. Expressed only when DNA is replicated, these five histones form the primary bulk of proteins constituting chromatin. Their expression is tightly coupled to cell-cycle progression and coordinated with DNA replication during S-phase when there is a need to package newly synthesized DNA. RD histone mRNA levels peak during S-phase with a 30 fold induction of available mRNA when the demand for histone proteins is greatest. In vertebrates, regulation of mRNA levels is the result of changes in rates and efficiency of transcription, mRNA 3' end formation, and mRNA stability with 3' end formation playing the most significant role (Fig. 1). In metazoans, RD histone mRNAs are the only messenger RNAs (mRNAs) that do not have poly(A) tails. Furthermore, their precursor messenger RNA (pre-mRNA) transcripts are monocistronic and lack introns.

Replication-independent histones, or histone replacement variants, are not cell-cycle regulated but instead expressed constitutively at basal levels throughout the cell-cycle. They are expressed from genes scattered through the genome and can substitute for RD histones under a number of circumstances including transcriptional activation, kinetochore assembly, and DNA repair (Sarma and Reinberg, 2005). A significant difference from the RD genes is that pre-mRNA transcripts from replication-independent genes often contain introns and the mRNAs are polyadenylated.

ORGANIZATION OF HISTONE GENES AND NUCLEAR BODIES

In general, metazoan RD histone genes exist in the genome as large clusters of either tandem repeats or jumbled clusters. The human RD histone genes are found in three clusters distributed between chromosome 1 and chromosome 6. The HIST1 cluster on chromosome 6 consists of 55

histone genes: 49 for the core histones and 6 for the linker H1. Two smaller clusters, HIST2 and HIST3, lie on chromosome 1 and contain 6 and 3 histone genes respectively (Marzluff et al., 2002). In contrast, histone genes in *Drosophila melanogaster* are located at just one locus on chromosome 2L 39DE (Lifton et al., 1978). These genes are organized in 100 tandem 5.1kB repeating units each encoding one copy of all five RD histones (Fig. 2). The *Drosophila* H3 and H4 genes, and the H2a and H2b genes, are in pairs with divergent transcription from shared promoters while the histone H1 gene has its own promoter. Grouping large numbers of histone genes into distinct clusters is an evolutionary conserved feature unique to RD histone genes, unlike many other large copy number gene families that are found scattered throughout the genome. This clustered arrangement is ideal for localizing *cis*-acting transcription and pre-mRNA processing factors at concentrations sufficient for increased transcription rates and efficient biosynthesis of RD histone pre-mRNA during S-phase.

Eukaryotes have evolved a complex system of cellular compartmentalization in order to regulate and ensure the efficiency of distinct biochemical processes. While some of these compartments such as mitochondria, the nuclear envelope, lysosomes, peroxisomes, and the endoplasmic reticulum, are clearly defined by membranes, others, like the nucleolus, can be characterized by protein-protein and RNA-protein interactions that bring together factors in a macromolecular body lacking a membrane but still possessing a specialized function. Since these bodies lack a lipid membrane, constituent components can rapidly exchange between the macromolecular structure and its surrounding nucleoplasm or cytoplasm environment.

A particularly well studied nuclear body present in multicellular organisms is the Cajal Body (CB), described by Ramon y Cajal over 100 years ago. One of the first discovered and most abundant proteins in CBs is p80 coilin, a 576 amino acid protein that has an organizational

role in the formation of the bodies (Liu et al., 2009) and which is used as a marker to identify CBs. These bodies play a key role in the biogenesis and organization of RNA processing machinery. Despite not being a center for splicing activity, the U1, U2, U4, U5 and U6 small nuclear RNAs (snRNAs) and their associated proteins that together form the splicesomal small nuclear ribonucleoprotein particles (snRNPs) can be found in CBs. Although the connection between CBs and splicing snRNPs is not completely understood, there are indications that upon return to the nucleus and association with CBs, the assembled snRNPs are subject to further modification. For example, the CB-specific RNA (scaRNA) U85 contains a C/D and H/ACA motif similar to that found in small nucleolar RNAs (snoRNAs) and in conjunction with the methyltransferase fibrillarin and pseudouridine synthase dyskerin/NAP57/CBF5 complex has been shown to guide 2′-O-methylation and pseudouridylation respectively of distinct nucleotide ribose moieties on the U5 snRNA (Matera and Ward, 1993; Almond and Carmo-fonseca, 1993). This could be considered analogous to the way ribosomal RNAs (rRNAs) are modified in the nucleolus by snoRNAs.

Work in *Xenopus* oocytes revealed the U7 snRNA, a snRNA specific to RD histone biogenesis, as a component of CBs (Wu and Gall, 1993). However, additional observations in mammals and *Drosophila* suggested not all CBs contained U7 snRNA and that in fact there were at least two different bodies forming. *In situ* hybridization showed a portion of U7 snRNA positive CBs localized to the histone gene loci on chromosomes 1 and 6 in humans (Frey and Matera, 1995) and that in *Drosophila* all U7 snRNA positive CBs co-localized to the histone genes on chromosome 2L (Liu et al., 2006). These specialized bodies were named histone locus bodies (HLB) due to their vicinity to the histone genes and are identifiable during G1, G2, and S-

phase. Harmony Salzler showed the bi-directional promoter region located between H3 and H4 is sufficient to promote assembly of ectopic HLBs in *Drosophila* (Salzler et al., 2013).

The protein p220 nuclear protein ataxia-telengiectasia locus (NPAT) is also found constitutively in HLBs. It is required for the G1 to S-phase transition (Ma et al., 2000) and expression of histone genes during S-phase (Zhao et al., 2000). The C-terminal half of NPAT contains domains critical for G1 to S-phase transition. During the start of S-phase, NPAT is a substrate of cyclin E-cyclin dependent kinase 2 (CDK2/cyclin E) and is phosphorylated at five sites within its C-terminal half resulting in an increased level of histone gene expression (Ma et al., 2000). Regions in the N-terminal portion of NPAT are critical for its role as a histone gene co-activator. NPAT contains an approximately 30 amino acid LisH domain at its N-terminus, which is known to be essential for histone gene transcription. Mutations at conserved residues in this region block histone gene transcription but do not prevent localization to the HLB (Wei et al., 2003). NPAT has also been shown to bind histone nuclear factor P (HiNF-P), a transcription factor for H4 genes, and disrupting this interaction inhibits the transcription of the H4 gene at the G1 to S-phase transition (Miele et al., 2005).

CANONICAL PRE-MRNA 3' END PROCESSING

The vast majority of mRNAs are polyadenylated at the 3' end in a two-stage reaction. In the first stage, the cleavage and polyadenylation machinery recognizes a conserved AAUAAA hexanucleotide sequence called the poly(A) signal (PAS) and a G/U rich element 3′ of PAS. Cleavage occurs between these two *cis*-acting elements before proceeding to a tightly coupled

second reaction where the cleaved pre-mRNA 3' end is appended with a poly(A) tail of approximately 200-250 adenosines added cotranscriptionally by poly(A) polymerase (PAP).

Cleavage coupled to polyadenylation requires at least five multi-subunit protein complexes to proceed including cleavage polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CF I_m), cleavage factor II (CF II_m), in addition to symplekin, PAP. CPSF consists of six protein subunits: CPSF30, CPSF73, CPSF100, CPSF160, Fip1, and WDR33. The CPSF complex recognizes the AAUAAA poly(A) signal through an interaction with CPSF160 (Mandel et al., 2008). The endonuclease that cleaves the transcript after a CA dinucleotide is CPSF73 (Mandel et al., 2006). CPSF30 binds a uridine rich sequence 5' of the PAS (Barabino et al., 2000). Fip1 was identified as a member of CPSF that binds several cleavage and polyadenylation factors including CPSF30 and CstF77 (Helmling et al., 2001) and CPSF160 (Kaufmann et al., 2004). CPSF100, the catalytically inactive homologue of CPSF73, forms a heterodimer with the endonuclease (Dominski et al., 2005; Kolev et al., 2008). The protein WDR33 was also found to bind CPSF73 and co-purify with the CPSF complex (Shi et al., 2009).

CstF is a heterotrimeric RNA binding complex consisting of CstF77, CstF50, and CstF64 which recognizes and binds the downstream G/U rich element through an RNA binding domain (RBD) in CstF64 (Takagaki and Manley, 1997). CstF77 was shown to interact with both CstF64 and CstF50 in addition to self-associating through its C-terminus. CstF50 can also form a homodimer through its N-terminus. CstF77 binds symplekin, an essential processing factor that likely bridges the CPSF and CstF complexes, potentially serving a crucial role as a scaffolding protein (Hatton et al., 2000). CstF64 also binds symplekin. The ability of CstF64 to interact with CstF77 or symplekin is mutually exclusive as increasing amounts of symplekin protein *in vitro*

inhibits the efficiency of CstF77 binding to CstF64, and the same region of CstF64 binds both CstF77 and symplekin.

In higher eukaryotes, CF I_m is necessary for stimulating cleavage and polyadenylation (Brown and Gilmartin, 2003). CF I_m is a heterodimer consisting of eithera 59, 68, or 72 kDa subunit (CF I_m 68) and a 25 kDa subunit (CF I_m 25) (Rüegsegger et al., 1998). CF I_m 68 contains an N-terminal RNA recognition motif (RRM) that mediates its interaction with CF I_m 25 and enhances the ability of the CF I_m 25 RBD to bind a UGUA sequence element located 5' of the PAS. Evidence indicates CF I_m plays a role in recognition of canonical and non-canonical poly(A) sites (Venkataraman et al., 2005) and a role for CF I_m in inhibiting usage of proximal polyadenylation sites has been described as a mechanism for regulating alternative polyadenylation (Gruber et al., 2014).

The CF II_m heterodimer comprised of Clp1 and Pcf11 is also required for cleavage and polyadenylation in higher eukaryotes. Clp1 is an RNA kinase that has been shown to be essential for cleavage but dispensable for polyadenylation. Pcf11 binds pre-mRNA in proximity to the PAS (Kim et al., 2004) and associates with the C-terminal domain (CTD) of RNA Polymerase II (Pol II) (Sadowski et al., 2003), linking transcription and 3' end processing of canonical mRNAs. Mutations that affect the ability of Pcf11 to interact with the CTD result in transcription termination defects and are lethal yet do not impact normal 3' end formation. The CF II_m complex co-purifies in relatively low amounts with the rest of the cleavage and polyadenylation machinery and is not stably associated with a complex assembled on pre-mRNA, suggesting that it may play a regulatory role in 3' end processing of pre-mRNAs.

Taken altogether, the protein components of the polyadenylation machinery form an extended network of interactions that ultimately position the cleavage site for precise

endonucleolytic cleavage of the pre-mRNA (Fig 4). After the cleavage of the pre-mRNA, PAP generates the poly(A) tail that is essential for nuclear export, translational regulation, and the general stability of the mRNA. It is entirely possible that additional unknown protein factors that participate in 3' end processing of canonical pre-mRNAs yet remain to be discovered.

EARLY INSIGHTS INTO HISTONE PRE-MRNA 3' END PROCESSING

Similar to other eukaryotic mRNAs, histone mRNAs have a 7-methylguanosine cap and require two *cis*-acting sequence elements in the pre-mRNA for efficient formation of the mature transcript 3' end. But the similarities end there. Unlike other eukaryotic mRNAs, histone mRNAs lack introns and could not be purified by binding to poly(dT)-cellulose suggesting that this class of mRNA lacked a poly(A) tail. The RD histone pre-mRNA sequence elements necessary for 3' end processing are a stem-loop structure 5' of the cleavage site and a purine-rich element 3' of the cleavage site named the histone downstream element (HDE) (Fig. 3). These two elements differ drastically from those required for canonical pre-mRNA 3' end formation.

Cloning of the tandem repeat histone genes in sea urchins (Cohn et al., 1976) led to the discovery of a common nucleotide sequence near the end of replication-dependent histone genes. This sequence contains a 16 nucleotide stem-loop structure present at the 3' end of the mature mRNA and a separate purine rich HDE 3′ of the stem-loop that is absolutely conserved in sea urchin histone genes but more loosely defined in vertebrates (Birnstiel et al., 1985). Additional cloning of RD histone genes in other organisms indicated the stem-loop is a highly conserved feature of metazoa with several features being invariant including sequences in the stem and nucleotides immediately pre and post stem-loop (Birchmeier et al., 1982). Only in *C. elegans*

does the 3' end of H1 mRNA lack the stem-loop and is instead polyadenylated. Because of similarities to bacterial transcriptional termination signals, researchers initially thought the stem-loop structure might function as a transcription terminator. However, experiments using synthetic chicken H2b histone pre-mRNA injected into *Xenopus* oocytes indicated the 3' ends of mature histone mRNAs were formed by a processing reaction as opposed to transcription termination (Krieg and Melton, 1984).

The mechanism of histone pre-mRNA processing was first investigated using a biochemical complementation assay. When injected into *Xenopus* oocytes, sea urchin histone genes are all expressed and properly processed with the exception of H3. Properly processed H3 histone mRNA was only formed if injected into the *Xenopus* oocyte in conjunction with extract from sea urchin embryos. Sucrose gradient fractionation of sea urchin embryo extract identified a fraction sedimenting at 12S that could restore normal processing to H3 pre-mRNA when co-injected into *Xenopus* oocytes, suggesting a key factor missing in the *Xenopus* oocyte was present in this fraction. The specific RNA molecule required to catalyze correct formation of H3 3' ends was isolated by deproteinizing the poly(A)$^-$ RNAs of the 12S fraction and preparing a second sucrose gradient from this pool. A small RNA of approximately 60 nucleotides in the 4S fraction was identified by S1 protection assays using ^{32}P labeled RNAs to identify protected regions of histone pre-mRNA (Strub and Birnstiel, 1986).

A more detailed analysis of histone pre-mRNA processing was possible after the development of efficient *in vitro* processing assays using mammalian nuclear extracts. *In vitro* transcribed and 7-methylguanosine capped mouse H4 histone pre-mRNA was incubated with HeLa nuclear extracts to show that histone mRNA maturation depended on endonucleolytic cleavage between the stem-loop structure and the HDE (Gick et al., 1986; Mowry and Steitz,

1987). Cleavage occurred about 4-5 nucleotides 3′ of the stem-loop and favored a CA dinucleotide substrate, similarly to the conditions in canonical cleavage/polyadenylation and cleavage occured in 20mM EDTA. Once cleavage occurs, the downstream cleavage product (DCP) 3' of the cleavage site is degraded (Walther et al. 1998)

ROLE OF RNA IN HISTONE PRE-MRNA 3' END PROCESSING

In eukaryotes, a number of small nuclear RNAs (snRNAs) are known to play key roles in RNA metabolism. U1, U2, U4, U5 and U6 snRNAs form RNA-protein complexes known as small nuclear ribonucleoprotein particles (snRNPs) that participate in the splicing reactions responsible for removing introns from pre-mRNA (Jurica and Moore, 2003; Nilsen, 2003). In a number of animal species, these snRNPs can be immunoprecipitated using sera from patients suffering from the auto-immune disease systemic lupus ertythermatosus. The anti-Sm sera specifically recognize the protein component of the snRNP and not the RNA (Lerner and Steitz, 1979; Billings et al., 1982). Enrichment of snRNPs from sea urchin ovary extracts using anti-Sm sera resulted in the identification of a very low abundance small RNA that migrated at the same size as the active one identified in the 4S fraction in the sea urchin/*Xenopus* biochemical complementation assay. When isolated, this RNA stimulated the formation of sea urchin mature H3 mRNA 3' ends in *Xenopus* oocytes. cDNA copies of the RNA component were cloned for sequencing and the unique RNA named U7 snRNA (Strub et al., 1984). The authors noted that the 5' end of U7 snRNA contained partial sequence complimentary to the conserved CAAGAAGAA sequence in the HDE located 3' from the stem-loop.

The importance of sequence complementarity between sea urchin U7 snRNA 5' end and the HDE of histone pre-mRNA was confirmed using complementation assays in *Xenopus* oocytes. Plasmids containing the wild type sea urchin *Psammechinus miliaris* H3 genes or a variety of mutant H3 genes with alterations in the HDE sequence were injected into sea urchin *Paracentrotus lividus* eggs where they are transcribed and subjected to the same regulatory pathways as endogenous histone genes after fertilization. Mutations in the HDE of H3 resulted in a reduction or loss of correctly processed 3' ends while compensatory changes in the sequence of the U7 snRNA restored processing to wild type levels (Schaufele et al., 1986). Kim Mowry and Joan Steitz identified and sequenced the U7 snRNA from humans. The region complementary to the histone pre-mRNA was identified in human U7 snRNA using HeLa cells to express mouse H2A with purine to pyrimidine block substitutions in the HDE, which abolished processing. These were then rescued by co-transfection of a synthetic U7 snRNA containing compensatory changes complementary to the mutated histone HDE (Bond et al., 1991).

There are two classes of snRNAs. Sm-class snRNAs have a 5' trimethylguanosine cap, a 3′ stem-loop structure proceeded by a Sm site for binding a heptameric ring composed of Sm proteins B/B', D1, D2, F, E, G, and D3. Sm-class snRNAs include the U1, U2, U4, U5, (U11, U12, $U4_{atac}$). They are transcribed by RNA polymerase II, as suggested by the cap-structure. A second class of snRNAs is transcribed by RNA polymerase III, and hence end in a short oligo(U) tail. Lsm-class snRNAs such as U6 have a 5' monomethylphosphate cap, a 3′ stem-loop structure, and end in a stretch of uridines, which form the Lsm site for binding a heptameric ring composed of Lsm proteins 2-8 (Matera et al., 2007). Sm/Lsm proteins contain two Sm sequence motifs, Sm1 and Sm2, which form a characteristic Sm-fold. Sm1 is comprised of an α helix and three anti-parallel β strands while the Sm2 sequence contributes two additional anti-paralell β

strands. Between the two motifs is a linker of variable length. Contacts between the subunits of the heptameric ring are primarily mediated by the anti-parallel interactions of the fourth and fifth β strands in Sm2 although the α helix and other amino acid residues also contribute at the interface of subunits (Khusial et al., 2005).

The U7 snRNA identified by Schaufle et al. contained a non-canonical Sm binding site with consensus of AAUUUGUCUAG which was similar to but distinct from those found in the Sm-class of splicesomal snRNAs (Grimm et al., 1993). The functionality of the Sm binding site for binding Sm proteins was confirmed in a series of experiments using plasmids expressing sea urchin U7 genes under a *Xenopus* U2 promoter that were injected into *Xenopus* oocytes. Anti-Sm sera precipitated the sea urchin U7 snRNA and associated proteins while key mutations or deletions within the U7 Sm binding site disrupted the formation of the snRNP (Gilmartin et al., 1988).

The unique Sm binding site in U7 snRNA associates with an unusual Sm protein ring. In order to isolate and identify the protein components of the U7 snRNP, Pillai et al. successfully adapted a purification scheme in which Sm-class snRNPs are first enriched from HeLa nuclear extracts by immunoprecipitation using anti-trimethylguanosine cap antibodies followed by anion exchange column chromotography on a Resource Q column. After identifying a fraction containing U1 and U7 snRNA, a biotinylated anti-sense oligonucleotide complementary to the 5'end of U7 snRNA was used to affinity purify the U7 snRNP. Although Sm proteins B/B', D1, D2, F, G, and D3 were detected by Western blot analysis in the fraction containing U1 and U7 snRNA (no E antibody available), D1 and D2 were not present in the U7 anti-sense oligonucleotide affinity purified material. Furthermore, a unique peptide band migrating at 14 kDa in the U7 snRNA column fraction was analyzed by mass spectrometry and identified as a

new Sm-like protein termed Lsm10. Lsm10, but not SmD1 or SmD2, was shown to coprecipitate with U7 snRNA in mammalian (293T) cells in a manner dependent on the sequence of the Sm-site (Pillai et al., 2001a). A second peptide band migrating around 50 kDa from an U7 snRNP anti-sense oligonucleotide affinity purified fraction was later identified as Lsm11 and shown to (Parsons et al.)cross-link directly to U7 snRNA (Pillai et al., 2003). Sequence alignments between known Sm/Lsm proteins and Lsm10/Lsm11 indicated key residues are conserved in the Sm1 and Sm2 domains.

The 170 amino acid N-terminal domain of Lsm11 is particular to this specific protein and not found in other Sm/Lsm family members. Processing assays in *Xenopus* oocytes indicated the unique N-terminus portion of Lsm11 preceding the two Sm domains was critical for processing. *In vitro* transcribed, capped, and polyadenylated mRNAs encoding a number of various HA-tagged N-terminal deletions of Lsm11 were injected into the oocytes followed by the subsequent injection of a labeled chimeric histone-U7 snRNA substrate (Stefanovic et al., 1995a; Stefanovic et al., 1995b). Immunoprecipitiation of cytoplasmic extracts prepared from the oocytes using an anti-HA antibody showed removing the first N-terminal 104 amino acids of Lsm11 resulted in a marked reduction of cleavage of the chimeric probe (Pillai et al., 2003). Additionally, when mouse nuclear extract was incubated with bacterially expressed recombinant N-terminally tagged 136 amino acids of Lsm11, processing efficiency of the extract diminished significantly suggesting a critical role for N-terminal Lsm11 in the functional binding of other factors required for 3’ end processing of histone pre-mRNAs.

ROLE OF SLBP IN HISTONE PRE-MRNA 3' END PROCESSING

The highly conserved stem-loop structure in histone pre-mRNA contains a six base pair stem and four nucleotide uridine loop. The 5' and 3' single stranded flanking regions are also conserved. A 32 kDa protein named the stem-loop binding protein (SLBP) binds the stem-loop and 3' end processing in *Drosophila* critically depends on this interaction (Dominski et al., 2005b). Sequestering SLBP using a stem-loop oligonucleotide competitor completely inhibits processing *in vitro*. In mammals, SLBP stabilizes the base pairing of the U7 snRNP with the HDE and is somewhat dispensable if the strength of the duplex formed between the HDE and 5' end of U7 snRNA is sufficient *in vitro* (Dominski et al., 1999). SLBP levels increase with the onset of S-phase, when it is necessary to promote the upregulation of 3' end processing (Whitfield et al., 2000). After cleavage, SLBP remains bound to the mature message and plays a key role in message stability (Zheng et al., 2003), nuclear export (Sullivan et al., 2009a), and translation (Gorgoni et al., 2005; Sanchez and Marzluff, 2002).

SLBP contains a conserved RBD domain (amino acids 125-200) (Wang et al., 1996) comprised of three helices, αA, αB, and αC (Tan et al., 2013). Only helices αA and αC make contacts with the stem-loop, with helix αA making contact with the 5' flanking region. Interestingly, SLBP only recognizes one nucleotide on the 5' arm of the stem and none on the 3' side. Helix αC is closest to the stem loop and interacts with a guanine base at the second position in the stem. It also contacts the single stranded 5' flanking sequence previously shown to be essential for SLBP binding (Dominski et al., 2003). The 5' stem and flanking sequence interactions occur through residues located in the N-terminus of helix αC while the C-terminal end of the helix touches a uridine in the loop. Since helix αC interacts with three distinct positions of the stem-loop (the 5' flanking region, the 5' arm of the stem-loop, and the loop), it

has been hypothesized it may play the role of a ruler to measure the length of the stem (Tan et al., 2013).

IDENTIFICATION OF SYMPLEKIN AS A HEAT LABILE FACTOR AND CPSF73 AS THE ENDONUCLEASE

Mildly heating HeLa nuclear extracts blocks histone pre-mRNA 3' end processing *in vitro*. A heat labile factor (HLF) of at least 40 kDa was inactivated at 48˚C but resistant to micrococcal digestion and precipitation by anti-Sm sera suggesting another unique factor is required for histone mRNA 3' end formation (Gick et al., 1987). Fractionation of HeLa nuclear extracts and characterization of the fractions containing HLF led to the identification of symplekin as the heat sensitive protein required for processing. Surprisingly, a number of other cleavage and polyadenylation factors also co-purified with the HLF fraction including five subunits of CPSF (CPSF73, CPSF100, CPSF30, Fip1, and CPSF160) and two subunits of CstF (CstF64 and CstF77) (Kolev and Steitz, 2005). Symplekin is a protein originally identified in tight junctions (Keon et al., 1996) and later shown to be a factor in cleavage and polyadenylation of canonical pre-mRNAs, where it has been suggested to play a role as a scaffolding protein (Takagaki and Manley, 2000b). But the identification of the specific endonuclease required for cleavage of histone pre-mRNA 3' ends remained elusive.

CPSF73 was identified as the endonuclease responsible for cleavage of histone pre-mRNA 3' ends in our lab by Dr. Zbigniew Dominski. In active nuclear extracts, CPSF73 cross-links to the cleavage site in histone pre-mRNA substrates during U7-dependent 3′-end processing in addition to cross-linking with the downstream cleavage product (DCP) that undergoes 5' to 3′ exonucleolytic degradation after release from the mature mRNA (Dominski et al., 2005c).

CPSF73 and its homologue CPSF100 belong to the β-CASP family, a separate group of metallo-β-lactamase superfamily enzymes (Dominski, 2007; Callebaut et al., 2002). As determined by phylogenic sequence comparisons, typical metallo-β-lactamases possess five motifs enriched in histidine and aspartic acid residues. In general they coordinate two metal ions, most often Zn^{2+}(Aravind, 1999). The metallo-β-lactamase domains in CPSF73 and CPSF100 are lacking the fifth motif but contain the three motifs distinct to the β-CASP family. The site required for catalytic activity of CPSF73 is located in the cleft between the partial metallo-β-lactamase domain and the β-CASP domain. It is possible that CPSF73 may require the formation of a heterodimer with CPSF100 for the activation of its catalytic endonuclease activity (Dominski et al., 2005a). Recombinant CPSF73 is inactive for *in vitro* processing. A lack of conserved histidine residues in motif B is thought to render CPSF100 catalytically inactive but despite this variance it is still required for cleavage of histone pre-mRNA transcripts (Kolev et al., 2008). The identification of CPSF73 as the endonuclease in histone pre-mRNA 3' end processing led to work demonstrating it is also the endonuclease in canonical cleavage and polyadenylation (Mandel et al., 2006). The precise mechanism by which the endonuclease was recruited to histone pre-mRNA remained unknown.

FLASH IS ESSENTIAL FOR HISTONE PRE-MRNA 3' END PROCESSING

Through a yeast two-hybrid screen where the extended N-terminal portion of Lsm11 critical for processing was used as bait, multiple cDNAs constituting the first 139 amino acids of the caspase-8 binding protein FLICE associated huge protein (FLASH) were isolated (Yang et al., 2009a). FLASH was originally described as a cytoplasmic pro-apoptotic protein that interacts with the adapter protein FADD in conjunction with the ligand bound CD95/FAS receptor.

FLASH was shown to be a member of the death inducing signal complex (DISC) (Imai et al., 1999), an integral complex essential for activation of caspase-8 in FAS mediated extrinsic apoptosis. FLASH was also found to be involved in the activation of NF-κB by the TRAF2 pathway (Choi et al., 2001). Subsequent work clearly showed FLASH to be a mostly nuclear protein that localized in punctate dots casting doubt on the earlier characterization of FLASH's cytoplasmic roles (Milovic-Holm et al., 2007; Barcaroli et al., 2006).

Work by Barcaroli et al. was the first to suggest a role for FLASH in histone gene expression. In mammalian cells they showed FLASH was often associated with CBs and invariantly co-localized with NPAT at HLBs in a cell-cycle dependent manner. FLASH protein levels rise and fall during the cell-cycle with an increase during G1, a clear peak during S-phase and decrease in G2 (Barcaroli et al., 2006b). Down regulation of FLASH by short hairpin RNA (shRNA) in a variety of mammalian cell types resulted in the accumulation of cells in S-phase. It is not unusual for proteins that function in RD histone gene expression to cause cells to be blocked in S-phase when their function is impeded. When overexpressed, FLASH stimulated the activity of the H4 histone promoter in a luciferase assay while FLASH depleted cells exhibited reduced transcription of endogenous RD histone genes. Indeed, chromatin immunoprecipitation (ChiP) using anti-HA antibody against exogenously expressed HA-tagged FLASH coprecipitated the RD histone H2B, H3, and H4 gene promoters. In mammalian cells, S-phase progression, RD histone gene expression, and recruitment of FLASH to the HLB depend on the interaction between FLASH and Ars2 (Kiriyama et al., 2009), a protein shown to participate in cellular proliferation (Gruber et al. 2009) and biogenesis of microRNAs (Lobbes et al., 2006)

In mammals, FLASH is a large protein consisting of almost 2000 amino acids yet the N-terminal 139 amino acids are sufficient to stimulate processing of radio-labeled H2a histone pre-

mRNA when added in dilute HeLa nuclear extracts. Mutational analysis of this region of FLASH identified a highly conserved LDLY motif from amino acids 55 to 58 that is critical for processing *in vitro* (Yang et al., 2009a). The robust interaction between FLASH and the N-terminus of Lsm11 was confirmed by both directed yeast two-hybrid and *in vitro* pull down assays with GST tagged N-terminal 139 amino acids of FLASH and ^{35}S-labeled Lsm11. In both cases deletion of Lsm11 amino acids 1-40 prevented binding with FLASH. An orthologue of FLASH was identified in *Drosophila* and RNA interference (RNAi) using double stranded RNA targeting FLASH in S2 cells resulted in a substantial increase in mis-processed histone mRNA. These data confirmed FLASH as a newly identified essential histone pre-mRNA 3' end processing factor (Fig. 5). These results had just been published when I joined Dr. Dominski and Dr. Marzluff to study histone pre-mRNA 3' end processing.

SUMMARY

Histone biogenesis is critical to cell division. RD histone gene expression is invariably linked to cell-cycle progression with levels of histones peaking during S-phase and coordinated with DNA replication. Although increased transcription and RNA stability contribute to the elevated levels of RD histone mRNA in S-phase cells, the U7-dependent endonucleolytic cleavage to form a mature 3' end is the single most significant regulatory step that increases the pool of available mRNA for synthesis of histone proteins.

Several key elements and factors are essential for the formation of histone mRNA 3' ends. Two *cis*-acting sequence elements in the histone pre-mRNA are critical: the stem-loop and the HDE. The stem loop is bound by SLBP, which is absolutely required for processing in

Drosophila but somewhat dispensable in mammalian systems. Increases in SLBP levels are critical to the upregulation of U7-dependent 3' end processing in S-phase. The HDE performs a critical function by recruiting the U7 snRNP through partial base pairing with the U7 snRNA 5' end. A subset of polyadenylation factors consisting of CPSF73, CPSF100, and symplekin are required for processing. CPSF73 is the endonuclease that cleaves the transcript but the specific mechanism of recruitment to the cleavage site remained a mystery.

THESIS GOALS

The goal of this dissertation is to characterize the factors required for histone pre-mRNA processing. I focused on examining the composite nature of the U7 snRNP and sequence elements of FLASH necessary to recruit the polyadenylation factors required for processing, including CPSF100, CPSF73, and symplekin. In chapter 2, I describe biochemical techniques for purifying histone pre-mRNA 3' end processing factors from mammalian nuclear extracts using elements unique to U7-dependent processing. I also identify regions of Lsm11 that are essential for recruitment of 3' end processing factors. These studies indicate a subset of cleavage and polyadenylation factors, including CPSF73, CPSF100, and symplekin, associate with the U7 snRNP in a FLASH dependent fashion. This subset of cleavage and polyadenylation factors is named the histone pre-mRNA cleavage complex (HCC).

Although the initial studies were conducted in mammalian systems, much of my research effort focused on describing the HCC in *Drosophila*, using nuclear extracts prepared from Kc and S2 cell cultures. Chapter 3 addresses the identification of critical residues in *Drosophila* FLASH required for histone pre-mRNA 3' end processing and recruitment of a unique

Drosophila HCC. It also describes a surprising result that FLASH is required for degradation of the downstream cleavage product (DCP), whereas in mammalian systems FLASH is dispensable for this reaction. Furthermore, I examine the role of CstF64 in histone processing and describe the isolation of a large macromolecular complex comprised of all the major complexes required for canonical cleavage and polyadenylation with the exception of CF II_m. This is the first instance where these polyadenylation complexes have been purified together in *Drosophila*. Its composition contrasts sharply with the subset of polyadenylation factors that assemble in the *Drosophila* HCC.

In addition to these studies, I conducted a yeast-two hybrid screen with the N-terminal fragment of Lsm11 (amino acids 1-168) and a human universal tissue library. In this screen, I isolated an alternatively spliced variant of FLASH containing its first five exons comprising the first 138 amino acids of the protein and the last two exons that represent the 52 amino acids of the C-terminus. This alternatively spliced variant of FLASH excluded two huge internal exons that together account for a total of 1792 amino acids. Interestingly, while the extreme N-terminal region contains all the elements for processing *in vitro*, the C-terminal region was shown to interact with NPAT. This interaction localizes FLASH to the HLB and is described in Chapter 4.

Figure 1. Histone mRNA levels peak during S-phase. Expression of RD histone genes is tightly coupled to cell-cycle progression with peak levels of available histone mRNA occurring during S-phase. Several mechanisms contribute to the increased pool of mRNA and the processing of pre-mRNA plays a significant role in this process.

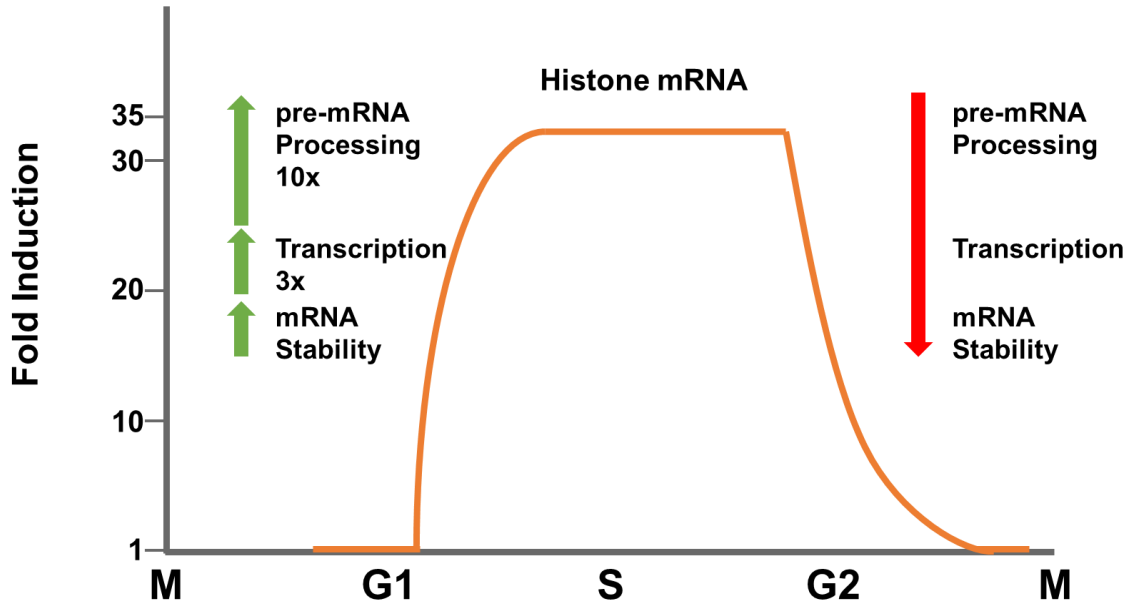

Histone mRNA
Fold Induction
35
30
20
10
1
pre-mRNA
Processing
10x
Transcription
3x
mRNA
Stability
pre-mRNA
Processing
Transcription
mRNA
Stability
M
G1
S
G2
M

Figure 2. Organization of histone genes in *Drosophila*. RD histone genes in *Drosophila* are located on chromosome 2L and grouped in approximately 100 tandemly repeated units consisting of one copy of each core histone gene. H3/H4 and H2a/H2b share divergent promoters.

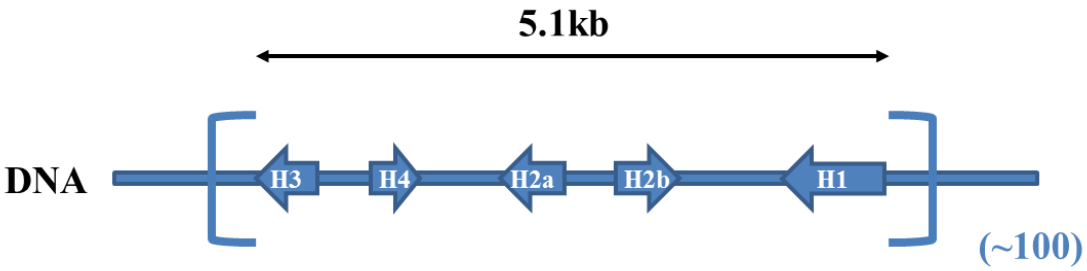
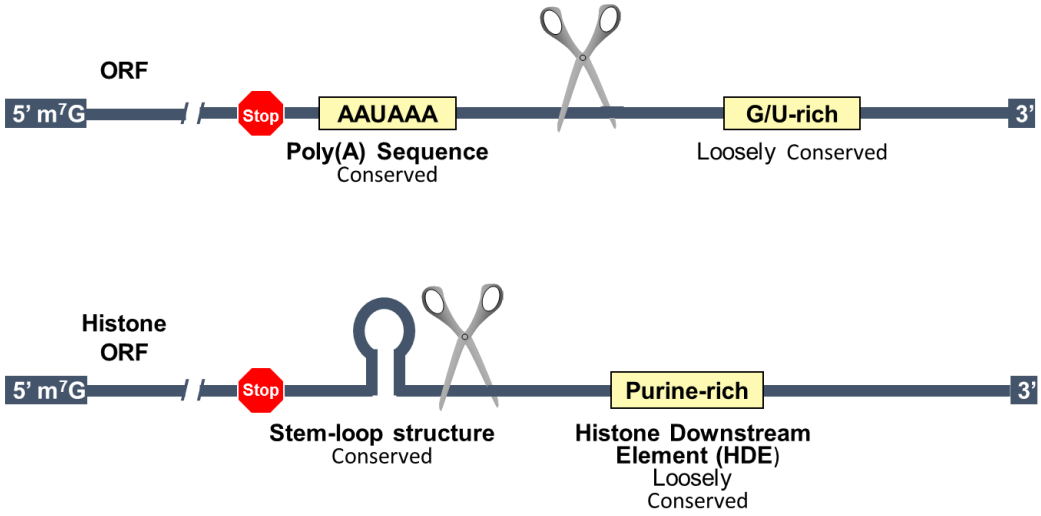

5.1kb
DNA
H3
H4
H2a
H2b
H1
(~100)

Figure 3. 3’ end processing of canonical and histone pre-mRNAs depend on different sequence elements. Although both pre-mRNAs have a 7-methylguanosine caps at the 5’ end and are cleaved during maturation, the sequence elements required for the two different processing reactions lack any similarity. Canonical pre-mRNAs have a highly conserved AAUAAA polyadenylation signal 5’ of the cleavage site and a less defined G/U rich element 3’ of the cleavage site. Histone pre-mRNAs have a highly conserved stem-loop and a purine rich HDE in the corresponding positions. Both pre-mRNAs undergo endonucleolytic cleavage between the two *cis*-acting sequence elements. Eukaryotic histone mRNAs have a unique 3’ end that lacks a poly(A) tail and instead ends in a highly conserved stem-loop structure whereas canonical pre-mRNA is appended at the 3’ end with a poly(A) tail after cleavage.

pre-mRNA



mRNA

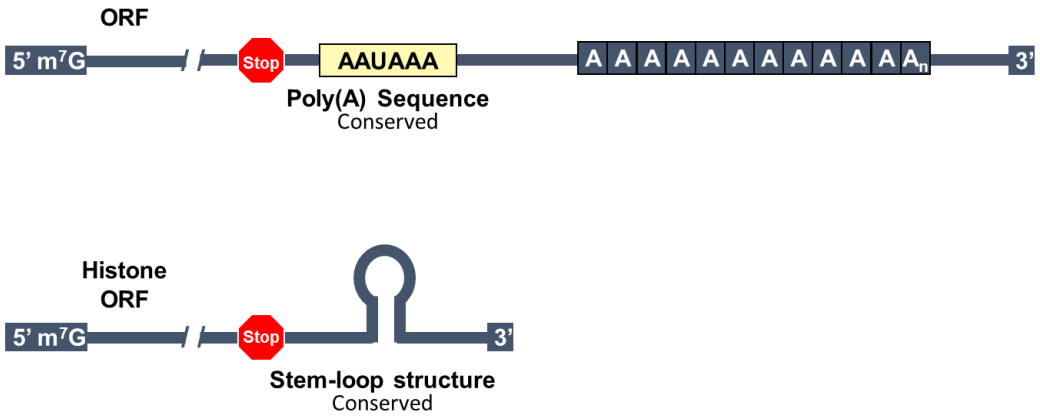


Figure 4. A large network of protein complexes positions the endonuclease in cleavage and polyadenylation. Endonucleolytic cleavage of pre-mRNA in cleavage and polyadenylation depends on recognition of key sequence elements in the 3'-UTR of pre-mRNAs. CPSF recognizes and binds the conserved poly(A) signal AAUAAA while CStF binds a G/U rich element 3' of the cleavage site. Symplekin bridges these two multi-subunit complexes. CF I_m has been functionally linked to a AUGUA motif 5' of the poly(A) signal and stimulates cleavage. Together this network of interactions juxtaposes endonuclease CPSF73 precisely at the cleavage site.

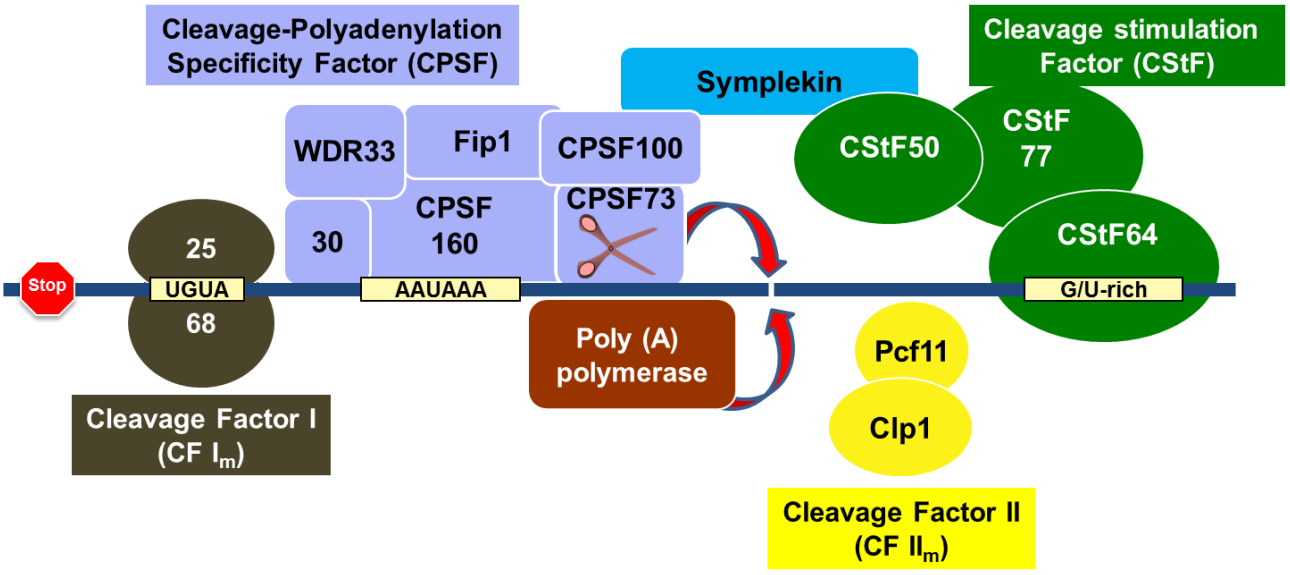
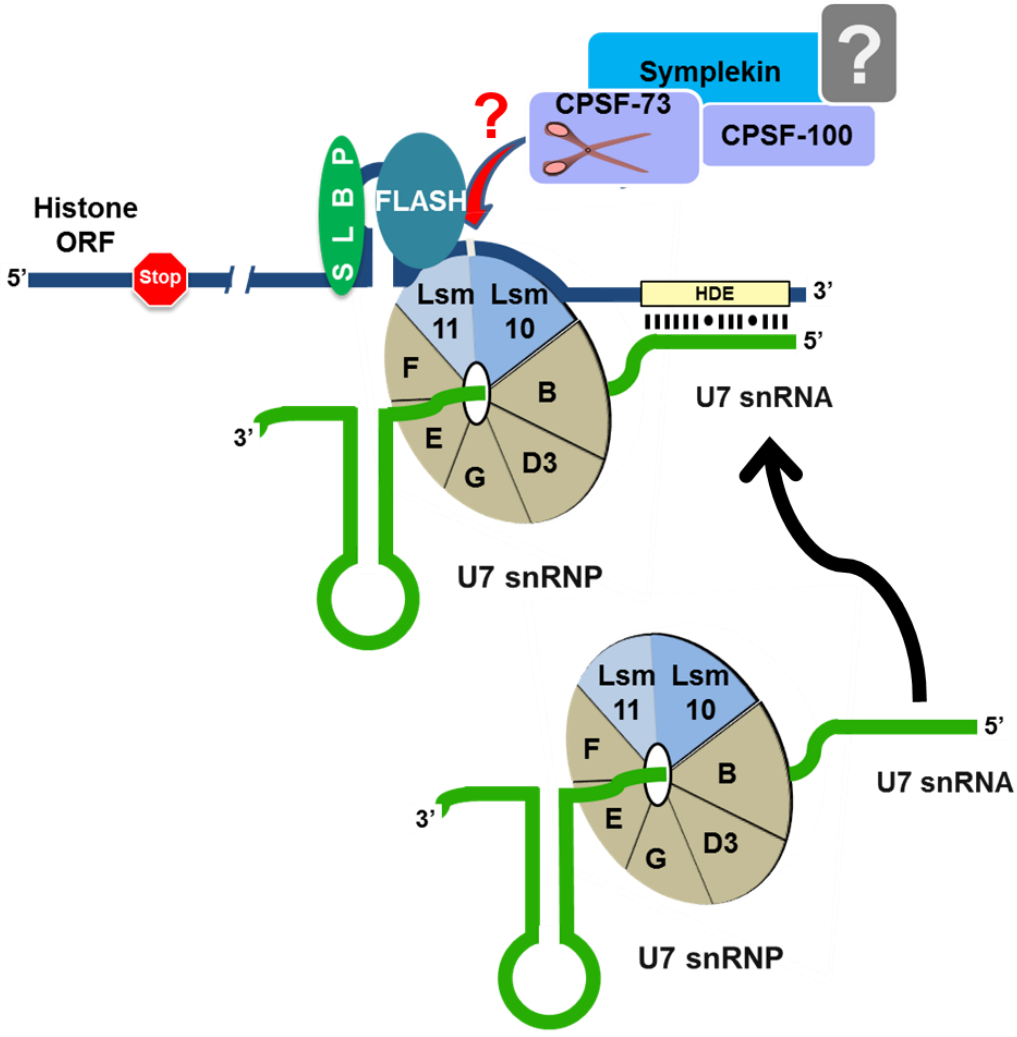

Cleavage-Polyadenylation Specificity Factor (CPSF)
Symplekin
Cleavage stimulation Factor (CStF)
WDR33
Fip1
CPSF100
CStF50
CStF 77
CPSF 160
CPSF73
30
25
CStF64
Stop
UGUA
AAUAAA
G/U-rich
68
Poly (A) polymerase
Pcf11
Clp1
Cleavage Factor I (CF I_m)
Cleavage Factor II (CF II_m)

Figure 5. Histone pre-mRNA 3' end processing depends upon a unique set of factors. The U7 snRNP is recruited to the histone pre-mRNA through partial complementarity between the HDE and 5' end of the U7 snRNA. SLBP, absolutely required for processing in *Drosophila* but less so in mammalian systems, binds the stem-loop. Lsm11, the largest member of the heptameric Lsm ring, interacts with the recently identified histone 3' end processing factor FLASH. The core cleavage complex comprised of the CPSF73, its catalytically inactive homologue CPSF100, and symplekin are all necessary for the endonucleolytic cleavage of the transcript to form mature mRNA. The accessory proteins accompanying the cleavage complex and the precise mechanism by which it is recruited remained unknown.


Symplekin
?
CPSF-73
CPSF-100
?
FLASH
S L B P
Histone
ORF
5'
Stop
Lsm 11
Lsm 10
HDE
3'
5'
F
B
E
G
D3
3'
U7 snRNA
U7 snRNP
Lsm 11
Lsm 10
5'
F
B
E
G
D3
3'
U7 snRNA
U7 snRNP

CHAPTER 2: IN MAMMALS, A SUBSET OF CLEAVAGE AND POLYADENYLATION FACTORS INCLUDING THE ENDOCUCLEASE CPSF73 ASSOCIATE WITH THE U7 SNRNP IN A FLASH DEPENDENT MANNER

INTRODUCTION

The U7 snRNP is essential for histone pre-mRNA 3' end processing in both mammalian and *Drosophila* cells. The specificity for histone pre-mRNA processing lies in the U7 snRNA and the unusual U7 snRNP heptameric Sm ring, which contains the two unique Sm-like proteins, Lsm10 and Lsm11. These replace the SmD1 and SmD2 proteins found in the canonical splicesomal Sm rings. Lsm11 has an extended N-terminus not contained by other Lsm proteins that is critical for processing (Pillai et al., 2003; Azzouz et al., 2005). With the discovery of CPSF73 as the endonuclease in histone pre-mRNA 3' end processing (Dominski et al., 2005c), and the requirement for at least two other polyadenylation factors, a major remaining question was how they are recruited to the pre-mRNA to carry out the cleavage.

A yeast two-hybrid screen using the N-terminal region of Lsm11 led to the identification of FLASH as a binding partner, and it was subsequently shown to be a critical factor for histone pre-mRNA processing just before I joined the lab. *In vitro* processing assays showed the first 139 amino acids of N-terminal FLASH were sufficient to stimulate 3' end pre-mRNA processing in HeLa and mouse myeloma nuclear extracts(Yang et al., 2009a). Although limiting in mammalian nuclear extracts, supplemented recombinant N-terminal FLASH co-purified with

biotinylated histone pre-mRNA incubated in nuclear extracts in a U7 snRNP dependent manner, indicating FLASH is associated with the histone 3' end processing machinery.

Two regions of N-terminal FLASH are essential to promote a stimulatory effect in the processing reaction: the Lsm11 binding site found through amino acids 100-139 and a highly conserved LDLY motif at amino acids 55-58. In mouse myeloma nuclear extracts, removing the Lsm11 binding site or altering key residues in this region that prevent binding eliminated the stimulatory effect of N-terminal FLASH on processing *in vitro*. A similar effect was noted when the conserved LDLY motif was mutated. Furthermore, the LDLY mutation had a mild inhibitory effect on processing when compared to control reactions (Yang et al., 2011). It was theorized this was due to the binding of endogenous Lsm11 and sequestration of the U7 snRNP in an inactive complex.

These studies led to the hypothesis that the LDLY motif was critical for the recruitment of CPSF73 to histone pre-mRNA in the context of a Lsm11/FLASH interaction. CPSF73 cleaves pre-mRNA in histone pre-mRNA 3' end processing and has been shown to stably associate with the two other polyadenylation factors essential for processing, CPSF 100 and symplekin (Dominski et al., 2005a; Kolev and Steitz, 2005). In this chapter I present data that show CPSF73, CPSF100, and symplekin associate with the Lsm11/FLASH complex, with the critical LDLY motif in FLASH fulfilling an essential role for this interaction. These results were achieved using a stable complex formed from bacterially expressed N-terminal Lsm11 and FLASH, adding it to HeLa or mouse myeloma nuclear extracts, then isolating the material in a single step biochemical purification. Surprisingly, a subset of additional polyadenylation factors associate with the Lsm11/FLASH platform including the remaining subunits of CPSF and CstF64 as the only component of CstF. This unique complex is named the histone pre-mRNA

cleavage complex (HCC) and differs from known polyadenylation machinery by the absence of CstF50, CstF77, CF I_m and CF II_m. I also map the region of Lsm11 required for binding FLASH and examine the other regions of Lsm11 required for interacting with the HCC. The HCC co-purifies with the U7 snRNP from mammalian nuclear extracts in the presence of recombinant FLASH and is recruited to histone pre-mRNA in a U7-dependent fashion under the same conditions. Together these data show that at least a portion of the available U7 snRNP exists as a stable composite structure containing FLASH and the HCC in nuclear extracts.

MATERIALS AND METHODS

3' end processing. The 86-nucleotide mouse H2a pre-mRNA substrate was generated by T7 RNA polymerase using an appropriate DNA template and was subsequently treated with calf intestinal phosphatase (NEB) to remove the 5' triphosphate. The dephosphorylated H2a pre-mRNA was labeled at the 5' end with ^{32}P using T4 polynucleotide kinase (NEB). The 3' end processing reaction was carried out in a nuclear extract prepared from mouse myeloma cells (Biovest International, Inc., MN), as previously (Dominski et al., 1995). The H2a/5m pre-mRNA, anti-U7, and anti-Mock oligonucleotides were synthesized by Dharmacon (Lafayette, CO) and had the following sequences:

H2a/5m (61-mer), 5'-biotin)(18S)(18S)CUCCCAAAAAGGCUCUUUUCAGAGCCACmCmCmAmCmUGAAUCAGAUAAAGAGCUGUGACACGGUA-3′;

anti-U7 (17-mer), 5'-mAmAmAmGmAmGmCmUmGmUmAmAmCmAmCmUmU(18S)(18S)(biotin)-3′;

anti-Mock (15-mer), 5'-
mCmGmAmGmCmUmCmGmAmUmUmCmGmCmC(18S)(18S)(biotin)-3′.

Note that in these sequences, "(18S)" represents an 18-atom spacer and "m" represents the 2′-O-methyl group.

Mutagenesis and protein expression. Mutations in the N-terminal fragments of FLASH (amino acids 29 to 139) and Lsm11 (amino acids 1 to 168) were generated using PCR and appropriately altered oligonucleotide primers. FLASH mutants were expressed in bacteria from the pET-42a vector, as described previously (Yang et al. 2009), and contained an N-terminal glutathione S-transferase (GST) tag. I expressed the N-terminal fragment of Lsm11 and its mutant versions in bacteria from the pDEST566 vector as fusions with an N-terminal maltose binding protein (MBP). Both FLASH and Lsm11 proteins additionally contained a 6×His tag that was used for purification on nickel beads (Qiagen), as described by the manufacturer.

GST pulldown assay. The wild type and various deletion mutants of the N-terminal fragment of human Lsm11 (amino acids 1 to 168) were synthesized in the presence of [^{35}S]methionine using the TnT kit (Promega), as recommended by the manufacturer. Their ability to interact with FLASH was tested using a GST-mediated pulldown assay, as described in detail previously (Yang et al., 2009a). Briefly, the ^{35}S-labeled Lsm11 proteins were mixed with the bacterially expressed GST-tagged N-terminal FLASH (amino acids 29 to 139) or with GST alone. Proteins bound to FLASH were adsorbed on glutathione (GSH) beads, separated on SDS-polyacrylamide gels, and detected by autoradiography.

Binding of nuclear proteins to the Lsm11/FLASH complex. The Lsm11/FLASH complex was formed by mixing 100 pmol of each recombinant protein in 100 μl of buffer D (100 mM KCl, 20

mM HEPES [pH 7.9], 20% glycerol, 0.2 mM EDTA [pH 8], 0.5 mM dithiothreitol [DTT]). The complex was subsequently incubated for 20 min at room temperature with 100 to 300 μl of a nuclear extract from HeLa or mouse myeloma cells. Bound proteins were purified on glutathione-Sepharose beads via the GST-tagged FLASH, electrophoretically resolved on 8 to 12% SDS–polyacrylamide gels, and detected by Western blotting and/or staining with either Coomassie blue or silver. To identify protein bands that correspond to recombinant FLASH and Lsm11, the Lsm11/FLASH complex was bound to glutathione beads in the absence of the nuclear extract and electrophoretically separated next to samples containing nuclear proteins.

Formation of processing complexes and purification of the U7 snRNP. Processing complexes were assembled in a final volume of 1 ml containing 750 μl of a highly active mouse myeloma nuclear extract, 1.25 μg of the 61-nucleotide H2a/5m pre-mRNA, and 20 mM EDTA. The samples were incubated for 5 min at 22°C followed by a 1-h rotation at 4°C. The RNA substrate and associated proteins were bound to streptavidin beads, washed several times with buffer D containing 20 mM EDTA, and separated on an SDS-polyacrylamide gel. The same method was used for the single-step purification of the endogenous U7 snRNP, with the exception that the anti-U7 2′-O-methyl oligonucleotide (1 μg) containing biotin at the 3' end was used instead of the H2a/5m pre-mRNA.

Antibodies. Antibodies against Lsm11 and FLASH were described previously (Yang et al., 2009a; Yang et al., 2009b). Antibodies against all polyadenylation factors were purchased from Bethyl Laboratories. Anti-viral protein R binding protein (anti-VprBP) and anti-damaged DNA binding protein 1 (anti-DDB1) were kindly provided by the laboratory of Y. Xiong (University of North Carolina at Chapel Hill).

RESULTS

Bacterially expressed FLASH and Lsm11 form a tight complex. Although human FLASH contains almost 2,000 amino acids, a short N-terminal region encompassing amino acids 52 to 139 is sufficient to stimulate processing of histone pre-mRNAs *in vitro* (Yang et al., 2011) (Fig. 6D, lanes 3 and 4). Amino acids 100 to 135 interact with Lsm11, whereas the highly conserved LDLY sequence located between amino acids 55 and 58 plays an essential but undetermined role in processing (Fig. 6B). Mutant proteins lacking the LDLY motif inhibit rather than stimulate processing *in vitro* by sequestering Lsm11 (and hence U7 snRNP) into an inactive processing complex (Yang et al., 2011) (Fig. 6D, lane 5). The Lsm11 binding site and the LDLY motif are functionally conserved in *Drosophila* FLASH and are essential *in vivo* (Burch et al., 2011).

Xiao Yang first tested whether bacterially expressed N-terminal regions of FLASH and Lsm11 interact with the same efficiency and specificity as previously shown for radioactively labeled proteins generated using *in vitro* transcription and translation system (TnT) (Yang et al., 2011; Burch et al., 2011). She bacterially expressed a series of GST-tagged FLASH deletions beginning at amino acid 52, 62, 88, 100, or 111 and ending at amino acid 139: FΔ51N, FΔ61N, FΔ87N, FΔ99N, and FΔ110N (Fig. 1B and andC).C). Of these FLASH variants, only FΔ28N and FΔ51N contain the essential LDLY motif (amino acids 55 to 58) and are active in processing (Fig. 6D, lanes 3 and 4) (Yang et al., 2011). I also bacterially expressed Lsm11 (amino acids 1 to 168) tagged at the N terminus with MBP. With the exception of FΔ110N, which lacks a substantial portion of the Lsm11 binding site (Fig. 6C), all of the remaining FLASH deletion mutants strongly interact with Lsm11 and the two proteins can be recovered either on glutathione beads via the GST tag attached to FLASH or on amylose beads via the MPB tag attached to Lsm11. Thus, bacterially expressed N-terminal fragments of FLASH and Lsm11 together form a

tight complex and the interaction between these two proteins involves amino acids 100 to 135 of FLASH, as previously concluded (Yang et al., 2011).

Size exclusion chromatography was used to analyze the oligomeric status of bacterially expressed N-terminal Lsm11 (amino acids 1 to 168), FΔ51N and their complex. These experiments demonstrated that Lsm11 exists in solution as a monomer, while FΔ51N alone under the same conditions forms a homo-oligomer, most likely a tetramer. The molecular mass of the FΔ51N/Lsm11 complex is consistent with four molecules of FΔ51N interacting with one molecule of Lsm11, although further detailed studies are required to confirm this stoichiometry in the complex.

The Lsm11/FLASH complex interacts with multiple polyadenylation factors. Our hypothesis was that a complex formed by FLASH and Lsm11 interacts with another processing factor and that the LDLY motif is essential for this interaction (Yang et al., 2011). To test this hypothesis, Xiao Yang mixed bacterially expressed N-terminal portions of human FLASH and Lsm11 and analyzed the ability of the Lsm11/FLASH complex to interact with CPSF73 and/or other polyadenylation factors present in nuclear extracts (Fig. 6E). Initially, we used FΔ51N (the shortest FLASH active in processing *in vitro*) either alone or together with Lsm11. The recombinant proteins were incubated either alone or with a HeLa nuclear extract for 20 min and subsequently collected on glutathione beads via the GST tag attached to FΔ51N. The two components of the FΔ51N/Lsm11 complex were recovered on glutathione beads, indicating that the presence of the nuclear extract does not interfere with formation of the complex (Fig. 7A, compare lanes 2 and 3). The same nuclear extract was also incubated with FΔ51N alone (Fig. 7A, lane 4). As visualized by silver staining of SDS gels, only a small number of HeLa nuclear proteins nonspecifically bind to glutathione beads in the absence of FΔ51N or Lsm11. One of

these HeLa proteins migrates slightly slower than FΔ51N (Fig. 7A, lane 1) and comigrates with the longer FLASH variant FΔ28N.

HeLa nuclear proteins bound to the FΔ51N/Lsm11 complex were screened by specific antibodies for the presence of CPSF73, CF I_m68, and CstF50, each representing a separate class of cleavage and polyadenylation factors. Importantly, the FΔ51N/Lsm11 complex interacted with readily detectable amounts of CPSF73 but not with CF I_m68 or CstF50 (Fig. 7B, lane 4). The precipitate also lacked COPS5, a 35 kDa component of the octameric COP9 signalosome (Wei et al., 2008), which served as a negative control for protein complexes unrelated to 3' end processing. No CPSF73 bound to FΔ51N alone, emphasizing the requirement for the Lsm11 partner (Fig. 7B, lane 3).

Xiao Yang extended this experiment by using FΔ51N and two larger FLASH deletion mutants, FΔ61N and FΔ87N (Fig. 6C). We also used more antibodies to look for the presence of other polyadenylation factors. No detectable amount of polyadenylation factors accumulated on glutathione beads in the presence of Lsm11 alone, which is tagged with MBP and does not bind to GSH beads in the absence of GST-tagged FLASH, hence serving as a negative control to measure the nonspecific background of nuclear proteins (Fig. 7C, bottom panel, lane 2). Among HeLa proteins associated with the complex composed of Lsm11 and the processing-proficient FΔ51N FLASH, we identified CPSF73 and all of the remaining CPSF subunits larger than 50 kDa, including CPSF160, CPSF100, Fip1, and WDR33, although the latter subunit was only weakly detected by Western blotting (Fig. 7C, lane 3). CPSF30, which is the smallest subunit of CPSF, was detected in independent experiments after resolving proteins in a higher-percentage polyacrylamide gel. The bound proteins also included symplekin and CstF64 (Fig. 7C, lane 3). Importantly, the binding of all these proteins was abolished when the complex contained the

processing-deficient FLASH mutant FΔ61N or FΔ87N (Fig. 7C, lanes 4 and 5). This result clearly demonstrates that the strong binding of CPSF73 and other polyadenylation factors to the complex is not mediated solely by Lsm11, as the same amounts of Lsm11 were precipitated by all three FLASH deletion mutants (Fig. 7C, bottom panel, lanes 3 to 5).

The material bound to the FΔ51N/Lsm11 complex lacked CF I_m68 and CstF50 and contained only weakly detectable amounts of CstF77 that were also present when the first 87 amino acids were deleted from FLASH (Fig. 7C, lanes 3 to 5). Subsequent studies demonstrated that CstF77 nonspecifically interacts with Lsm11 and also has a weak affinity for streptavidin beads (described below and see Fig. 11A). The complex bound to the FΔ51N/Lsm11 complex also did not contain any detectable amounts of CstF64 Tau. This paralogue of CstF64 associates with canonical pre-mRNAs (Shi et al., 2009) and can partially substitute for CstF64 in cleavage and polyadenylation and in 3' end processing of histone pre-mRNAs (Ruepp et al., 2011).

The FΔ51N/Lsm11 complex bound the same set of proteins in a nuclear extract prepared from mouse myeloma cells (Fig. 7D, lane 3). Among the mouse proteins bound to the complex, WDR33 was readily detectable, as were the two forms of mouse symplekin. Again, with the exception of CstF77, the binding of all of the proteins was greatly reduced when the processing-deficient FΔ87N FLASH mutant was used instead of FΔ51N (Fig. 7D, lane 4). Altogether, the results demonstrate that in mammalian nuclear extracts the FΔ51N/Lsm11 complex interacts with a specific combination of cleavage and polyadenylation factors: symplekin, all CPSF subunits, and CstF64 as the only CStF component.

The LDLY sequence in FLASH is essential for the efficient recruitment of polyadenylation factors. The LDLY motif is absolutely essential for the activity of FLASH in processing (Yang et al., 2011), and the above results suggest that this motif might also be critical for binding the

multitude of polyadenylation factors to the Lsm11/FLASH complex. To test this assumption, we used the processing-deficient LDLY-4A mutant in which the LDLY sequence was replaced with 4 alanines (Fig. 6C). The mutation was made in the context of the longer FLASH FΔ28N (amino acids 29 to 139) fused N terminally to GST. No detectable amounts of polyadenylation factors were bound to glutathione beads in the absence of any recombinant protein or in the presence of either Lsm11 (amino acids 1 to 168) or FΔ28N alone. The complex of FΔ28N and Lsm11 efficiently interacted with all the polyadenylation factors previously identified for the Lsm11/FΔ51N complex: symplekin, CPSF160, CPSF100, CPSF73, and CPSF30 (Fig. 8A and and B). Most strikingly, the complex of Lsm11 and LDLY-4A did not interact with any of these proteins (Fig. 8A and B). We conclude that the essential function of the LDLY motif in processing is to recruit a subset of polyadenylation factors to histone pre-mRNA.

To confirm the identity of polyadenylation factors associated with FLASH and Lsm11 and to potentially identify other interacting proteins, we used mass spectrometry. We assembled the complex of Lsm11 (amino acids 1 to 168) and either the FΔ28N FLASH (amino acids 29 to 139) or its LDLY-4A mutant and incubated each complex with a HeLa nuclear extract. The wild-type complex interacted with a number of proteins readily detectable in a polyacrylamide gel by silver staining as individual bands (Fig. 8C, lane 2). These bands and the remaining parts of the same lane and the corresponding regions in the lane containing the mutant complex (Fig. 8C, lane 3) were excised and analyzed by mass spectrometry. This analysis fully confirmed the results obtained by Western blotting, with the slowest-migrating silver-stained band in the lane containing the wild-type complex (Fig. 8C, lane 2) yielding a large number of peptides for CPSF160 and WDR33 (band 1). The faster-migrating bands contained symplekin, CPSF100, CPSF73, Fip1, and CstF64 (bands 2 to 7). All of these proteins were virtually undetectable in the

precipitate bound to the complex of Lsm11 and LDLY-4A (Fig. 8D). Consistent with the results of Western blotting, proteomic analysis of the material bound to the wild-type complex did not detect any peptides from CstF50 and CF I_m68 and only one peptide from CstF77. In addition, no traces of other polyadenylation factors, including the Pcf11 subunit of CF II_m, were detected. Finally, the complex of the FΔ28N FLASH and Lsm11 does not bind any of previously identified proteins specifically involved in 3' end processing of histone pre-mRNAs, including ZFP100 and SLBP.

Both the wild-type and mutant Lsm11/FLASH complexes bound readily detectable amounts of DDB1 (damaged DNA binding protein 1) and its binding partner VprBP (viral protein R binding protein) (McCall et al., 2008) that comigrates with CPSF160 and WDR33 (Fig. 8C, lanes 2 and 3). Subsequent studies revealed that VprBP and DDB1 bind directly to Lsm11, and this interaction is not linked to 3' end processing of histone pre-mRNA *in vitro*.

Independent proteomic analyses were conducted using other HeLa nuclear extract preparations and a nuclear extract prepared from mouse myeloma cells. The same polyadenylation factors invariably bound to the wild-type Lsm11/FLASH complex, regardless of whether the extract was very active or poorly active in processing and whether the EDTA was omitted or included at the final concentration of 20 mM, as used in the *in vitro* processing reaction. The unique combination of polyadenylation factors bound to the complex of FLASH and Lsm11 was also not affected by pretreatment of the nuclear extract with RNase A.

Regions of Lsm11 required for recruiting polyadenylation factors. Our previous analysis indicated that deletion of the first 40 amino acids of Lsm11 (LΔ40N) almost completely eliminated its interaction with FLASH (Yang et al., 2009a). I used the GST-binding assay and various ^{35}S-labeled deletion mutants of Lsm11 expressed using the TnT system (Fig. 9A and B)

to define the minimal region of Lsm11 required to form a complex with FLASH. A short fragment encompassing the first 40 amino acids of Lsm11 (L40N) is not sufficient to interact with FΔ28N fused to GST (Fig. 9C, top panel, lane 3). However, extending this region to 65 amino acids resulted in a minimal Lsm11 (L65N) that interacted with FΔ28N as strongly as the entire N-terminal Lsm11 encompassing amino acids 1 to 168 (Fig. 9C, top panels, lanes 6 and 9). Further mapping of Lsm11 using ^{35}S-labeled fragments and GST-tagged FΔ28N narrowed the region of interaction to between amino acids 19-55 of Lsm11 (L19-55N in Fig. 10 lane 3).

I bacterially expressed some of the fragments of Lsm11 as fusions with MBP and tested their ability to interact with FΔ28N fused to GST and bind polyadenylation factors. As expected, only residual amounts of LΔ40N and polyadenylation factors were collected on glutathione beads (Fig. 9D, both panels, lane 4), consistent with the virtual inability of this Lsm11 mutant to form a complex with FLASH. Importantly, FΔ28N and the L65N mutant, while forming a very tight complex, were even less efficient in binding polyadenylation factors (Fig. 9D, both panels, lane 5). When the length of Lsm11 was increased from 65 to 130 N-terminal amino acids, the corresponding complex was nearly as efficient in binding symplekin, CstF64, and CPSF subunits as the wild-type complex containing the entire 168-amino-acid N-terminal Lsm11 (Fig. 9D, both panels, compare lanes 3 and 6). Shortening Lsm11 to the first 105 amino acids (L105N in Fig. 9B) resulted in a significant loss of binding activity by the Lsm11/FLASH complex (Fig. 9E, both panels, lanes 3 and 4). Thus, a complex containing the first 130 amino acids of Lsm11 is sufficient for binding the polyadenylation factors.

Polyadenylation factors are present on the endogenous U7 snRNP. We demonstrated that large amounts of the recombinant Lsm11/FLASH complex bind a unique combination of polyadenylation factors in mammalian nuclear extracts. To determine whether FLASH associates

with endogenous Lsm11 in the context of the U7 snRNP, I purified the U7 snRNP on streptavidin beads using a 2′-O-methyl oligonucleotide that contains a biotin tag at the 3’ end and a sequence complementary to the first 15 nucleotides of the U7 snRNA (anti-U7). I also used a 3′-biotinylated 2′-O-methyl oligonucleotide with an unrelated sequence that served as a negative control (anti-Mock). The experiment was carried out with a nuclear extract from mouse myeloma cells, which contains about a 10-fold higher concentration of the U7 snRNP than extracts from HeLa cells (unpublished observations).

As judged by the high enrichment of Lsm11 in the material bound to the anti-U7 oligonucleotide, this antisense oligonucleotide is very efficient in purifying the U7 snRNP (Fig. 11A, lane 2). The anti-Mock oligonucleotide did not result in any background of Lsm11, indicating that it has no affinity for the U7 snRNA (Fig. 11A, lane 3). Most importantly, readily detectable amounts of polyadenylation factors, including several CPSF subunits and CstF64, were present in the anti-U7 precipitate but not in the anti-Mock precipitate. Compared with the input, the amounts of these proteins were small and this reflects the extremely low concentration of endogenous U7 snRNP and FLASH and the large abundance of the polyadenylation factors in nuclear extracts. A specific antibody against the N-terminal region of FLASH encompassing amino acids 1 to 139 detected a band migrating at about 35 kDa in the anti-U7 precipitate that was not present in the control precipitate (Fig. 11A, lanes 2 and 3). This band most likely represents a FLASH degradation product encompassing the first ~300 amino acids of the protein, i.e., the region that is sufficient for the interaction with Lsm11 and which contains the region used to generate the anti-FLASH antibody (Yang et al., 2009a).

Only background amounts of CstF77 were detected in both the anti-U7 and anti-Mock precipitates, indicating that CstF77 is not part of a complex with CstF64 in the U7 snRNP (Fig.

11A, lanes 2 and 3). In agreement with the data obtained for the recombinant Lsm11/FLASH complex, the U7 snRNP-bound proteins also did not contain detectable amounts of CstF50 or CF I_m68. We conclude that in mammalian cells at least a fraction of the U7 snRNP exists as a preassembled complex containing the same set of polyadenylation factors that interact with the recombinant Lsm11/FLASH complex.

Addition of recombinant FΔ28N but not the LDLY-4A mutant increased the amount of polyadenylation complex bound to the U7 snRNP but had no effect on the amount of CstF77 detected (Fig. 11B, lanes 2 and 3). Note that lane 3 compared to lane 2 was loaded with a smaller amount of U7 snRNP, as indicated by a weaker signal for Lsm11, yet symplekin, CPSF160, CPSF73, CPSF30, and CstF64 were clearly more abundant. These results confirm that a fraction of the endogenous U7 snRNP is associated with a subset of polyadenylation factors and support the notion that these factors are recruited to the U7 snRNP by FLASH interacting with Lsm11.

Composition of the processing complex assembled on histone pre-mRNA. We next investigated whether the same cleavage and polyadenylation factors, including the endonuclease CPSF73, can be detected *in vitro* in a complex assembled on histone pre-mRNA. An obstacle in achieving this goal is that the cleavage reaction proceeds very rapidly in nuclear extracts incubated at room temperature, causing disruption of the processing complex. Previously, we used low concentrations of NP-40 (0.05 to 0.1%) to block the cleavage reaction and to arrest processing complexes during the assembly process (Yang et al., 2009a). This strategy allowed us to capture SLBP and the U7 snRNP in association with histone pre-mRNA, but we failed to detect any polyadenylation factors. One possibility was that while NP-40 blocks processing, it does so by destabilizing the interaction of polyadenylation factors with FLASH bound to the U7 snRNP.

To avoid this potential limitation, Dr. Dominski designed a new histone pre-mRNA substrate termed H2a/5m in which 5 nucleotides surrounding the major and three minor cleavage sites were modified with a 2′-O-methyl group (Fig. 11C). Our previous studies demonstrated that 2′-O-methyl nucleotides are resistant to hydrolysis by CPSF73 when it acts as a 5' exonuclease (Yang et al., 2009). The H2a/5m substrate also contained two point mutations within the HDE that improved its interaction with the U7 snRNA and biotin at the 5' end for subsequent purification of bound nuclear components on streptavidin beads. When incubated in a mouse nuclear extract, the H2a/5m pre-mRNA assembled into a stable complex containing highly enriched amounts of SLBP and, as judged by the presence of Lsm11, the U7 snRNP. This result indicates that the cleavage reaction was at least partially blocked (Fig. 11D, lane 2).

As determined using specific antibodies, the complex assembled on the H2a/5m pre-mRNA contained the same polyadenylation factors that associate with the recombinant Lsm11/FLASH complex or the U7 snRNP, including symplekin, CPSF subunits, and CstF64 but lacking CstF50 and CF I_m68 (Fig. 11D, lane 2). Addition of recombinant FLASH (amino acids 29 to 139) only slightly increased the amount of these factors in the processing complex, although the recombinant FLASH was efficiently delivered to the H2a/5m pre-mRNA through the interaction with Lsm11. Importantly, blocking the U7 snRNA by a 2′-O-methyl oligonucleotide identical to the anti-U7 but lacking the biotin tag prevented the association of the U7 snRNP, recombinant FΔ28N, and all of the polyadenylation factors with the pre-mRNA substrate without affecting the binding of SLBP (Fig. 11D, lane 4). Again, only very small amounts of CstF77 were detected in both the presence and the absence of the anti-U7 oligonucleotide (Fig. 11D, compare lanes 2 to 4), indicating that this protein nonspecifically associates with the substrate and/or streptavidin beads and is not part of the complex with other

polyadenylation factors. We conclude that a preassembled U7 snRNP containing the CPSF73 3' endonuclease and multiple polyadenylation factors is delivered to histone pre-mRNA through the interaction between the HDE and the U7 snRNA for 3' end processing (Fig. 11E).

DISCUSSION

Cleavage of animal replication-dependent histone pre-mRNAs critically depends on the U7 snRNP that interacts with the histone downstream element (HDE) located several nucleotides 3′ of the cleavage site. Histone pre-mRNAs do not share any sequence elements with canonical pre-mRNAs, and their cleavage is not followed by the polyadenylation step. Previous studies showed that 3' end processing of histone pre-mRNAs requires at least three components of the cleavage/polyadenylation machinery: the endonuclease CPSF73, CPSF100, and symplekin (Dominski et al., 2005c; Kolev and Steitz 2006; Wagner et al., 2007). For canonical pre-mRNAs, the CPSF73 endonuclease is brought to the vicinity of the cleavage site primarily by CPSF160, which recognizes the upstream AAUAAA element (Mandel et al., 2008). However, it has not been determined how CPSF73 and the two other common subunits are recruited to the processing machinery assembled on histone pre-mRNAs.

A subset of polyadenylation factors interacts with the Lsm11/FLASH complex. Of the known components of the U7 snRNP-based processing machinery, Lsm11 was the most likely candidate for a protein that either directly or indirectly recruits CPSF73 to histone pre-mRNA. First, Lsm11 is the largest component of the U7-specific Sm ring and contains an extended N-terminal region of about 170 amino acids that is essential for processing (Azzouz et al., 2005). Furthermore, this extended region interacts with FLASH (Yang et al., 2009a), a recently

identified factor required for endonucleolytic cleavage of histone pre-mRNAs by CPSF73 (Yang et al., 2011). Finally, the base pair interaction between the HDE and the 5' end of the U7 snRNA brings Lsm11 near the cleavage site (Yang et al., 2009b).

In this study, Xiao Yang bacterially expressed the N-terminal fragments of FLASH and Lsm11 and showed that these two recombinant proteins interact with each other. Our initial gel filtration experiments demonstrated that the Lsm11/FLASH complex is significantly larger than a simple heterodimer of the two proteins and suggest that it may consist of four molecules of FLASH and one molecule of Lsm11. This conclusion is consistent with the mechanism of 3' end processing in which a single particle of the U7 snRNP interacts with histone pre-mRNA and with the observation that FLASH can self-associate through the N-terminal region (Kiriyama et al., 2009).

The complex of recombinant FLASH and Lsm11 tightly interacts with a unique combination of polyadenylation factors present in mammalian nuclear extracts, whereas FLASH and Lsm11 individually are unable to bind the same factors with efficiency. Among nuclear proteins associated with the Lsm11/FLASH complex were symplekin, CstF64, and all subunits of CPSF, including the recently discovered component WDR33 (Shi et al., 2009) and the endonuclease CPSF73. This is the first biochemical demonstration of an interaction between proteins specifically devoted to processing of histone pre-mRNAs (FLASH and Lsm11) with subunits of the cleavage and polyadenylation machinery. The two other subunits of CStF, CstF77 and CstF50, as well as the cleavage factors I and II, were absent, indicating that FLASH and Lsm11 interact with a specific subset of polyadenylation factors.

The interaction of the Lsm11/FLASH complex with CPSF subunits, CstF64, and symplekin was abolished by mutating the LDLY sequence in FLASH, a highly conserved cluster

of amino acids found in FLASH orthologues of both vertebrates and invertebrates. Thus, all of these polyadenylation factors likely exist in the cell as a single complex that is specifically crafted to directly interact with the Lsm11/FLASH complex. Due to the presence of the CPSF73 endonuclease, we refer to this complex as the histone pre-mRNA cleavage complex (HCC). The LDLY motif is absolutely essential for the activity of FLASH in processing of histone pre-mRNAs in mammalian nuclear extracts (Yang et al., 2011), and deletion of the corresponding LDIY sequence in *Drosophila* FLASH disrupts 3' end processing in cultured S2 cells (Burch et al., 2011). These results demonstrate that the role of this motif in processing is to recruit the CPSF73 endonuclease and other components of the HCC to the Lsm11/FLASH complex.

The details of the interaction between the Lsm11/FLASH complex and the HCC are not known. In human Lsm11, I showed amino acids 19 to 55 interact with FLASH and an additional region(s) located between amino acids 66 and 130 is required for binding the HCC by the Lsm11/FLASH complex. In the simplest model, the LDLY motif in FLASH and amino acids 66 to 130 in Lsm11 have individually a weak affinity for one or more components of the HCC. When these two sequence elements are juxtaposed together in the Lsm11/FLASH complex, they function in a cooperative manner, resulting in a very tight binding of the HCC. An important role in binding the HCC may also be played by the interface of the interacting Lsm11 and FLASH and/or potential structural rearrangements in each protein triggered by the formation of the Lsm11 and FLASH complex.

Intriguingly, the HCC is very similar in its composition to the heat labile factor (HLF) that was identified by Kolev and Steitz in HeLa cell nuclear fractions capable of restoring the ability of heat-treated nuclear extracts to process histone pre-mRNAs (Kolev and Steitz, 2005). The HLF contains symplekin as the most heat-sensitive component and cofractionated with all of

the known CPSF subunits and CstF64. Similar to the complex recruited to histone pre-mRNA, the HLF lacks CstF50 and subunits of CF Im and CF IIm but was reported to contain CstF77, which we detect as a nonspecific contaminant. An attractive possibility is that the HCC and the HLF are identical and CstF77 is not a genuine part of the HLF but simply copurifies as a component of a different complex sharing similar chromatographic properties to the HCC. Taken together, our findings and the previous results of Kolev and Steitz demonstrate that the functional overlap between the canonical cleavage/polyadenylation and 3' end processing of histone pre-mRNAs is much more extensive than previously anticipated, strongly suggesting that these two 3' end processing reactions are evolutionarily related (Gilmartin, 2005).

U7 snRNP as a preformed 3' end processing unit. The HCC is also associated with the endogenous U7 snRNP in the absence of a pre-mRNA substrate. Based on our experiments with recombinant Lsm11 and FLASH, the assembly of this "active" U7 snRNP is dependent on FLASH, which is limiting in many nuclear extracts. Indeed, addition of recombinant wild-type FLASH, but not the LDLY-4A mutant, to a nuclear extract stimulated the association of the polyadenylation complex with the U7 snRNP. Overall, our results reveal an unexpected complexity of the U7 snRNP, previously believed to consist solely of the U7 snRNA and the Sm ring.

Previous studies of U7 snRNP purified by several chromatographic steps did not reveal novel components of this snRNP other than Lsm10 and Lsm11, suggesting that the entire HCC dissociated from U7 snRNP during purification, perhaps as a result of FLASH proteolysis (Pillai et al., 2001a; Smith et al., 1991). More recent proteomic studies identified CF I_m68 as the only polyadenylation factor that associates with the U7 snRNP (Ruepp et al., 2010). We found no evidence that this protein is recruited by the Lsm11/FLASH complex or is part of the

endogenous U7 snRNP purified by the anti-U7 oligonucleotide. Further studies are required to determine the source of this contradiction.

I also isolated a processing complex assembled on a histone pre-mRNA that was resistant to cleavage as a result of the placement of five 2′-O-methyl nucleotides around the cleavage site. This processing complex contained SLBP, U7 snRNP, and the same combination of polyadenylation factors that bind the Lsm11/FLASH complex, suggesting that the "active" form of U7 snRNP carrying the HCC is directly delivered to histone pre-mRNA through a single-step base-pairing interaction between the HDE and U7 snRNA. Addition of recombinant FLASH only slightly stimulated recruitment of the HCC to the modified histone pre-mRNA, suggesting that the interaction between the HCC and the U7 snRNP in the context of the entire processing machinery assembled on histone pre-mRNA is intrinsically less stable and/or the modified substrate was undergoing limited cleavage in neighboring sites that lack the 2′-O-methyl modification.

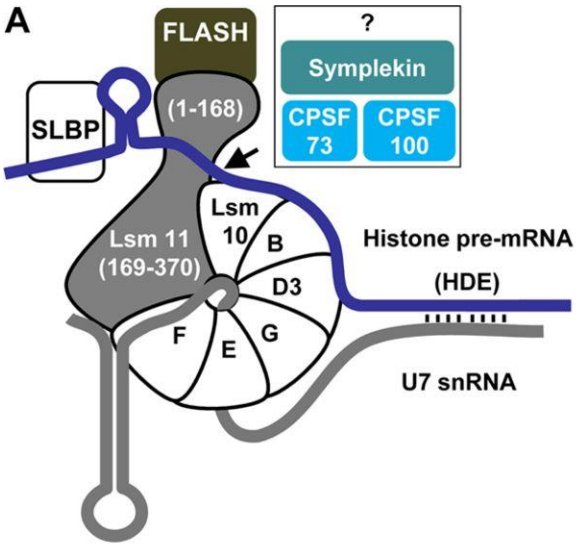
Potential roles of components of the HCC in 3' end processing of histone pre-mRNAs. CstF64 was previously shown to interact through overlapping regions with either CstF77 or symplekin, but it is unable to interact with these two proteins simultaneously (Takagaki and Manley, 2000; Ruepp et al., 2011). The presence of CstF64 and symplekin in the complex bound to the Lsm11/FLASH complex likely explains the concomitant lack of CstF77 and CstF50, as CstF50 interacts with CstF77 (Takagaki and Manley, 2000; (Bai et al., 2007). Thus, our data suggest that the previously puzzling mutually exclusive interaction of CstF64 with either symplekin or CstF77 allows formation of two alternative complexes in animal cells: one containing all three CStF subunits that functions in cleavage and polyadenylation and the other composed of CstF64 and symplekin that operates on histone pre-mRNAs (Fig. 11E). This

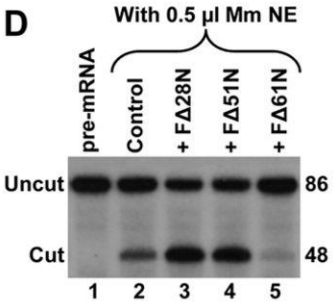
interpretation is strongly supported by the recent in vivo studies of Schumperli and coworkers who in an elegant set of experiments addressed the role of the CstF64-symplekin interaction in 3' end processing of canonical and histone pre-mRNAs. These authors designed a symplekin mutant that does not bind CstF64 and tested its ability to substitute for the endogenous symplekin depleted from HeLa cells by RNA interference (RNAi) (Ruepp et al., 2011). Importantly, this mutant protein could efficiently rescue cleavage/polyadenylation, but formation of mature histone mRNAs remained impaired, demonstrating that in vivo the symplekin-CstF64 interaction is required for 3' end processing of histone pre-mRNA but not canonical pre-mRNAs. Intriguingly, CstF64 is the only known component of the cleavage and polyadenylation machinery that is upregulated several fold as cells transit from G0 to S phase (Martincic et al., 1998). In addition, depletion of CstF64 causes cell-cycle arrest and apoptosis-like cell death (Takagaki and Manley, 1998), and this phenotype could be directly related to its role in generating mature histone mRNAs. Overall, our results suggest that the biochemical function of the CstF64/symplekin heterodimer is to interact with FLASH and Lsm11 and to bring other polyadenylation factors, including CPSF73 and CPSF100, to the U7 snRNP. This "active" U7 snRNP is ultimately recruited to histone pre-mRNA for 3' end processing.

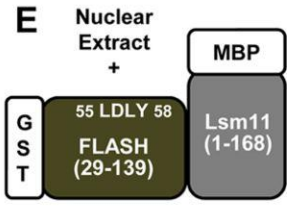
With the exception of CPSF73, which contacts the cleavage site in histone pre-mRNA and functions as the 3′ endonuclease (Dominski et al., 2005c; Dominski, 2010), the role of the remaining CPSF subunits in the U7-dependent processing is largely unknown. CPSF100 is a

noncatalytic homologue, binding partner, and perhaps a regulator of CPSF73 (Kolev et al., 2008; Dominski). CPSF160, CPSF30, Fip1, and WDR33 may be essential for maintaining the integrity of the polyadenylation complex or play regulatory roles in vivo by connecting 3' end processing of histone pre-mRNAs with other cellular processes. However, it is also possible that at least some of these subunits are simply passive bystanders that play no role in the U7-dependent processing.

Figure 6. Biochemical assay to identify proteins interacting with a complex of FLASH and Lsm11. (A) The known critical factors in 3' end processing of histone pre-mRNA (blue line) are shown. Vertical lines indicate base paring between the histone downstream element (HDE) and the 5' end of U7 snRNA (gray line). Histone pre-mRNAs are cleaved by CPSF73 typically five nucleotides after the stem-loop (arrow). The reaction also requires CPSF100, symplekin, and potentially other factors (question mark). (B) Amino acid sequence of the N-terminal region of human FLASH (amino acids 1 to 139). The first amino acids of various N-terminal truncations are indicated with arrowheads. The Lsm11-binding site is indicated with a double-headed arrow, and the LDLY motif is underlined. (C) Diagram of FLASH mutants used in this study. The GST tag is fused N terminally to each protein and is not indicated. The ability of each FLASH protein to support processing or bind Lsm11 is indicated. (D) *In vitro* processing of histone pre-mRNA with a limiting amount of mouse nuclear extract (NE) in the absence of any recombinant protein (lane 2) or in the presence of 100 ng of various FLASH proteins, as indicated (lanes 3 to 5). Lane 1 contains the input substrate. (E) Schematic of the assay for isolation of factors that bind to the Lsm11/FLASH complex. The N-terminal FLASH (amino acids 29 to 139) fused to GST and the N-terminal Lsm11 (amino acids 1 to 168) fused to MBP were incubated with a nuclear extract, and interacting proteins were purified on glutathione beads.







C

		55 58	100	135		Lsm11 binding	3' end cleavage
FΔ28N	29	LDLY	Lsm11 binding site		139	Yes	Yes
FΔ51N	52	LDLY	Lsm11 binding site		139	Yes	Yes
FΔ61N	62		Lsm11 binding site		139	Yes	No
FΔ87N	88		Lsm11 binding site		139	Yes	No
FΔ99N	100		Lsm11 binding site		139	Yes	No
FΔ110N	111				139	No	No
LDLY-4A	29	AAAA	Lsm11 binding site		139	Yes	No

Figure 7. The Lsm11/FLASH complex binds a subset of factors involved in cleavage and polyadenylation. (A) GST-mediated binding of the FΔ51N/Lsm11 complex to glutathione (GSH) beads in the absence (lane 2) or in the presence (lane 3) of a HeLa nuclear extract (NE). Binding of FΔ51N alone in the presence of the same nuclear extract is shown in lane 4. Proteins bound to GSH beads were resolved in an 8% SDS–polyacrylamide gel and stained with silver. Proteins in the nuclear extract bound to GSH beads in the absence of recombinant proteins are shown in lane 1 (indicated with asterisks). (B) FΔ51N (lane 3), alone or together with Lsm11 (lane 4), was incubated with a HeLa nuclear extract, and the material bound to GSH beads was tested for the presence of CPSF73, CF I_m68, CstF50, and COPS5 using specific antibodies. Lane 2 contains proteins bound to GSH beads in the absence of recombinant proteins, and HeLa NE input (20%) is shown in lane 1. (C and D) Proteins of a HeLa nuclear extract (C) or mouse myeloma (Mm) nuclear extract (D) bound to GSH beads in the presence of indicated recombinant proteins were analyzed by Western blotting using specific antibodies. Coomassie blue staining of the two recombinant proteins bound to GSH beads is shown in the bottom panels. Lane 2 in each panel shows proteins collected on GSH beads in the presence of Lsm11, which is tagged with MBP and does not bind to GSH beads, hence serving as a negative control to measure the nonspecific background (see the bottom panels).

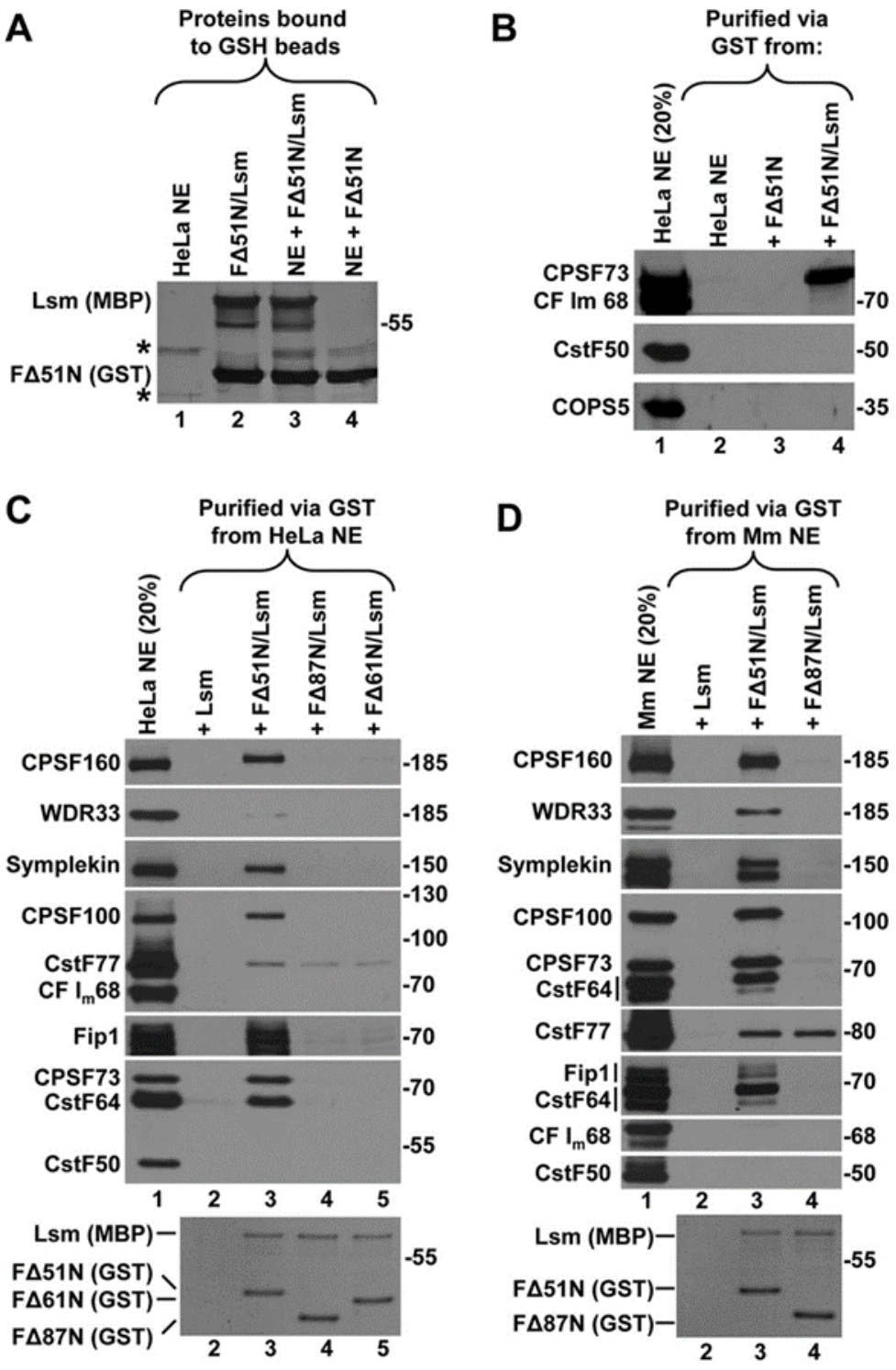
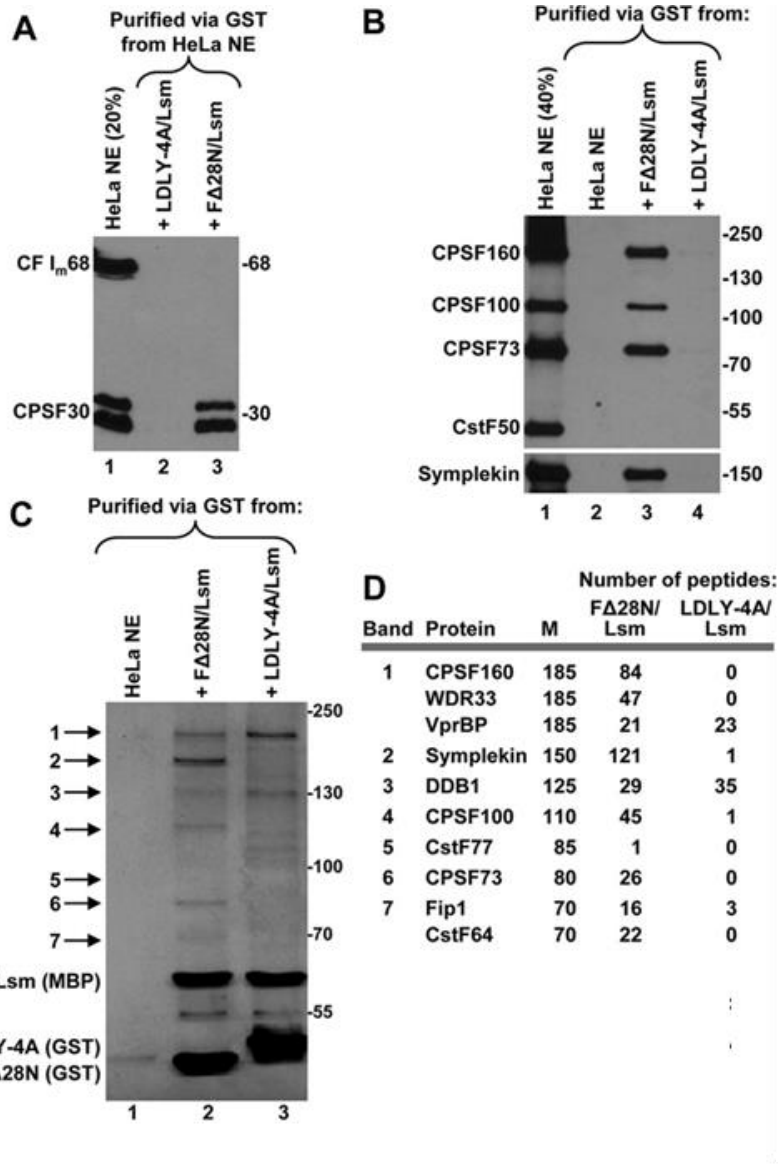

A
Proteins bound to GSH beads
HeLa NE
FΔ51N/Lsm
NE + FΔ51N/Lsm
NE + FΔ51N
Lsm (MBP)
FΔ51N (GST)
-55
1 2 3 4
B
Purified via GST from:
HeLa NE (20%)
HeLa NE
+ FΔ51N
+ FΔ51N/Lsm
CPSF73
CF Im 68
CstF50
COPS5
-70
-50
-35
1 2 3 4
C
Purified via GST from HeLa NE
HeLa NE (20%)
+ Lsm
+ FΔ51N/Lsm
+ FΔ87N/Lsm
+ FΔ61N/Lsm
CPSF160
WDR33
Symplekin
CPSF100
CstF77
CF Im68
Fip1
CPSF73
CstF64
CstF50
-185
-185
-150
-130
-100
-70
-70
-70
-55
1 2 3 4 5
Lsm (MBP)
FΔ51N (GST)
FΔ61N (GST)
FΔ87N (GST)
-55
2 3 4 5
D
Purified via GST from Mm NE
Mm NE (20%)
+ Lsm
+ FΔ51N/Lsm
+ FΔ87N/Lsm
CPSF160
WDR33
Symplekin
CPSF100
CPSF73
CstF64
CstF77
Fip1
CstF64
CF Im68
CstF50
-185
-185
-150
-100
-70
-80
-70
-68
-50
1 2 3 4
Lsm (MBP)
FΔ51N (GST)
FΔ87N (GST)
-55
2 3 4

Figure 8. The LDLY motif in FLASH is required for the interaction of the complex with polyadenylation factors. (A and B) HeLa nuclear proteins bound to GSH beads in the presence of the indicated Lsm11/FLASH complexes were analyzed using specific antibodies. Lane 2 in panel B contains HeLa nuclear proteins bound to GSH beads in the absence of recombinant proteins. (C and D) The identity of HeLa nuclear proteins bound to GSH beads in the presence of the indicated complexes (lanes 2 and 3) was determined by mass spectrometry. HeLa proteins were electrophoretically separated in a 10% SDS–polyacrylamide gel and stained with silver (C). Note that the LDLY-4A mutant migrates slower than the wild-type FΔ28N. Lanes 2 and 3 were divided into 11 equal sections and treated with trypsin, and the complete proteome of each gel section was determined by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The most abundant proteins identified by mass spectrometry were assigned to individual silver-stained bands (indicated with arrows) and are listed in panel D. Lane 1 in panel C contains HeLa nuclear proteins bound to GSH beads in the absence of recombinant proteins, and it was not analyzed by mass spectrometry. Column M in panel D shows molecular masses in kDa.



D

Band	Protein	M	Number of peptides: FΔ28N/ Lsm	LDLY-4A/ Lsm
1	CPSF160	185	84	0
	WDR33	185	47	0
	VprBP	185	21	23
2	Symplekin	150	121	1
3	DDB1	125	29	35
4	CPSF100	110	45	1
5	CstF77	85	1	0
6	CPSF73	80	26	0
7	Fip1	70	16	3
	CstF64	70	22	0

Figure 9. Sequences in Lsm11 required for binding the polyadenylation complex by the Lsm11/FLASH complex. (A) Amino acid sequence of the N-terminal region of human Lsm11 (amino acids 1 to 168). The first or last amino acids of various truncations are indicated with arrowheads, and the FLASH-interacting region is underlined. (B) Diagrams of Lsm11 truncations used in the study. The MBP tag is fused N terminally to Lsm11 and is not indicated. (C) Binding of ^{35}S-labeled Lsm11 truncations by either GST or FΔ28N tagged with GST, as indicated. GST-tagged proteins collected on GSH beads were stained with Coomassie blue (bottom panels). (D and E) Polyadenylation factors bound to GSH beads in the presence of the indicated FLASH and Lsm11 variants (lanes 3 to 6 in panel D and lanes 3 and 4 in panel E) were identified by Western blotting using specific antibodies. The recombinant proteins were stained with silver to monitor their recovery on GSH beads (bottom panels). The asterisk in panel E indicates a HeLa nuclear protein that nonspecifically binds to GSH beads and comigrates with FΔ28N (bottom panel, lane 2).

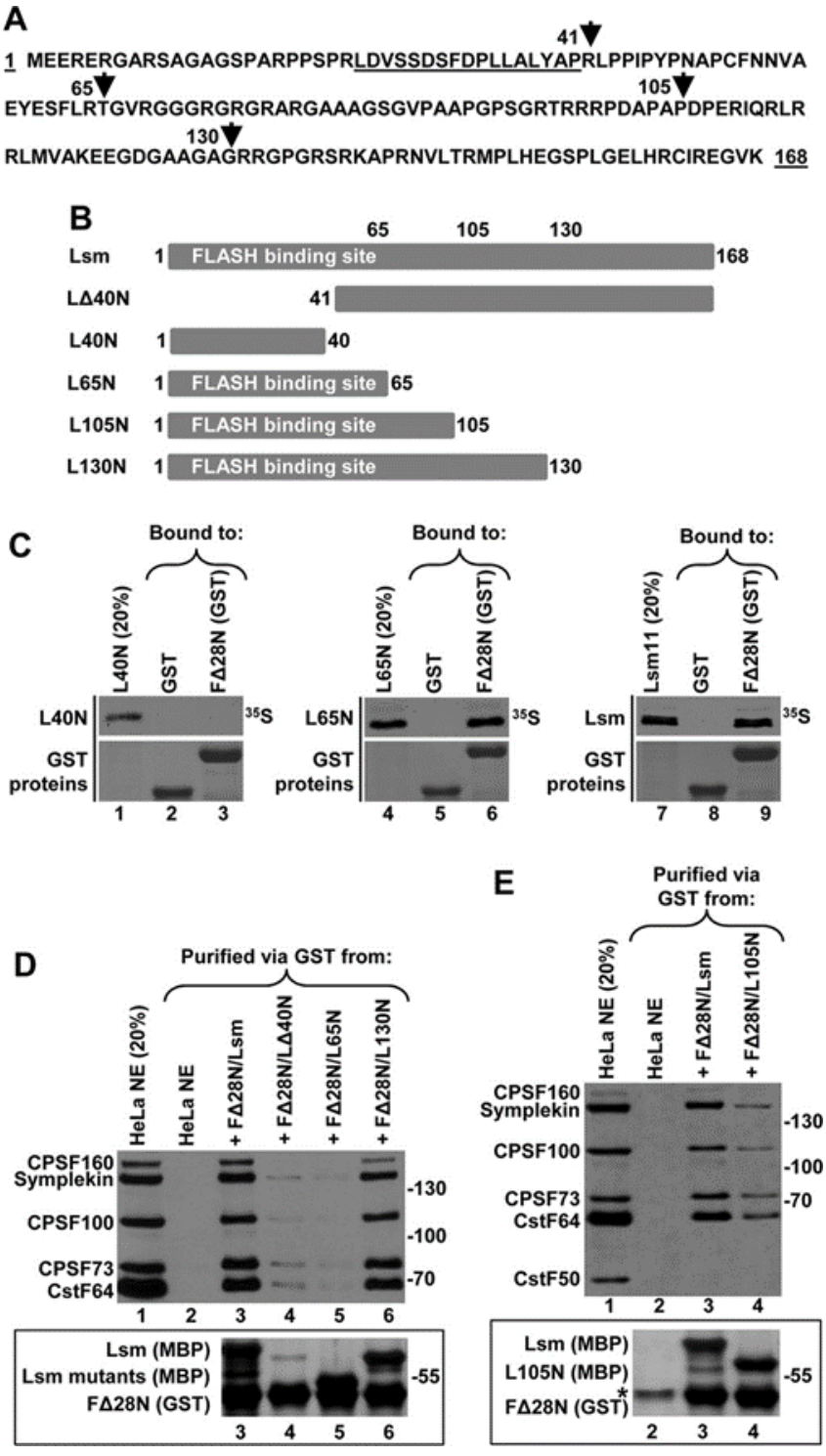

A
41
1 MEERERGARSAGAGSPARPPSPRLDVSSDSFDPLLALYAPRLPPIPYPNAPCFNNVA
65
105
EYESFLRTGVRGGGRGRGRARGAAAGSGVPAAPGPSGRTRRRPDAPAPDPERIQRLR
130
RLMVAKEEGDGAAGAGRRGPGRSRKAPRNVLTRMPLHEGSPLGELHRCIREGVK 168
B
65 105 130
Lsm 1 FLASH binding site 168
LΔ40N 41
L40N 1 40
L65N 1 FLASH binding site 65
L105N 1 FLASH binding site 105
L130N 1 FLASH binding site 130
C
Bound to:
L40N (20%)
GST
FΔ28N (GST)
L40N
GST proteins
35S
1 2 3
Bound to:
L65N (20%)
GST
FΔ28N (GST)
L65N
GST proteins
35S
4 5 6
Bound to:
Lsm11 (20%)
GST
FΔ28N (GST)
Lsm
GST proteins
35S
7 8 9
D
Purified via GST from:
HeLa NE (20%)
HeLa NE
+ FΔ28N/Lsm
+ FΔ28N/LΔ40N
+ FΔ28N/L65N
+ FΔ28N/L130N
CPSF160
Symplekin
CPSF100
CPSF73
CstF64
-130
-100
-70
1 2 3 4 5 6
Lsm (MBP)
Lsm mutants (MBP)
FΔ28N (GST)
-55
3 4 5 6
E
Purified via GST from:
HeLa NE (20%)
HeLa NE
+ FΔ28N/Lsm
+ FΔ28N/L105N
CPSF160
Symplekin
CPSF100
CPSF73
CstF64
CstF50
-130
-100
-70
1 2 3 4
Lsm (MBP)
L105N (MBP)
FΔ28N (GST)*
-55
2 3 4

Figure 10. Refined mapping of Lsm11 regions required for binding to FLASH. Binding of ^{35}S-labeled Lsm11 fragments to GST tagged N-terminal FLASHΔ28N (FΔ28N). Phosphoscreen images at top show radiolabeled Lsm11 fragments with 20% input in lane 1, GST control shown in lane 2, and FΔ28N in lane 3. Proteins collected on glutathione sepharose beads are shown below on Coomassie stained gel.

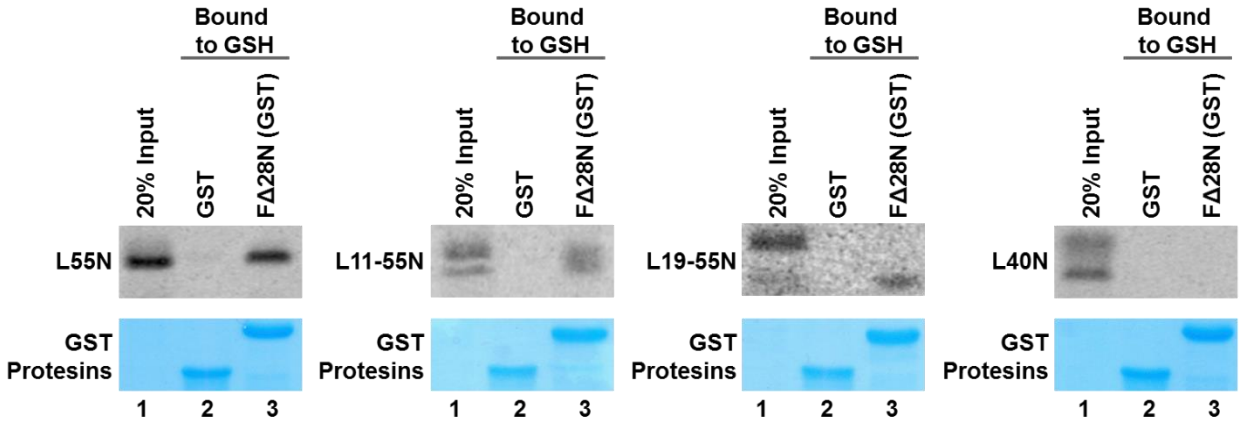
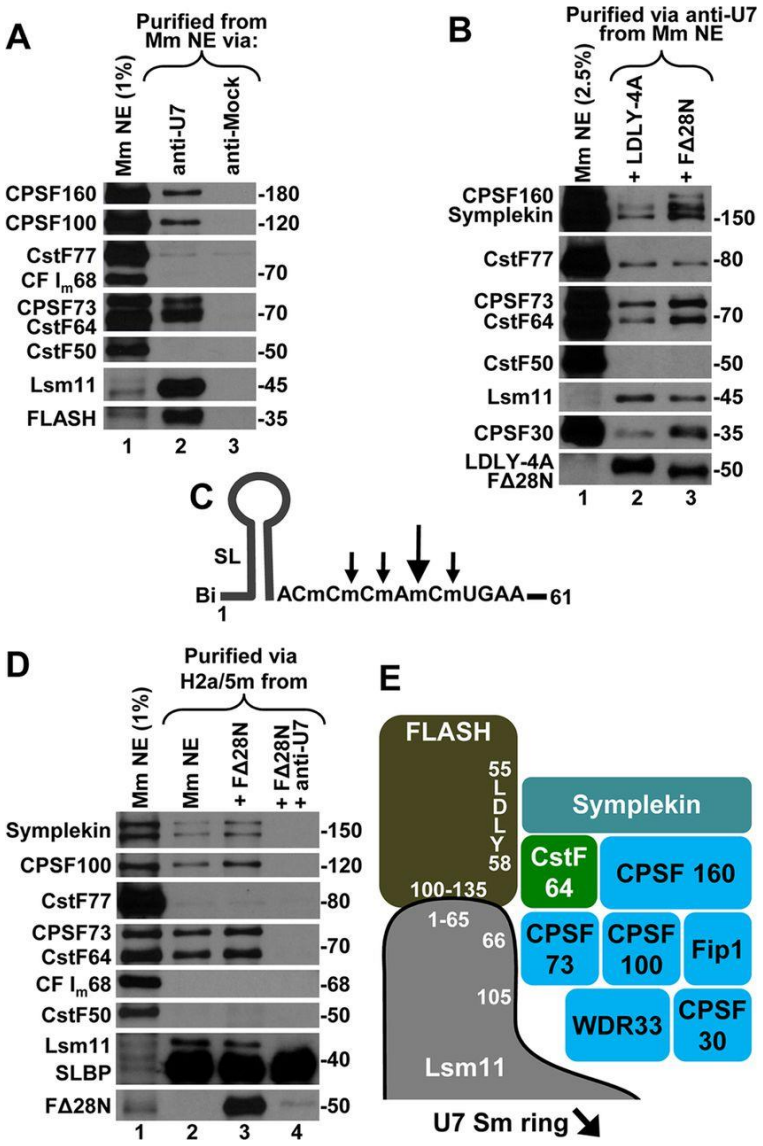

Bound to GSH
20% Input
GST
FΔ28N (GST)
L55N
GST Protesins
1 2 3
Bound to GSH
20% Input
GST
FΔ28N (GST)
L11-55N
GST Protesins
1 2 3
Bound to GSH
20% Input
GST
FΔ28N (GST)
L19-55N
GST Protesins
1 2 3
Bound to GSH
20% Input
GST
FΔ28N (GST)
L40N
GST Protesins
1 2 3

Figure 11. Polyadenylation factors bind to the endogenous U7 snRNP and histone pre-mRNA. (A) A mouse myeloma nuclear extract was incubated with the anti-U7 oligonucleotide (lane 2) or with an unrelated oligonucleotide (lane 3, anti-Mock), both containing a biotin tag at the 3' end. Bound proteins were collected on streptavidin beads and analyzed by Western blotting using specific antibodies. (B) The anti-U7 oligonucleotide (lane 2) was used to purify the U7 snRNP from a mouse myeloma nuclear extract supplemented with either LDLY-4A (lane 2) or wild-type FΔ28N (lane 3). Purified proteins were analyzed using specific antibodies. Note that the recombinant FLASH is not present in the input nuclear extract. (C) A depiction of the 61-nucleotide H2a/5m histone pre-mRNA in which 5 nucleotides adjacent to the major cleavage site (large arrow) and minor cleavage sites (small arrows) are modified with a 2′-O-methyl group. The H2a/5m pre-mRNA contains the upstream stem-loop, the HDE fully complementary to the U7 snRNA, and biotin at the 5' end. (D) The H2a/5m pre-mRNA was incubated with a mouse myeloma nuclear extract under control (lane 2) or in the presence of recombinant FΔ28N without (lane 3) or with the anti-U7 oligonucleotide (lane 4). Proteins bound to the H2a/5m substrate were collected on streptavidin beads and analyzed by Western blotting using specific antibodies. FΔ28N is not present in the input nuclear extract, and the band detected in the same region in lane 1 is a result of cross-reactivity with an antibody used concomitantly with the anti-FLASH antibody. (E) Composition of the polyadenylation complex recruited to the U7 snRNP and histone pre-mRNA by the Lsm11/FLASH complex. Amino acid regions in FLASH and Lsm11 involved in formation of the complex and binding the polyadenylation factors are indicated. The interaction between CstF64 and symplekin is mutually exclusive with the formation of the CStF complex, (CstF64, CstF77, and CstF50), which is involved in 3' end processing of canonical pre-mRNAs by cleavage and polyadenylation.


A
Purified from Mm NE via:
Mm NE (1%)
anti-U7
anti-Mock
CPSF160
CPSF100
CstF77
CF Im68
CPSF73
CstF64
CstF50
Lsm11
FLASH
-180
-120
-70
-70
-50
-45
-35
1 2 3
B
Purified via anti-U7 from Mm NE
Mm NE (2.5%)
+ LDLY-4A
+ FΔ28N
CPSF160
Symplekin
CstF77
CPSF73
CstF64
CstF50
Lsm11
CPSF30
LDLY-4A
FΔ28N
-150
-80
-70
-50
-45
-35
-50
1 2 3
C
SL
Bi
1
ACmCmCmAmCmUGAA
61
D
Purified via H2a/5m from
Mm NE (1%)
Mm NE
+ FΔ28N
+ FΔ28N + anti-U7
Symplekin
CPSF100
CstF77
CPSF73
CstF64
CF Im68
CstF50
Lsm11
SLBP
FΔ28N
-150
-120
-80
-70
-68
-50
-40
-50
1 2 3 4
E
FLASH
55
LDLY
58
100-135
1-65
66
105
Lsm11
U7 Sm ring
Symplekin
CstF 64
CPSF 160
CPSF 73
CPSF 100
Fip1
WDR33
CPSF 30

CHAPTER 3: THE U7 SNRNP AND FLASH ASSOCIATE WITH A DIFFERENT SUBSET OF CLEAVAGE AND POLYADENYLATION FACTORS LACKING CPSF30 AND FIP1 IN *DROSOPHILA*

INTRODUCTION

In Chapter 2, I showed bacterially expressed FLASH and Lsm11 N-terminal fragments, two proteins unique to histone pre-mRNA 3' end processing, interact to form a platform that stably associates with a subset of polyadenylation factors called the histone pre-mRNA cleavage complex (HCC) in mammalian nuclear extracts. The HCC contains all six subunits of CPSF, including the endonuclease CPSF73, in addition to symplekin and CstF64. The only proteins in the HCC that have been shown to be essential for processing are CPSF73, its catalytically inactive binding partner CPSF100, and symplekin. Unlike the larger polyadenylation complex required for 3' end processing of canonical pre-mRNAs, the HCC lacks two subunits of CstF, CstF50 and CstF77, and the two cleavage factors CF I_m and CF II_m. Interaction between the mammalian Lsm11/FLASH complex and HCC is critically dependent on a highly conserved LDLY motif in FLASH. The LDLY motif is essential for processing *in vitro* and mutating these amino acids abolishes the ability of FLASH to stimulate processing in mammalian nuclear extracts. The HCC associates with the U7 snRNP and is recruited to histone pre-mRNA in a FLASH dependent manner. Thus, at least a portion of the native U7 snRNP has a composite structure in which FLASH bound to the U7 snRNP through Lsm11 interacts with a subset of polyadenylation factors, including the endonuclease. This complex is brought to the histone pre-mRNA by partial base pairing, thereby behaving as an RNA guided endonuclease. In this

chapter, I show that in *Drosophila* a unique HCC associates with the U7 snRNP and is recruited to histone pre-mRNA.

Cleavage of the pre-mRNA yields a mature histone mRNA and a downstream cleavage product (DCP), which is degraded exonucleolitycally in the 5'-3' directon. Degradation of the DCP is dependent on the U7 snRNP and is important to release the U7 snRNP from the HDE for additional rounds of processing. The degradation reaction is catalyzed by CPSF73, the endonuclease activity can also function as a 5' to 3' exoncuclease (Dominski et al 2005). In mammals, FLASH does not have a role in DCP degradation, yet here I show it is essential for degradation in *Drosophila*.

FLASH is a large 220 kDa protein in mammalian cells and has been shown to interact with Ars2 through a domain in its central region. Ars2 was originally identified in a screen searching for genes conferring resistance to sodium arsenite in a hamster cell line. In *Drosophila*, Ars2 was shown to be a member of the Cap Binding Complex through an interaction with CBP80 and to play a role in the processing of pri-microRNAs to pre-microRNAs (Sabin et al., 2009; Gruber et al., 2009). Mutations of the Ars2 orthologue in *Arabidopsis* also show defects in microRNA biogenesis (Lobbes et al., 2006). Additionally, it is required for cellular proliferation (Gruber et al., 2009) and depletion of Ars2 in mammalian cells results in a small portion of *mis*-processed histone mRNA transcripts that are polyadenylated on the 3' end, suggesting a possible involvement in histone pre-mRNA processing. Here I examine whether the interaction between FLASH and Ars2 is conserved in *Drosophila*.

In *Drosophila* the interaction between FLASH and Lsm11 is conserved (Yang et al., 2009a) and FLASH is required for histone pre-mRNA 3' end processing *in vivo* (Burch et al., 2011). Similar to processing in mammals, CPSF73 is the endonuclease in *Drosophila* and the

only other polyadenylation factors shown to be essential for processing *in vivo* are CPSF100, and symplekin (Wagner et al., 2007). Whether or not there was a complex analogous to the mammalian HCC in *Drosophila* that associated with the U7 snRNP was unknown.

For this chapter, I prepared large scale *Drosophila* Kc nuclear extracts that are active in 3' end processing to characterize the U7 snRNP and associated complexes. Using a number of different protocols and purification approaches, I show the U7 snRNP co-purifies with a set of polyadenylation factors similar in composition to the mammalian HCC. The *Drosophila* HCC associates with the U7 snRNP and is recruited to the pre-mRNA in a U7-dependent fashion. Recruitment of this composite U7 snRNP to histone pre-mRNA is completely independent of the stem-loop or SLBP binding to the stem-loop. I contrast the *Drosophila* HCC with a larger super complex of cleavage and polyadenylation factors immunoprecipitated using an antibody we raised against CstF64. Furthermore, I depleted small scale *Drosophila* S2 cell cultures of FLASH by RNA interference and prepared nuclear extracts for *in vitro* 3' end processing assays to show that the LDLY sequence in mammals is functionally conserved in insects. I use these same extracts to demonstrate a critical role for FLASH in *Drosophila* DCP degradation. Finally, I show Ars2 does not interact in a complex with FLASH in *Drosophila* cells.

MATERIALS AND METHODS

***Drosophila* nuclear extracts and 3' end processing.** The 106-nt *Drosophila* H3 pre-mRNA used as a substrate for 3' end processing was generated by T7 RNA polymerase, treated with calf intestinal phosphatase (NEB), and labeled at the 5' end with ^{32}P using T4 polynucleotide kinase (NEB). 3' end processing reactions were carried out in nuclear extracts prepared from either S2

or Kc *Drosophila* cells, as previously described (Dominski et al., 2002; Dominski et al., 2005b). Unless otherwise indicated, the extracts were prepared from freshly harvested cells. The cells were grown at room temperature in the serum-free D-22 medium (US Biological) using spinner flasks.

Synthetic RNAs

The following RNAs (written in the 5'–3′ orientation) were used in this study:

-Biot-dH3 pre-mRNA (63-mer):

(Bi)(18S)(18S)CUCUAUAAUCGGUCCUUUUCAGGACCACAAACCAGAUUCAAUGAGAUAAAAUUUUCUGUUGCC

-Biot-dH3ΔSL (47-mer):

(Bi)(18S)(18S)CUCUAUAAUCACAAACCAGAUUCAAUGAGAUAAAAUUUUCUGUUGCC

-Biot-mH2a pre-mRNA (62-mer):

(Bi)(18S)(18S)CUCCCAAAAAGGCUCUUUUCAGAGCCACCCACUGAAUCAGAUAAAGAGUUGUGUCACGGUAG

-Biot-SL (31-mer):

(Bi)(18S)(18S)GUGCCAAAAAGGCUCUUUUCAGAGCCACCCA

-Biot-αU7 (20-mer):

mCmAmAmAmGmAmGmAmAmUmAmAmAmAmAmUmUmUmUmC(18S)(18S)(Bi)

-Biot-αMock (17-mer):

mAmAmAmGmAmGmCmUmGmUmAmAmCmAmCmUmU(18S)(18S)(Bi)

-αU7 (17-mer):

mCmAmAmAmGmAmGmAmAmUmAmAmAmAmAmUmU

-WT SL (31-mer):

GUGCCAAAAAGGCUCUUUUCAGAGCCACCCA

-Mut SL (31-mer):

GUGCCAAUUAGCCUCUUUUCAGAGGCACCCA

where (Bi) biotin; (18S) 18-atom spacer; and (m) 2′-O-methyl group. The Biot-dH3ΔSL RNA was synthesized by IDT. The remaining RNAs were ordered from Dharmacon.

RNAi in *Drosophila* S2 cells. For preparation of nuclear extracts from RNAi-depleted cells, a total of 10^6 *Drosophila* S2 cells were plated in serum-free SF900 medium (GIBCO) and treated daily with 10 µg of double-stranded (ds) RNAs against selected *Drosophila* genes. After 5 d of treatment, cells were collected and used to prepare small-scale nuclear extracts. Each preparation yielded ~100 µL of a nuclear extract. For analyzing the expression of GFP reporters, *Drosophila* S2 cells were plated as described above and treated with dsRNA for 48–72 h. Following the treatment, cells were transfected with appropriate GFP reporter plasmids, as described (Wagner et al., 2007; Burch et al., 2011) and 24 h later viewed on an Olympus IX81-ZDC inverted fluorescence microscope.

Degradation of *Drosophila* DCP. Degradation reactions were carried out in nuclear extracts prepared from *Drosophila* S2 cells (see above). Extracts were incubated with 5’ end labeled DCP

substrate for 120 minutes in the presence of 10mM EDTA and reactions stopped by treating mixtures for 1 h with 0.5mg/mL of proteinase K. Products were diluted in a 7M urea gel and resolved on 8% polyacrylamide 7M urea gels.

Mutagenesis and protein expression. Mutations in the N-terminal fragment of *Drosophila* FLASH (amino acids 1–178) were generated using PCR and appropriately altered oligonucleotide primers. *Drosophila* FLASH fused N-terminally to glutathione S-transferase (GST) was expressed in bacteria from the pET-42a vector. *Drosophila* Lsm11 (amino acids 1–153) fused N-terminally to maltose binding protein (MBP) was expressed in bacteria from the pDEST566 vector. Both recombinant proteins were purified on nickel beads (QIAGEN) using the His-tag present in each protein.

Binding of *Drosophila* nuclear proteins to *Drosophila* Lsm11/FLASH complex. The *Drosophila* Lsm11/FLASH complex was formed by mixing 100 pmol of each recombinant protein in 100 µL of buffer D (100 mM KCl, 20 mM HEPES at pH 7.9, 20% glycerol, 0.2 mM EDTA at pH 8, 0.5 mM DTT). All the subsequent steps were carried out as described previously (Yang et al., 2013), with the exception that 100 µL of *Drosophila* nuclear extracts from either Kc or S2 cells was used instead of mammalian nuclear extracts.

Formation of processing complexes and purification of the *Drosophila* U7 snRNP. Processing complexes were assembled in a final volume of 4 mL that contained 3 mL of a nuclear extract from Kc cells, 2.5 µg of the 63-nt Biot-dH3 pre-mRNA, 50 µg of yeast tRNA (Invitrogen), and 10 mM EDTA. The samples were incubated for 5 min at 22°C followed by 1 h of rotation at 4°C. The RNA substrate and associated proteins were bound to streptavidin beads, washed several times with buffer D/10 mM EDTA for a total of 1.5 h, and separated on a SDS/polyacrylamide gel. A similar approach was used for the single-step purification of the

endogenous U7 snRNP with the exception that the Biot-αU7 2′-O-methyl oligonucleotide containing biotin at the 3’ end was used instead of the Biot-dH3 pre-mRNA.

Immunoprecipitations. A single sample typically contained 1 mL of a nuclear extract (~10 mg of total protein) and 3–5 μg of an affinity-purified specific antibody, and it was rotated for 4–14 h at 4°C. Following incubation, the extract was spun down at 10,000 rpm in a table-top microcentrifuge and rotated for 2 h with ~30 μL of protein A plus agarose beads (Pierce) to purify protein complexes bound to the antibody. The beads were washed with buffer D (see above) for a total of 2 h, moved to a new tube, and resuspended in an SDS sample buffer. The material immobilized on the beads was separated in a single SDS/polyacrylamide gel for analysis by silver staining/mass spectrometry or in three or four gels for analysis by Western blotting using various antibodies.

Mass spectrometry. Protein associated with the *Drosophila* U7 snRNP, Biot-dH3 pre-mRNA, or precipitated by specific antibodies was separated on SDS/polyacrylamide gels, excised from the gel, treated with trypsin, and identified by LC-MS/MS using the Nano-Acquity LC system (Waters) and Orbitrap Velos mass spectrometer (Thermo Electron Corp.), as described (Yang et al., 2013).

Antibodies. Custom-made antibodies against *Drosophila* Lsm11 and *Drosophila* CstF50 were kindly provided by J. Gall and John Lis, respectively. Antibodies against the N-terminal portion of Ars2 (amino acids 1–167 of total 769 residues), full-length *Drosophila* CstF64 (amino acids 1–437), and the N-terminal portion of human CF I_m68 (amino acids 1–191), all N-terminally tagged with 6xHis and GST, were generated in rabbits by Pacific Immunology and affinity-purified using their respective protein antigens. Antibodies against *Drosophila* FLASH, CPSF73, CPSF100, and symplekin were described previously (Sullivan et al., 2009b; Yang et al., 2009a).

RESULTS

Characterization of Kc nuclear extracts used in biochemical studies on 3' end processing. The 3' end processing machinery that cleaves replication-dependent histone pre-mRNAs in *Drosophila* shares a number of similarities with the mammalian machinery (Fig. 12A), including the essential role played by the interaction of Lsm11 and FLASH (Burch et al., 2011). In mammalian cells, these two proteins form a platform that serves to recruit multiple polyadenylation factors to the U7 snRNP (Yang et al., 2013). In this study, we used large-scale nuclear extracts from Kc cells and several biochemical approaches to determine the role of the Lsm11/FLASH interaction and composition of the U7-dependent processing machinery in *Drosophila*.

In our previous studies on 3' end processing of histone pre-mRNAs in *Drosophila*, we typically used large-scale nuclear extracts prepared from frozen Kc cells since they had virtually the same processing activity as nuclear extracts prepared from freshly harvested cells (Dominski et al., 2002; Dominski et al., 2005b). However, closer biochemical analysis revealed that these extracts contained significantly reduced levels of two essential processing factors: SLBP and FLASH (Fig. 12B, compare. lanes 1 and 2). The reduced level of SLBP was primarily caused by its rapid leakage to the cytoplasm during cell thawing. FLASH, in addition to leaking to the cytoplasm, was degraded to shorter fragments that are fully functional in 3' end processing *in vitro* (see below). As a result of leakage, cytoplasmic extracts prepared from frozen cells are active in processing, indicating that they contain complete processing machinery. Clearly, relatively small amounts of SLBP, FLASH, and other processing factors are sufficient to support

robust processing activity of nuclear and cytoplasmic extracts when tested with a limiting amount of the dH3 pre-mRNA substrate.

Since nuclear extracts prepared from freshly harvested Kc cells were highly enriched in full-length processing factors, our subsequent biochemical experiments were carried out using these extracts. In agreement with our previous results (Dominski et al., 2002, 2005b), 3' end processing of histone pre-mRNAs carried out in these extracts is absolutely dependent on SLBP. The presence of 10 μM wild-type stem–loop RNA competitor (WT SL, 500-fold excess over the pre-mRNA substrate), which sequesters the entire pool of SLBP in the extract inhibited cleavage of the dH3 pre-mRNA, whereas the same amount of a mutant stem–loop RNA (Mut SL) had no effect (Fig. 12C, compare lanes 4 and 5). This mutant RNA contains alterations in the 5'-flanking adenosines and the second base pair of the stem (Fig. 12E) and does not bind SLBP (Yang et al., 2009b; Tan et al., 2013). Processing was also efficiently inhibited by 5 μM anti-U7 2′-O-methyl oligonucleotide (αU7) complementary to nucleotides 5–19 of the *Drosophila* U7 snRNA (Fig. 12C, lane 7) and by an antibody directed against the N-terminal portion of *Drosophila* FLASH (Fig. 12D, lane 3). A nonspecific 2′-O-methyl oligonucleotide targeted to the first 17 nt of the mouse U7-snRNA (αMock) (Fig. 12C, lane 8) and a nonspecific antibody (Fig. 12D, lane 4) did not reduce *in vitro* 3' end processing. All these reagents had the same effect on processing in *Drosophila* S2 nuclear extracts prepared from cells grown either in monolayers (ML S2) or suspension cultures (Sus S2). We conclude that the essential role of SLBP, the U7 snRNP and FLASH in 3' end processing of histone pre-mRNAs is a universal feature of *Drosophila* nuclear extracts. This conclusion is consistent with the essential role of these factors in processing *in vivo* (Sullivan et al., 2001; Godfrey et al., 2006; Burch et al., 2011).

Functional conservation of *Drosophila* FLASH. A short N-terminal fragment of human FLASH encompassing amino acids 52–139 strongly stimulates processing in diluted or poorly active mammalian nuclear extracts (Yang et al., 2009a, 2011). This fragment of FLASH tightly binds the N-terminal fragment of human Lsm11 (amino acids 1–170), and together they interact with a large nuclear complex consisting of multiple polyadenylation factors, including the endonuclease CPSF73 (Yang et al., 2013). We refer to this complex as the Histone pre-mRNA Cleavage Complex (HCC). By introducing various deletions and mutations into the N-terminal fragment of human FLASH, we demonstrated that amino acids 100–135 bind Lsm11, whereas amino acids 55–58 with the highly conserved LDLY sequence are required for the recruitment of the HCC (Yang et al., 2011).

Our previous studies in *Drosophila* cultured cells showed that the region located between amino acids 65 and 77 of *Drosophila* FLASH is essential for processing of histone pre-mRNAs *in vivo*, but its role has not been determined (Burch et al., 2011). This region contains an LDIY sequence (Fig. 13A), which is likely the functional counterpart of the LDLY motif in mammalian FLASH. Surprisingly, the N-terminal fragment of *Drosophila* FLASH (FN-ter, amino acids 1–178) did not stimulate 3' end processing in *Drosophila* Kc nuclear extracts (Fig. 13B, lane 4). Moreover, a deletion mutant of this fragment lacking the first 77 amino acids (FΔ77), including the LDIY motif, did not act in a dominant-negative fashion *in vitro* (Fig. 13B, lane 5), although the equivalent mutant of human FLASH strongly inhibited processing in mammalian extracts (Yang et al., 2011). Note that as indicated above, FLASH is an essential component of *Drosophila* processing machinery, and the presence of an anti-FLASH antibody inhibits cleavage of the dH3 pre-mRNA in this and all other tested *Drosophila* nuclear extracts.

In contrast to mammalian cells, where FLASH is severely limiting, *Drosophila* cultured cells contain relatively high levels of this protein, either full length or its shorter proteolytic fragments, which could be precipitated by specific antibodies and visualized on SDS/polyacrylamide gels by Coomassie and silver staining (Fig. 13B). We reasoned that the endogenous *Drosophila* FLASH is in excess to Lsm11 and quantitatively bound to the U7 snRNP, providing a potential explanation for the failure of the recombinant N-terminal FLASH to stimulate processing. To test this possibility, we reduced the endogenous level of FLASH by RNA interference (RNAi). *Drosophila* S2 cells were grown in monolayers for 5 d in the absence or presence of double-stranded (ds) RNA against *Drosophila* FLASH and used to prepare small-scale nuclear extracts. A nuclear extract from untreated cells contained readily detectable amounts of FLASH (Fig. 13C, lane 1) and was proficient for processing of the dH3 histone pre-mRNA (Fig. 13D, lane 2). Recombinant N-terminal FLASH (FN-ter, amino acids 1–178) increased the efficiency of cleavage approximately twofold in this nuclear extract (Fig. 13D, lane 3; Fig. 2F), whereas the FΔ77 FLASH mutant typically had on average a minor inhibitory effect (Fig. 13D, lane 4; Fig. 13F). It is unclear why the FN-ter and FΔ77 FLASH proteins affected processing only in the small-scale nuclear extracts from S2 cells and showed no activity when added to large-scale nuclear extracts prepared from suspension Kc cells (Fig. 13B, lanes 4, 5). S2 cells adapted to grow in suspension cultures yielded large-scale nuclear extracts that behave in the same manner as Kc nuclear extracts. One possible explanation is that extracts from monolayer cells differ from extracts from suspension cells in the abundance of FLASH and other processing factors and their accessibility for recombinant FLASH.

The nuclear extract I prepared from S2 cells treated with dsRNA against FLASH contained a significantly reduced level of FLASH (Fig. 13C, lane 2), whereas the amounts of

SLBP and the U7 snRNP (Fig. 13C, lane 2), the two other critical processing factors, were unchanged. This extract alone poorly processed histone pre-mRNA (Fig. 13D, lane 5), and the addition of the recombinant FN-ter protein (amino acids 1–178) increased its activity more than fivefold, to a level that was higher than that of the control extract (Fig. 13D, cf. lanes 2 and 7; Fig. 13F). The stimulation of the processing efficiency by the N-terminal FLASH was highly reproducible and observed with every nuclear extract independently depleted of FLASH (Fig. 13F). The FΔ77 FLASH (amino acids 78–178) lacking the LDIY mutant was inactive (Fig. 13D, lane 6; Fig. 13F). To determine whether the lack of activity is caused by the absence of the LDIY motif, I created the LDIY-4A point mutant in which this motif was replaced by four alanines. This protein had only a residual activity in processing (Fig. 13E, lane 4; Fig. 13F), demonstrating that the LDIY motif in FLASH is critical for 3' end processing of histone pre-mRNAs in *Drosophila* and supporting the notion that it is a functional counterpart of the LDLY motif in mammalian FLASH.

Based on unpublished reports that a second LDIY motif located between amino acids 45-48 was capable of rescuing the LDIY-4A mutant *in vivo* in *Drosophila* flies, (Deirdre Tatomer et al., submitted), I expressed a double mutant fragment of FLASH (amino acids 1-178) with both LDIY motifs changed to 4 alanines (LDIY-4A(x2)). N-terminal FLASH (amino acids 1-178) clearly rescues 3' end processing efficiency in FLASH depleted nuclear extracts (Fig. 14C compare lanes 2-3, 4-5, and 6-7) while the LDIY-4A mutant rescues only 10-15% of processing (Fig. 14C lanes 8-9 compared to lanes 4-5). Surprisingly, *in vitro* processing assays showed a virtual elimination of processing ability in FLASH depleted S2 nuclear extracts when comparing the LDIY-4A and LDIY-4A(x2) stimulated reactions (Fig. 14C, compare lanes 8-9 with lanes 10-11). In light of Deirdre's unpublished data indicating the single LDIY motif located at amino

acids 45-48 is capable of rescuing the fly, this additional result suggests a possible redundant role for the first LDIY that although is considerably less efficient *in vitro* is sufficient to rescue the fly *in vivo*.

The interaction between a complex of the N-terminal FLASH and Lsm11 and *Drosophila* polyadenylation factors. We next tested whether various recombinant N-terminal FLASH fragments (GST fusion) bound to the N-terminal Lsm11 (MBP fusion) can interact with polyadenylation factors in *Drosophila* nuclear extracts, as previously described for the mammalian proteins (Yang et al., 2013). The bound proteins were purified on glutathione beads and analyzed by Western blotting using antibodies against the three *Drosophila* polyadenylation factors known to be essential for 3' end processing of histone pre-mRNAs in *Drosophila* and mammalian cells: symplekin, CPSF100, and CPSF73 (Sullivan et al., 2009b). The level of all recombinant protein absorbed on glutathione beads was carefully monitored by staining with Coomassie Blue.

Initially, we used small-scale nuclear extracts prepared from untreated and FLASH-depleted S2 cells since their processing activity was increased by recombinant FLASH. A complex of the N-terminal regions of *Drosophila* FLASH (FN-ter, amino acids 1–178) and Lsm11 (amino acids 1–153) interacted with readily detectably amounts of symplekin and CPSF100 in extracts from both untreated (Fig. 15A, lane 3) and FLASH-depleted S2 cells (Fig. 15A, lane 7). Among the interacting proteins we also identified CPSF73, although its detection on Western blots was obscured by a bacterial protein present in the preparation of recombinant Lsm11 that strongly cross-reacted with the anti-CPSF73 antibody (see Fig. 15C, lanes 3, 4). Importantly, the interaction of all these polyadenylation factors was nearly abolished by deleting the first 77 amino acids from the FN-ter FLASH (FΔ77) (Fig. 15A, lanes 4, 8).

Of multiple large-scale Kc nuclear extracts tested by this approach, only some behaved in the same fashion as the small-scale extracts prepared from S2 cells. In these Kc extracts, the FN-ter/Lsm11 complex, but not the mutant FΔ77/Lsm11 complex, interacted with symplekin, CPSF100 (Fig. 15B, compare lanes 4 and 5) and CPSF73 (see also Fig. 15C, lane 3). Consistent with our previous results with the mammalian system, the N-terminal FLASH alone was less efficient in binding the polyadenylation factors than its complex with Lsm11 (Fig. 15B, compare lanes 3 and 4). To determine whether the FΔ77 FLASH mutant complexed with the N-terminal Lsm11 is unable to bind the polyadenylation factors primarily due to deletion of the LDIY motif rather than any other sequence in the first 77 amino acids of FLASH, we used the LDIY-4A mutant. The interaction of the LDIY-4A/Lsm11 complex with symplekin, CPSF100, and CPSF73 in the same Kc nuclear extract was significantly reduced (Fig. 15C, compare lanes 3 and 4), demonstrating that the LDIY motif plays an equivalent function to the LDLY motif in mammalian FLASH and is essential for recruiting the CPSF73 endonuclease and other polyadenylation factors to the U7 snRNP.

FLASH is essential for degradation of the DCP after cleavage of histone pre-mRNA. In mammalian extracts, degradation of the downstream cleavage product (DCP) released after cleavage depends on the 5’ to 3’ exonuclease activity of CPSF73 (Dominski et al., 2005c). The DCP contains a HDE and degradation of this cleavage product is U7-dependent. Adding α-FLASH antibody or mutated fragments of recombinant FLASH lacking the conserved LDLY motif to mouse myeloma extracts inhibits 3’ end processing of labeled H2a substrate. But it does not inhibit degradation of a separate 5’-labeled DCP under the same conditions, demonstrating FLASH is not required for degradation of the DCP *in vitro*. Furthermore, supplementing the

extract with wild type N-terminal FLASH showed no change in degradation efficiency (Yang et al., 2011).

To see if this is the same in *Drosophila*, I prepared small scale nuclear extracts from *Drosophila* S2 cells. Blocking the interaction between the U7 snRNP and HDE with an anti-U7 oligo prevented degradation of the DCP in extracts from untreated S2 cells, as expected (Fig. 16, compare lanes 2 and 3). Surprisingly, the addition of F-Nter had a stimulatory effect in the extract, while adding α-dFLASH antibody clearly blocked degradation of the DCP (Fig. 16, compare lane 2 to lanes 4 and 5). In S2 extracts depleted of FLASH using RNA interference, degradation of the DCP was reduced significantly compared to the untreated extract (Fig 16, lane 6 compared to lane 2). Upon addition of F-Nter, degradation is restored to levels surpassing that of the untreated extract (Fig. 16, compare lanes 7 and 2). This effect is at least partially dependent on the conserved LDIY motif as changing these residues to alanines reduces the stimulatory effect of adding recombinant FLASH (Fig 16, compare lanes 7 and 8). These results indicate FLASH is an essential factor for degradation of the DCP in *Drosophila*.

In *Drosophila* Ars2 does not stably associate in a complex with FLASH and is dispensable for histone pre-mRNA 3' end processing. In mammalian cells, FLASH interacts with Ars2, and this interaction is required for S-phase progression, expression of replication-dependent histone mRNAs, and efficient localization of FLASH to Histone Locus Bodies (HLBs) (Kiriyama et al., 2009). In addition, Ars2 is essential for cell proliferation (Gruber et al., 2009), and depletion of Ars2 from human cells increases the number of aberrant replication-dependent histone mRNAs that end with a poly(A) tail (Gruber et al., 2012), suggesting that Ars2 may cooperate with FLASH in U7-dependent 3' end processing.

I tested whether *Drosophila* Ars2 and FLASH form a stable complex in *Drosophila* cells. Xiao Yang generated a rabbit antibody against the N-terminal portion of Ars2 to isolate complexes containing Ars2 from Kc nuclear extracts and identify interacting proteins by mass spectrometry. As determined by Western blotting, the antibody precipitated Ars2 but not FLASH (Fig. 17A, lane 3). Mass spectrometry analysis of a large-scale precipitation with the anti-Ars2 antibody additionally identified large amounts of Cap Binding Protein 80 (CBP80) in the precipitate (Fig. 17B). This 80 kDa subunit of the nuclear RNA Cap Binding Complex (CBC) and the smaller 20 kDa subunit interact with tagged versions of Ars2 in human (Gruber et al., 2009) and *Drosophila* cells (Sabin et al., 2009), and they likely interact with Ars2 in plants (Laubinger et al., 2008). The anti-Ars2 precipitate also contained homologs of other mammalian proteins, including the helicase Mtr4 (Fig. 4B), shown to exist in a complex with Ars2 in mammalian cells (Lubas et al., 2011). However, both Western blotting (Fig. 17A, lane 3; Fig. 18A, lane 2) and mass spectrometry (Fig. 17B) demonstrated that the antibody against Ars2 failed to coprecipitate FLASH, suggesting that these two proteins do not form a stable complex in *Drosophila* cells. The anti-Ars2 antibody also did not precipitate Lsm11 (Fig. 18A, lane 2). The conclusion that FLASH and Ars2 are not part of the same complex was also supported by a reciprocal experiment with anti-FLASH antibodies directed against the N-terminal (Fig. 17A, lane 2) and C-terminal regions of *Drosophila* FLASH. Finally, no Ars2 was detected in the material precipitated by an anti-Lsm11 antibody, although this antibody precipitated FLASH (Fig. 17A, lane 5). The inability of the anti-FLASH and anti-Ars2 antibodies to coprecipitate Ars2 and FLASH, respectively, was highly reproducible and observed for a broad range of nuclear extracts, including those prepared from S2 cells grown in monolayers (Fig. 17A, lanes 6–

9) or in high-volume suspension cultures. Collectively, these results suggest that in *Drosophila*, Ars2 does not interact with FLASH and is not part of the U7 snRNP.

Since it was possible that *Drosophila* Ars2 functions in processing of histone pre-mRNAs via interacting with a different processing component, we next reduced the endogenous level of Ars2 in S2 cells by RNA interference (RNAi) and tested whether this treatment results in *in vivo* production of replication-dependent histone mRNAs that end with a poly(A) tail rather than the stem–loop. This phenotype was readily observed in S2 cells depleted of FLASH (Yang et al., 2009a) and other factors essential for U7-dependent 3' end cleavage of histone pre-mRNAs in *Drosophila*, including symplekin, CPSF100, and CPSF73 (Wagner et al., 2007; Sullivan et al., 2009b; Yang et al., 2009a) and was also observed in Ars2-depleted mammalian cells (Gruber et al., 2012). Xiao Yang monitored the extent of depletion by Western blotting, using the antibody against the N-terminal portion of Ars2 (Fig. 17C). The antibody detects a single band at ~140 kDa in the whole cell lysate from untreated S2 cells, and the corresponding band is virtually missing in the lysate from S2 cells depleted of Ars2 by treatment with dsRNA (Fig. 17C, lane 6). RNAi with dsRNA specific for *Drosophila* FLASH or a number of proteins unrelated to 3' end processing of histone pre-mRNA had no effect on the abundance of Ars2 in S2 cells (Fig. 17C, lanes 2–5).

Cells depleted of FLASH produced longer polyadenylated dH3 histone mRNA resulting from a defect in the U7-dependent 3' end processing (Yang et al., 2009a) (Fig. 17D, lane 1). Importantly, depletion of Ars2 had no visible effect on this process (Fig. 17D, compare lanes 1 and 6), although it resulted in a noticeable reduction of the overall level of the dH3 mRNA relative to the level of *Drosophila* 7SK RNA, a transcript generated by RNA Pol III. The same general reduction in expression of replication-dependent histone mRNAs was previously

observed in Ars2-depleted human cells (Kiriyama et al., 2009; Gruber et al., 2012) and is also seen in *Drosophila* S2 cells depleted of mxc (Fig. 17D, lane 3), the *Drosophila* orthologue of NPAT (White et al., 2011). NPAT is specifically required for transcription of histone genes in mammalian cells (Ma et al., 2000).

Since Ars2 may be essential for transcription of histone genes, thereby making it difficult to assess its role in 3' end processing of histone pre-mRNAs *in vivo*, I prepared nuclear extracts from Ars2-depleted S2 cells. In contrast to FLASH-depleted S2 extracts, extracts prepared from Ars2-depleted cells were consistently very active in processing the dH3 pre-mRNA, further arguing that Ars2 is not a component of the *Drosophila* U7-dependent processing machinery.

Nuclear proteins associated with *Drosophila* FLASH. To better understand the role of FLASH in 3' end processing of histone pre-mRNAs and to identify its binding partners in *Drosophila* nuclear extracts, we carried out a series of immunoprecipitations using antibodies against FLASH and other *Drosophila* processing factors. The antibody against the N-terminal portion of *Drosophila* FLASH efficiently precipitates U7 snRNA (Yang et al., 2009a) and Lsm11 (Fig. 18A, lane 3), indicating that a substantial fraction of the endogenous U7 snRNP is stably associated with FLASH. Consistently, an anti-Lsm11 antibody coprecipitates *Drosophila* FLASH from Kc nuclear extracts (Fig. 17A, lane 5).

I next investigated whether the endogenous *Drosophila* FLASH is associated with any of the polyadenylation factors known to be components of the mammalian HCC. As shown by Western blotting, the anti-FLASH antibody, but not the anti-Ars2 antibody, precipitated two polyadenylation factors that are essential for 3' end processing of *Drosophila* histone pre-mRNAs: symplekin and CPSF73 (Fig. 18A, lane 3). To confirm this result and to identify other proteins associated with *Drosophila* FLASH, I scaled up precipitation with the anti-FLASH. In

addition to FLASH, silver staining revealed several specific protein bands in the anti-FLASH precipitate that were identified by mass spectrometry as *Drosophila* orthologues of CPSF160 (CG10120), symplekin (CG2097), CPSF100 (CG1957), and CPSF73 (CG7698), which comigrated with the 70 kDa heat shock protein (HSC70, CG4262) (Fig. 5B). The precipitate also contained peptides from the *Drosophila* orthologue of WDR33 (CG1109), although this protein was not visible as a separate band on the silver-stained gel. All these proteins were either undetectable or detected with very low scores in the anti-Ars2 precipitate (Fig. 18A, lane 2). Consistent with FLASH and symplekin forming a common complex, an antibody against *Drosophila* symplekin (Fig. 18C, lane 3), but not a mock antibody directed against an unrelated *Drosophila* protein (Fig. 18C, lane 4), coprecipitated FLASH.

I carried out additional immunoprecipitation experiments with the anti-FLASH antibody and analyzed the precipitated material in higher-percentage gels to look for smaller proteins. These experiments failed to identify any additional polyadenylation factors associated with FLASH, including the *Drosophila* orthologue of CPSF30 (CG3642), which based on its amino acid sequence is expected to migrate at ~30 kDa. The analysis did not include the area of the gel near the 50 kDa size marker that contained large amounts of the immunoglobulin heavy chain. Altogether, these results demonstrate that FLASH and Lsm11 in *Drosophila* nuclear extracts are tightly associated with each other and suggest that they recruit several polyadenylation factors to the U7 snRNP.

Affinity purification of *Drosophila* U7 snRNP from Kc nuclear extracts. I next affinity-purified *Drosophila* U7 snRNP from Kc nuclear extracts using a 2′-O-methyl oligonucleotide (Biot-αU7) complementary to the 5’ end of the *Drosophila* U7 snRNA (Dominski et al., 2005b).

This oligonucleotide contains biotin at the 3’ end, allowing adsorption of the associated complexes on streptavidin beads.

I initially carried out a pilot experiment with a small amount of a Kc nuclear extract (500 µL) and determined that the Biot-αU7 oligonucleotide binds both specifically and efficiently to the U7 snRNP. The material purified by this method was highly enriched in Lsm11 and also contained readily detectable amounts of FLASH (Fig. 18D, lane 3), confirming that FLASH is an integral part of the U7 snRNP. As expected, the material did not contain SLBP, which could be specifically purified by the 31-nt Biot-SL RNA, containing the sequence of the stem–loop and biotin on the 5’ end (Fig. 18D, lane 4). A nonspecific 2′-O-methyl oligonucleotide complementary to the first 17 nt of the human U7 snRNA and tagged with biotin at the 3’ end (Biot-αMock) bound neither the U7 snRNP nor SLBP (Fig. 18D, lane 2). Importantly, the material purified by the Biot-αU7 also contained detectable amounts of symplekin, CPSF100, and CPSF73 but lacked CstF50 (Fig. 18E, lane 2). These results strongly support the conclusion from the immunoprecipitation experiments that the *Drosophila* U7 snRNP associates with FLASH and a complex of polyadenylation factors that resembles the mammalian HCC, which also lacks the 50 kDa subunit of CStF (Yang et al., 2013).

To determine whether additional proteins associate with the U7 snRNP, we used larger amounts of both the Biot-αU7 and Kc nuclear extract (1.0 µg and 5.0 mL, respectively) and analyzed the purified material by mass spectrometry. Nuclear proteins bound to the Biot-αU7 were collected on streptavidin beads, separated on an 8% SDS/polyacrylamide gel, and visualized by silver staining (Fig. 18F, lane 1). Control purification was carried out with the Biot-αMock oligonucleotide (Fig. 18F, lane 2). Several clearly visible silver-stained protein bands were detected only in the lane containing material bound to the Biot-αU7 oligonucleotide

(Fig. 18F, lane 1). These bands and other regions of the gel were analyzed by mass spectrometry (Fig. 18F, right panel). The heavily stained band near the top of lane 1 that is completely missing in lane 2 was identified as FLASH (Fig. 18F, band 2). The less-intense band migrating immediately above FLASH (band 1) was identified as *Drosophila* CPSF160, whereas the bands migrating faster than FLASH were identified as symplekin (band 3), CPSF100 (band 4), WDR33 (band 5), and CstF64 (band 9) (Fig. 18F). Lane 1 also contained CPSF73, although it was masked by massive amounts of another protein identified as the *Drosophila* orthologue of Insulin-like Growth Factor 2 mRNA Binding Protein 1 (IGF2BP1, CG1691). This protein likely directly interacts with the Biot-αU7 oligonucleotide (see below). All the polyadenylation factors purified by the Biot-αU7 oligonucleotide were undetectable by mass spectrometry in lane 2, whereas IGF2BP1 was present in this lane in trace amounts.

To analyze smaller proteins purified via the Biot-αU7, two-thirds of the same material was separated on a 12% SDS/polyacrylamide gel (Fig. 18G, lane 1) and compared with the material bound to the Biot-αMock oligonucleotide (Fig. 18G, lane 2). Two silver-stained bands were clearly visible at ~30 and 22 kDa in lane 1, and they were absent from lane 2. These bands were identified as *Drosophila* Lsm11 and SmB, the two largest components of the U7-specifc Sm ring, confirming that the Biot-αU7 oligonucleotide is very efficient and specific in purifying the *Drosophila* U7 snRNP. Independent experiments also identified two smaller members of the ring in the material purified by the Biot-αU7 oligonucleotide: *Drosophila* Lsm10 and SmD3.

The material purified via the Biot-αU7 contained large amounts of two major forms of the *Drosophila* orthologue of Insulin-like Growth Factor 2 mRNA Binding Protein 1 (IGF2BP1) and several of its apparent degradation products. The high abundance of IGF2BP1 greatly exceeded the amount of Lsm11 and other core U7 snRNP components in the precipitate,

suggesting that this protein is not part of the U7 snRNP. Subsequent studies and multiple immunoprecipitation experiments with the anti-FLASH antibody support the notion that IGF2BP1 directly binds the excess Biot-αU7 oligonucleotide and plays no role in 3' end processing of histone pre-mRNAs. Lane 1 also contained a band migrating at ~60 kDa that was identified as a *Drosophila* homolog of the mammalian p62 helicase (CG10279). This protein was also present in the negative lane, and hence it likely nonspecifically binds 2′-O-methyl oligonucleotides. Altogether, the results of both immunoprecipitation and affinity purification experiments combined with mass spectrometry demonstrate that *Drosophila* U7 snRNP has a composite structure, and in addition to the core components of its Sm ring contains FLASH and at least six polyadenylation factors.

Characterization of processing complexes formed on histone pre-mRNA in *Drosophila* nuclear extracts. I next tested whether processing complexes formed on *Drosophila* pre-mRNA in *Drosophila* nuclear extracts contain the same subset of polyadenylation factors that associate with the *Drosophila* U7 snRNP. A 63-nt dH3 pre-RNA containing all necessary processing elements (the SL and HDE) and biotin at the 5' end (Biot-dH3) was incubated for 20 min at room temperature with a Kc nuclear extract (750 µL), and factors that bound to this substrate were collected on streptavidin beads. To prevent cleavage, which would result in release of the U7 snRNP from the complexes, the reaction was supplemented with 0.1% NP-40. This mild detergent at low concentrations blocks processing in *Drosophila* nuclear extracts (Fig. 19A, lanes 3, 4), as it does in mammalian nuclear extracts (Yang et al., 2009b, 2013). However, processing complexes that assembled in the Kc nuclear extract under these conditions, while containing FLASH, SLBP, and U7 snRNP (as determined by the presence of Lsm11), lacked detectable amounts of polyadenylation factors. Purification of *Drosophila* U7 snRNP by either the Biot-

αU7 oligonucleotide (Fig. 19B, lane 3) or anti-FLASH antibody in the presence of NP-40 demonstrated that this detergent destabilized the interaction between U7 snRNP and polyadenylation factors, including symplekin (Fig. 19B, compare lanes 2 and 3). This is the most likely cause of the toxicity of NP-40 on 3' end processing in both *Drosophila* and mammalian nuclear extracts.

Instead of using NP-40, I significantly shortened the incubation time to capture processing complexes at early stages of assembly prior to cleavage of the substrate, in a manner similar to methods previously used for canonical pre-mRNAs containing the AAUAAA sequence (Veraldi et al., 2000; Shi et al., 2009). After 4–5 min of incubation in a nuclear extract, the Biot-dH3 pre-mRNA associated with both SLBP and the U7 snRNP, as determined by the presence of Lsm11 (Fig. 19C, lane 3). Thus, at least a fraction of the input RNA remains unprocessed under these conditions. The Biot-SL and the Biot-αU7 oligonucleotides bound only to SLBP and U7 snRNP, respectively (Fig. 19C, lanes 4, 5), demonstrating that binding of the two factors to sites flanking the cleavage site in the Biot-dH3 is specific. We next carried out larger-scale experiments, either in the absence or in the presence of various RNA competitors (Fig. 19D). Western blots with available antibodies against various processing factors demonstrated that the Biot-dH3 pre-mRNA in the absence of RNA competitors assembled into a complex containing SLBP, FLASH, the U7 snRNP (as determined by the presence of Lsm11), and the three essential polyadenylation factors we assayed: CPSF73, CPSF100, and symplekin (Fig. 19D, lane 3). The presence of 2.5 μM αU7 oligonucleotide, which blocks the 5' end of the U7 snRNA and abolishes the recruitment of the U7 snRNP to the substrate, eliminated all these proteins except SLBP from the complex (Fig. 19D, lane 6). Thus, the composite U7 snRNP

carrying FLASH and polyadenylation factors is directly recruited to histone pre-mRNA through the base-pairing interaction between the U7 snRNA and the HDE.

In the presence of 1 μM WT SL RNA that sequesters SLBP, binding of SLBP to the Biot-dH3 pre-mRNA was eliminated. However, the association of the composite U7 snRNP, including FLASH and the three polyadenylation factors, with the HDE was unaffected (Fig. 19D, lane 4). Note that this concentration of the WT SL competitor is sufficient to block the cleavage reaction (Fig. 12C, lane 3), whereas the Mut SL RNA (Fig. 12E) at the same and higher concentrations has no effect on 3' end processing (Fig. 12C, lane 5) and does not prevent binding of SLBP to the Biot-dH3 pre-mRNA (Fig. 19D, lane 5). The failure of the WT SL competitor to reduce the amount of the U7 snRNP in the complex assembled on the dH3 pre-mRNA was unexpected since previous models based on processing in mammalian and *Drosophila* nuclear extracts suggested that SLBP primarily functions by stabilizing the interaction of the U7 snRNP with the HDE in histone pre-mRNAs (Dominski et al., 1999, 2005b).

To confirm this surprising result by a different approach, I assembled processing complexes on a derivative of the Biot-dH3 pre-mRNA that lacked 16 nt encompassing the stem–loop structure. The HDE of this new substrate designated Biot-dH3ΔSL was identical to that in the Biot-dH3 pre-mRNA. In addition to the dH3ΔSL RNA, we also used the mouse-specific Biot-H2a-614 pre-mRNA (Yang et al., 2009b). This substrate contains the stem–loop structure that normally interacts with *Drosophila* SLBP, but its HDE has only a limited ability to base-pair with the *Drosophila* U7 snRNA. Due to this reduced compatibility between the HDE and the *Drosophila* U7 snRNP, the H2a-614 pre-mRNA is only poorly processed in *Drosophila* nuclear extracts (Dominski et al., 2002, 2005b).

As expected, the Biot-dH3ΔSL substrate, in contrast to the Biot-dH3 and Biot-H2a-614 pre-mRNAs, was unable to bind SLBP from the Kc nuclear extract (Fig. 19E, compare lane 4 with lanes 2 and 3). Importantly, consistent with our conclusion based on using the stem–loop competitor, the absence of the stem–loop structure in the Biot-dH3ΔSL substrate and SLBP in the processing complex did not prevent the interaction between the composite U7 snRNP and the HDE. Both FLASH and symplekin were detected in processing complexes assembled on the Biot-dH3ΔSL, and their amount was comparable to that detected in the complexes assembled on the Biot-dH3 pre-mRNA (Fig. 19E, compare lanes 3 and 4). The mouse-specific Biot-H2a-614 pre-mRNA interacted with the *Drosophila* U7 snRNP only weakly, consistent with the nature of the HDE in this substrate and its inability to form a strong duplex with the U7 snRNA (Fig. 19E, lane 2). I conclude that the interaction of the *Drosophila* SLBP with the stem–loop structure in histone pre-mRNA has no major effect on the recruitment of the U7 snRNP to the HDE.

Preparative scale purification (5.0 mL of Kc nuclear extract per each reaction) was next performed to analyze proteins associated with the Biot-dH3 pre-mRNA by mass spectrometry. The control reaction contained no RNA competitor, and the two other reactions contained either the αU7 oligonucleotide (1.25 μM) to block the *Drosophila* U7 snRNP or the WT SL RNA (1 μM) to sequester SLBP in the extract (Fig. 19F). On silver-stained gels, the three samples shared several bands, and most of these likely represented proteins nonspecifically bound to streptavidin beads and/or excess of the Biot-dH3 RNA. Compared with the control reaction (Fig. 19E, lane 1), the WT SL RNA competitor resulted in a disappearance of only one major protein band that corresponds to SLBP (Fig. 19F, lane 3). The reaction supplemented with the αU7 oligonucleotide lacked several proteins migrating in the range of ~50–200 kDa, whereas the amount of SLBP was unchanged (Fig. 19F, lane 2), consistent with the Western data. Mass

spectrometry identified these proteins (from the largest to the smallest) as CPSF160, FLASH, symplekin, CPSF100, WDR33, and CstF64 (Fig. 6F, right panel). Lanes 1 and 3, but not lane 2, also contained CPSF73 (band 7), although this polyadenylation factor was masked by an abundant protein present in all three lanes.

The material purified from the Kc nuclear extract via the Biot-dH3 pre-mRNA also contained CstF77 (CG17170) (Fig. 19F, lane 1, band 6). Results from mass spectrometry analysis suggest that the association of this protein with the Biot-dH3 substrate was only partially reduced by the presence of the αU7 oligonucleotide (Fig. 19F, right panel). CstF77 was not detected in the material purified by the anti-FLASH antibody or the Biot-αU7 oligonucleotide (see above), collectively arguing that this polyadenylation factor associates with the RNA substrate in a nonspecific manner. We also note that CstF77 was previously identified as a contaminant of the mammalian U7 snRNP affinity-purified from mouse nuclear extracts (Yang et al., 2013).

Mass spectrometry analysis of the remaining gel sections failed to identify any other major differences in protein composition between the three lanes. I conclude that the composite U7 snRNP containing FLASH and several polyadenylation factors is recruited to *Drosophila* histone pre-mRNA for 3' end processing. Strikingly, SLBP does not play any detectable role in the recruitment of this composite U7 snRNP, although its binding to the stem–loop in histone pre-mRNA is absolutely required for cleavage.

CstF64 as a component of the *Drosophila* U7 snRNP. The mammalian HCC that interacts with FLASH and Lsm11 is composed of symplekin, all CPSF subunits, and CstF64 (Yang et al., 2013). The two remaining CStF subunits, CstF50 and CstF77, are conspicuously missing from

the HCC. The equivalent complex in *Drosophila* also contains symplekin, most CPSF subunits, and CstF64 as the only component of the CStF subcomplex.

To confirm that CstF64 is part of the *Drosophila* U7 snRNP and to look more closely at its role in 3' end processing of histone pre-mRNAs in *Drosophila*, Xiao Yang generated antibodies against bacterially expressed full-length protein (amino acids 1–437). The antibody recognizes a distinct band on Western blots of *Drosophila* nuclear extract that migrates at ~50 kDa (Fig. 20A, lane 1), consistent with the mobility of CstF64 identified by mass spectrometry. I tested the material purified from Kc nuclear extracts via the Biot-αU7 oligonucleotide using the anti-CstF64 antibody and confirmed that it indeed contains CstF64 (Fig. 20A, lane 2). CstF64 was also readily detected in processing complexes assembled on the Biot-dH3 and Biot-dH3ΔSL RNAs. Lower amounts of CstF64 were associated with the Biot-mH2a pre-mRNA, consistent with the limited complementarity of this mouse-specific substrate to the *Drosophila* U7 snRNA. Thus, consistent with the mass spectrometry results, CstF64 is a stable component of the *Drosophila* U7 snRNP and is recruited together with other polyadenylation factors to histone pre-mRNA for 3' end processing.

I next tested the ability of the anti-CstF64 antibody to precipitate processing complexes from Kc nuclear extracts. Western blotting demonstrated that the anti-CstF64 antibody is very efficient in precipitating both CstF64 and CstF50 (Fig. 20B, lane 3), indicating that the antibody readily recognizes the entire CStF subcomplex known to consist of these three subunits in mammalian cells (Takagaki et al. 1990). Surprisingly, the anti-CstF64 precipitate also contained large amounts of CPSF73, CPSF100, and symplekin (Fig. 20B, compare lanes 1 and 3). Precipitation of all these polyadenylation factors was blocked by recombinant CstF64 used as the antigen to generate the antibody (Fig. 20B, lane 4). Importantly, the anti-CstF64 precipitate

contained readily detectable amounts of FLASH (both full-length and numerous degradation products) and Lsm11, but not Ars2 (Fig. 20C, lane 3), further confirming that in *Drosophila* nuclear extracts, a fraction of CstF64 exists in a complex with the U7 snRNP. Note that while at least 10% of Lsm11 was precipitated by the anti-CstF64 (Fig. 20C, lane 3), <1% of the total CstF64 was bound to U7 snRNP (Fig. 20A, lane 2), as expected from the very low amount of U7 snRNP compared with the high abundance of polyadenylation factors.

CstF64 as a component of a supercomplex of cleavage and polyadenylation factors. The ability of the anti-CstF64 antibody to precipitate a large fraction (~50%) of symplekin and various CPSF subunits (Fig. 20B) suggested that in *Drosophila* distinct polyadenylation subcomplexes may associate to form a macromolecular entity resembling that previously described in mammalian cells (Takagaki and Manley, 2000). To identify additional components of this hypothetical complex, I separated the precipitated proteins on either 8% (Fig. 20D, E) or 15% SDS/polyacrylamide gels (Fig. 20F) and analyzed their identity by mass spectrometry. The anti-CstF64 precipitate (Fig. 20D, lane 3) but not the control anti-Mock precipitate (Fig. 20D, lane 2) contained several proteins that were readily stained with silver. As determined by mass spectrometry, the precipitate contained all three subunits of CStF, although CstF64 and CstF50 were partially masked by the antibody heavy chain and hence not readily visible. I also identified all subunits of CPSF in the anti-CstF64 precipitate, including the smallest CPSF30 (CG3642) that migrates with the expected mobility of ~30 kDa (Fig. 20F, lane 1). Precipitation of all these proteins was eliminated by blocking the anti-CstF64 antibody with recombinant CstF64 (Fig. 20E, lane 2; Fig. 20F, lane 3). Among CPSF subunits, WDR33 and Fip1 (CG1078) were relatively weakly stained, which may indicate that they are limiting in *Drosophila* nuclear extracts (Fig. 20D, lane 3; Fig. 20E, lanes 1, 3). Note that Fip1 exhibits much slower

electrophoretic mobility than its mammalian counterpart (Kaufmann et al., 2004), consistent with its significantly larger size (701 amino acids vs. 550 amino acids for human Fip1).

The anti-CstF64 precipitate also contained significant amounts of the *Drosophila* orthologues of CF I_m68 (CG7185) (Fig. 20D, lane 3; Fig. 20E, lane 1) and CF I_m25 (CG3689) (Fig. 20F, lane 1). In mammalian cells, these two proteins belong to a third subcomplex, CI Im (Rüegsegger et al., 1996; Rüegsegger et al., 1998), which was not previously detected in a tight association with the subunits of CStF or CPSF in the absence of pre-mRNA substrate. Coprecipitation of the CF I_m was not affected by pre-treating the nuclear extract with RNase A (Fig. 20E, compare lanes 1 and 3; Fig. 20F, compare lanes 1 and 2).

To confirm this unexpected finding, I tested the precipitate with an antibody generated against the first 191 amino acids of human CF I_m68. This antibody cross-reacts with the *Drosophila* CF I_m68 (Fig. 20G), likely by recognizing the highly conserved N-terminal RRM-type RNA binding domain. Western blot analysis using this antibody confirmed that the anti-CstF64 antibody, indeed, coprecipitated CF I_m68 (Fig. 20H). Thus, *Drosophila* nuclear extracts contain a large, preassembled 3' end processing complex consisting of most known polyadenylation factors, with the notable exception of the *Drosophila* orthologues of the two subunits of CF II_m, Clp1 and Pcf11 (de Vries et al., 2000).

DISCUSSION

We recently showed that in mammalian nuclear extracts supplemented with the bacterially expressed N-terminal portion of FLASH, a subset of polyadenylation factors that we refer to as the Histone Cleavage Complex (HCC) associates with the U7 snRNP. The association depends

on the interaction between FLASH and Lsm11, the largest protein of the U7-specific Sm ring (Yang et al., 2013). The mammalian HCC contains symplekin, all six CPSF subunits, and CstF64, but lacks the two remaining CStF subunits. This composite U7 snRNP, containing the CPSF73 endonuclease among other components of the HCC, is subsequently recruited to the Histone Downstream Element (HDE) in histone pre-mRNA for the cleavage reaction (Yang et al., 2013). FLASH is present at a very low level in mammalian cells and is highly susceptible to proteolysis during preparation of nuclear extracts, likely explaining the failure of initial studies to detect its association with U7 snRNP and polyadenylation factors (Smith et al., 1991; Pillai et al., 2001).

FLASH and U7 snRNP are relatively abundant in a broad range of *Drosophila* nuclear extracts and likely not limiting for 3' end processing of histone pre-mRNAs *in vitro*. In this study, I took advantage of this fact to partially purify the U7 snRNP from the *Drosophila* Kc nuclear extracts and directly analyzed its proteome by mass spectrometry. My main goal was to determine whether *Drosophila* U7 snRNP associates with the same set of polyadenylation factors that form the HCC in mammalian nuclear extracts and whether this association depends on the interaction between *Drosophila* FLASH and Lsm11.

***Drosophila* U7 snRNP has a composite structure.** I affinity-purified the *Drosophila* U7 snRNP from Kc nuclear extracts via a 2′-O-methyl oligonucleotide complementary to the 5' end of the U7 snRNA. The purified material, in addition to Lsm11 and other proteins of the U7-specifc ring, contained readily detectable amounts of endogenous FLASH, confirming that this protein indeed constitutes an integral component of the U7 snRNP. The purified U7 snRNP also contained several polyadenylation factors. Among them I identified symplekin, four out of six CPSF subunits (CPSF160, CPSF100, CPSF73, and WDR33), and CstF64 as the only component

of the CStF subcomplex. I confirmed that CstF64 is indeed a component of the U7 snRNP in *Drosophila* by demonstrating that an antibody directed against full-length CstF64 efficiently coprecipitates both Lsm11 and FLASH.

Consistent with FLASH being an integral component of U7 snRNP, the anti-FLASH antibody precipitates Lsm11, and the antibody directed against Lsm11 precipitates FLASH. Importantly, the anti-FLASH precipitate contained the same polyadenylation factors that were identified in the affinity-purified U7 snRNP fraction. Thus, based on using two independent purification protocols, I conclude that the *Drosophila* U7 snRNP has a composite structure, consisting of the U7 snRNP core (U7 snRNA and the Sm ring), FLASH, and the *Drosophila* counterpart of the HCC. Compared with the mammalian HCC, the *Drosophila* HCC seems to lack orthologues of two CPSF components: Fip1 and CPSF30. My immunoprecipitation experiments with an anti-CstF64 antibody indicate that Fip1 may be present in low levels in *Drosophila* nuclear extracts, and it is possible that this subunit is also underrepresented in the HCC and escaped our detection. CPSF30, on the other hand, is readily detectable as a component of the *Drosophila* CPSF complex involved in cleavage and polyadenylation, and its absence in the HCC is surprising. Interestingly, RNAi-mediated knockdown of *Drosophila* CPSF30 in S2 cells in contrast to knockdown of CPSF160 does not result in codepletion of CPSF73, CPSF100, and symplekin, suggesting that CPSF30 may not be an essential component of CPSF in *Drosophila* cells, explaining the absence of this subunit in the HCC (Sullivan et al., 2009b).

I found no evidence for an association of the *Drosophila* U7 snRNP with the *Drosophila* orthologue of CF Im68. CF Im68 was recently reported to stably associate with the mammalian U7 snRNP (Ruepp et al., 2010), although previous results did not find CF Im68 as part of the mammalian HCC and the endogenous U7 snRNP (Yang et al., 2013) or the highly related Heat

Labile Factor (HLF) purified from mammalian cells (Kolev and Steitz, 2005). Similarly, this study argues that CF Im68 has no role in 3' end processing of histone pre-mRNAs in *Drosophila* nuclear extracts.

The composition of the HCC is much different from the composition of a related *Drosophila* complex that functions in cleavage and polyadenylation. This complex consists of symplekin, all three CStF subunits, all six CPSF subunits, and the two CF Im subunits of 25 and 68 kDa. The supercomplex was purified by precipitation with the anti-CstF64 antibody and is not disrupted by treatment with RNase A, indicating that its assembly involves multiple protein–protein interactions rather than binding of individual subcomplexes (CStF, CPSF, and CF Im) to a single pre-mRNA molecule. A similar complex was previously described in mammalian nuclear extracts (Takagaki and Manley, 2000).

In 3' end processing of mammalian histone pre-mRNAs, a critical interaction between the HCC and FLASH bound to Lsm11 is mediated by the LDLY motif located in the N-terminal region of FLASH (Yang et al., 2013). Recombinant N-terminal FLASH lacking this motif is unable to interact with the HCC or to stimulate 3' end processing in mammalian nuclear extracts. It is yet unknown which component of the mammalian HCC directly contacts the LDLY motif, although symplekin is the most likely candidate to perform this function. *Drosophila* FLASH contains a highly related motif, LDIY, within its N-terminal region (amino acids 71–74), and our previous studies demonstrated that deleting the first 77 amino acids impairs the wild-type function of FLASH in processing in *Drosophila* cultured cells (Burch et al., 2011). I now show that recombinant N-terminal FLASH with the LDIY motif being replaced by four alanines is unable to restore processing activity of nuclear extracts prepared from S2 cells depleted of the endogenous FLASH. I also showed that Lsm11 and this LDIY mutant form a complex that is

unable to interact with CPSF73, CPSF100, and symplekin. Thus, the mammalian LDLY and *Drosophila* LDIY motifs are functionally conserved and play a critical role in bringing the HCC to the U7 snRNP.

Degradation of *Drosophila* DCP is FLASH dependent. Following cleavage to form the 3' end of histone mRNA, the downstream cleavage product is degraded by the 5' to 3' exonucleolytic activity of CPSF73. Degradation is U7-dependent (Yang et al. 2009), but in mammals, FLASH is dispensable for this reaction (Yang et al., 2011). Here I present data that shows FLASH is essential for the degradation of the DCP in *Drosophila*. In *Drosophila*, histone pre-mRNA 3' end processing always results in cleavage 4 nts after the stem-loop. Repositioning the HDE a distance further 3' of its usual position results in a loss of processing. In contrast, mammalian cleavage always occurs a fixed distance 5' of the HDE, regardless of its position if the duplex formed with the HDE is sufficiently strong. Interactions with the U7 snRNP may rigidify and position the cleavage site for processing, independent of contributions from the stem-loop or SLBP. These same interactions may position the 5' end of the DCP optimally for CPSF73 exonuclease activity in mammals. Degradation in *Drosophila* could require additional factors to juxtapose the DCP 5' end and CPSF73 for initiation. FLASH may fulfill this crucial function.

Ars2 is dispensable for U7-dependent processing. Previous studies that used a sensitive GFP-based reporter gene driven by a histone gene promoter (Wagner et al., 2007) and direct analysis of histone mRNA 3′ ends (Sullivan et al., 2009b) indicated that only three polyadenylation factors are essential for 3' end processing of histone pre-mRNAs in *Drosophila* cells: (1) CPSF73, which functions as the endonuclease in cleavage and polyadenylation and 3' end processing of histone pre-mRNAs; (2) CPSF100, which is a homolog of CPSF73; and (3)

symplekin, which plays an unknown role in processing but likely functions as a scaffold or adaptor linking CPSF73 and CPSF100 with the rest of each processing machinery.

The lack of any essential role in 3' end processing of *Drosophila* histone pre-mRNAs for CstF64 is surprising. CstF64 directly interacts with symplekin (Takagaki et al., 1990) and plays an important role in 3' end processing of histone pre-mRNAs in human cells (Ruepp et al., 2011). In addition, our recent studies in mammalian nuclear extracts suggested that CstF64 in conjunction with symplekin may directly link the HCC with FLASH bound to Lsm11 (Yang et al., 2013). This is further evidence that U7-dependent processing machineries in mammalian and *Drosophila* cells are not identical (Fig. 21)

Direct analysis of endogenous histone mRNAs in Ars2-depleted S2 cells did not reveal any role for Ars2 in 3' end processing of *Drosophila* histone pre-mRNAs *in vivo*. In the absence of Ars2, the U7-dependent processing is fully operative, and no replication-dependent histone mRNAs with a poly(A) tail accumulate. This is in sharp contrast to depletion of Ars2 in human cells, which resulted in redirecting a fraction of nascent histone pre-mRNAs to the cleavage and polyadenylation pathway (Gruber et al., 2012). The lack of any major role in 3' end processing of histone pre-mRNAs for Ars2 in *Drosophila* is supported by our biochemical studies, which failed to detect this protein in association with U7 snRNP or any other processing factors known to operate on histone pre-mRNA. Overall, our results suggest that the interaction between Ars2 and FLASH and the role of Ars2 in promoting correct 3' end cleavage of histone pre-mRNAs are specific for higher metazoans (Fig. 21).

U7 snRNP as an RNA-guided nuclease and the role of SLBP in processing. The composite *Drosophila* U7 snRNP containing FLASH and multiple polyadenylation factors is recruited to histone pre-mRNA for 3' end processing, resembling the situation recently described for

mammalian extracts (Yang et al., 2013). An antisense oligonucleotide complementary to the 5' end of the *Drosophila* U7 snRNA that prevents the interaction between the U7 snRNP and the HDE completely eliminates FLASH and the polyadenylation factors from the complex formed on histone pre-mRNA. Thus, the U7 snRNA can be considered as a vehicle that identifies a target substrate by base-pairing and delivers the endonuclease CPSF73, CPSF100, symplekin, and several accessory polyadenylation factors to the site of cleavage. This mode of action is highly reminiscent of snoRNAs, which use an RNA-guided mechanism to locate target sequences for enzymatic modifications (Kiss et al., 2002).

In purified fractions of the U7 snRNP and processing complexes, FLASH stains with silver significantly stronger than the polyadenylation factors. It is therefore possible that a fraction of the U7 snRNP contains FLASH, but not the HCC (Fig. 22). Alternatively, more than one molecule of FLASH is required to interact with Lsm11, as suggested (Yang et al., 2013), or binding of the HCC to the Lsm11/FLASH complex is relatively weak and partially disrupted during purification of the U7 snRNP.

The most notable difference between the mechanism of 3' end processing of histone pre-mRNAs in *Drosophila* and mammals is that *Drosophila* processing is absolutely dependent on SLBP (Dominski et al., 2002), and cleavage invariably occurs 4 nt after the stem (Fig. 21) (Dominski et al., 2005b). These observations suggest that in *Drosophila*, SLBP and U7 snRNP are equally important in determining the site of cleavage. In mammalian processing, SLBP is an auxiliary rather than an obligatory factor and functions to stabilize the interaction between histone pre-mRNA and the U7 snRNP, which is capable of directing cleavage of the pre-mRNA on its own (Streit et al., 1993; Dominski et al., 1999). Consistently, in mammalian nuclear

extracts, the site of cleavage is determined by the U7 snRNP (Scharl and Steitz, 1994; Scharl and Steitz, 1996).

An excess of the stem–loop RNA added to *Drosophila* nuclear extracts results in a complete inhibition of *in vitro* processing due to preventing the interaction between SLBP and the stem–loop in histone pre-mRNAs (Dominski et al., 2002). Strikingly, this treatment had no effect on the recruitment of the composite *Drosophila* U7 snRNP to histone pre-mRNA. The core U7 snRNP, FLASH, and all the polyadenylation factors that form the HCC, including the CPSF73 endonuclease, are present in the processing complex assembled on histone pre-mRNA in the absence of free SLBP or on a pre-mRNA that lacks an SLBP binding site. Thus, in contrast to a previous model (Dominski et al., 2005b), the most critical role in recruiting the composite U7 snRNP to *Drosophila* histone pre-mRNA is played by the sequence of the HDE. SLBP (and hence the stem–loop) is yet indispensable for processing in *Drosophila* nuclear extracts and may function by delivering an unknown essential factor(s) to the U7 snRNP positioned near the cleavage site.

Mass spectrometry did not detect any major differences in the composition of processing complexes assembled either in the presence or the absence of *Drosophila* SLBP. One possibility is that the SLBP-interacting factor exists in *Drosophila* nuclear extracts at a very low concentration or its interaction with SLBP is weak or transient and cannot be detected biochemically. What could be the nature and function of this factor? It may, for example, represent a hypothetical activator of the CPSF73 endonuclease, which was speculated to depend

on a structural rearrangement prior to acquiring the catalytic activity (Mandel et al., 2006; Dominski, 2010). Alternatively, it may function by bypassing the requirement for CstF64 or Ars2 in 3' end processing in *Drosophila*. Elucidating the precise role of *Drosophila* SLBP in activating cleavage of histone pre-mRNA is the main focus of our ongoing studies.

Figure 12. 3' end processing of histone pre-mRNAs in *Drosophila.* (A) Known factors involved in 3' end processing of histone pre-mRNAs in *Drosophila*. (Black and gray lines) Histone pre-mRNA and *Drosophila* U7 snRNA, respectively. (Vertical lines) Base-pair interactions between the two RNAs. In the U7-specifc Sm ring, the two spliceosomal proteins SmD1 and SmD2 are replaced by Lsm10 and Lsm11 (indicated with dark gray). Lsm11 binds FLASH and in mammalian nuclear extracts these two proteins interact with CPSF73 (the endonuclease), CPSF100, symplekin, and other polyadenylation factors that together form the Histone pre-mRNA Cleavage Complex (HCC). It is unknown whether the interaction between Lsm11 and FLASH functions in the same manner in *Drosophila* (double arrow) and whether the HCC contains any factors in addition to CPSF73, CPSF100, and symplekin, as indicated with the question mark. (B) Levels of FLASH, SLBP, and symplekin in nuclear extract prepared from frozen (lane 1) or freshly collected Kc cells (lane 2), as determined by Western blotting. (C) 3' end processing of *Drosophila* histone H3 (dH3) pre-mRNA in a Kc nuclear extract prepared from freshly harvested cells in the presence of indicated RNA competitors (lanes 3–8). The input alone and processing in the absence of any RNA competitor is shown in lanes 1 and 2, respectively. (D) 3' end processing of the dH3 pre-mRNA in a Kc nuclear extract prepared from freshly harvested cells in the presence of indicated antibodies (lanes 3 and 4). The input alone and processing in the absence of any RNA competitor is shown in lanes 1 and 2, respectively. Inhibition of processing by the αU7 oligonucleotide is shown in lane 5. (E) Nucleotide sequence of the wild-type SL (WT SL) RNA. Regions altered in the Mut SL RNA that are essential for binding SLBP are underlined.

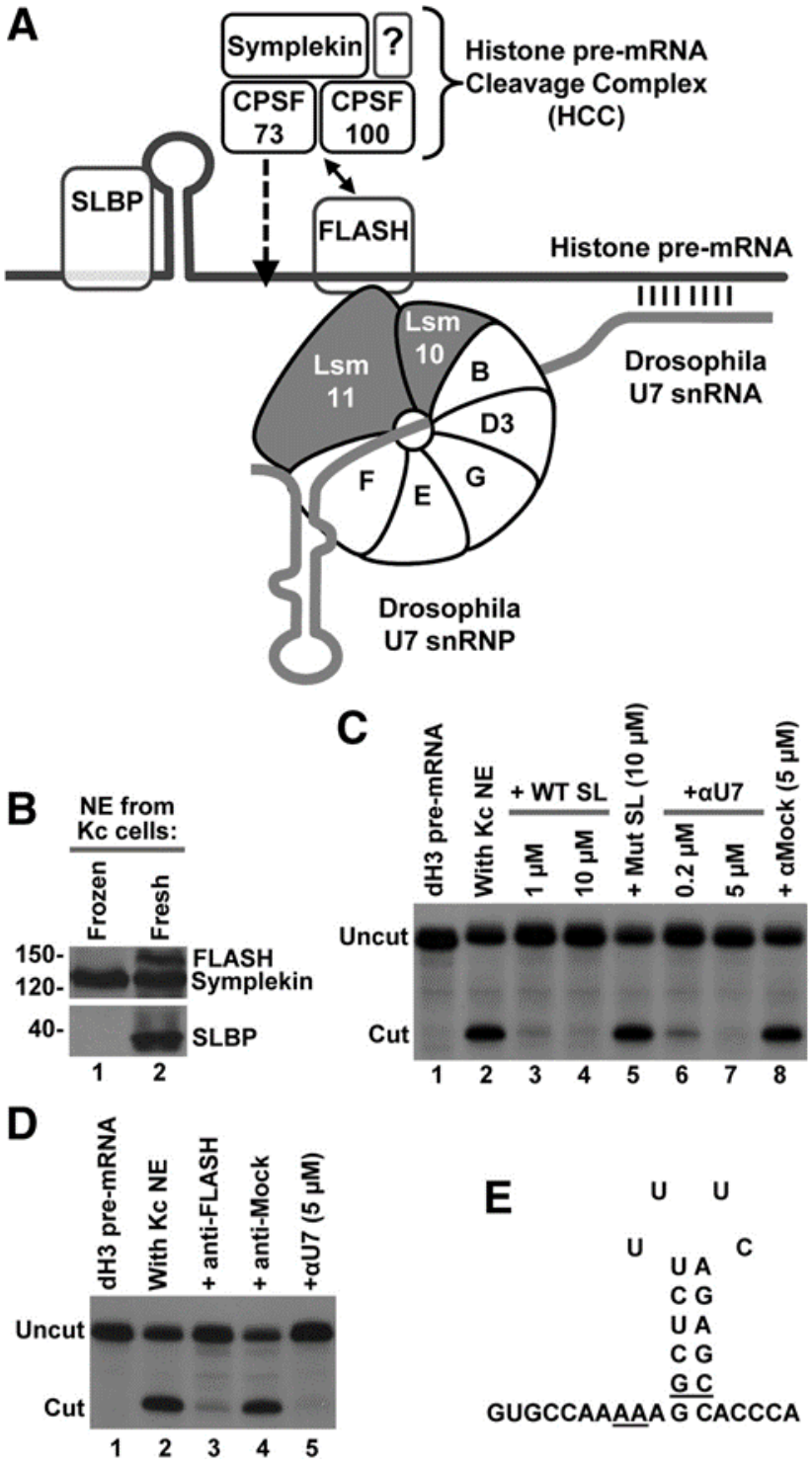

A
Symplekin
?
CPSF 73
CPSF 100
Histone pre-mRNA Cleavage Complex (HCC)
SLBP
FLASH
Histone pre-mRNA
Lsm 10
Lsm 11
B
D3
F
E
G
Drosophila U7 snRNA
Drosophila U7 snRNP
B
NE from Kc cells:
Frozen
Fresh
150-
120-
FLASH
Symplekin
40-
SLBP
1 2
C
dH3 pre-mRNA
With Kc NE
+ WT SL
1 µM
10 µM
+ Mut SL (10 µM)
+αU7
0.2 µM
5 µM
+ αMock (5 µM)
Uncut
Cut
1 2 3 4 5 6 7 8
D
dH3 pre-mRNA
With Kc NE
+ anti-FLASH
+ anti-Mock
+αU7 (5 µM)
Uncut
Cut
1 2 3 4 5
E
U U
U C
U A
C G
U A
C G
G C
GUGCCAAAAA G CACCCA

Figure 13. The role of the LDIY motif in FLASH in processing. (A) The amino acid sequence of the first 178 residues of *Drosophila* FLASH. The Lsm11-binding site and the critical LDIY motif (replaced with four alanines in the LDIY-4A mutant) are underlined. (Arrow) The start point of the FΔ77N deletion mutants. (B) 3' end processing of the dH3 pre-mRNA in a Kc nuclear extract in the presence of 100 ng of indicated recombinant proteins. The input pre-mRNA and control processing (no protein added) are shown in lanes 1 and 2, respectively. (C) The abundance of endogenous FLASH and SLBP in untreated S2 cells (No KD, lane 1) or treated with a specific double stranded (ds) RNA (FLASH KD, lane 2), as determined by Western blotting using an antibody directed against the N-terminal region of *Drosophila* FLASH. (*) A protein cross-reacting with the antibody that serves as a loading control. (D) 3' end processing of the dH3 pre-mRNA in nuclear extracts prepared from untreated S2 cells (lanes 2–4) or cells depleted of FLASH (lanes 5–7). Control processing in the absence of any recombinant protein is shown in lanes 2 and 5. Processing in lanes 3–4 and 6–7 was carried out in the presence of 100 ng of indicated recombinant proteins. (E) 3' end processing of the dH3 pre-mRNA in a nuclear extract prepared from S2 cells partially depleted of FLASH. Reactions were carried out in the presence of 100 ng of indicated recombinant FLASH proteins (lanes 2–4). The control reaction (lane 1) contains no exogenous protein. (F) Statistical analysis of 3' end processing of the dH3 pre-mRNA in small-scale nuclear extracts from untreated S2 cells (left panel) and FLASH-depleted S2 cells (right panel). Processing was carried out in the absence of any recombinant protein (buffer) or in the presence of indicated variants of FLASH. The P-value

was calculated using a paired version of the Student's t-test with a two-tailed distribution. The data represent two to four separate processing reactions and three independent nuclear extract preparations from untreated and treated S2 cells. Efficiency was measured as a ratio of the final processed product to the input (unprocessed + processed). Error bars indicate the standard deviation for measured efficiencies.

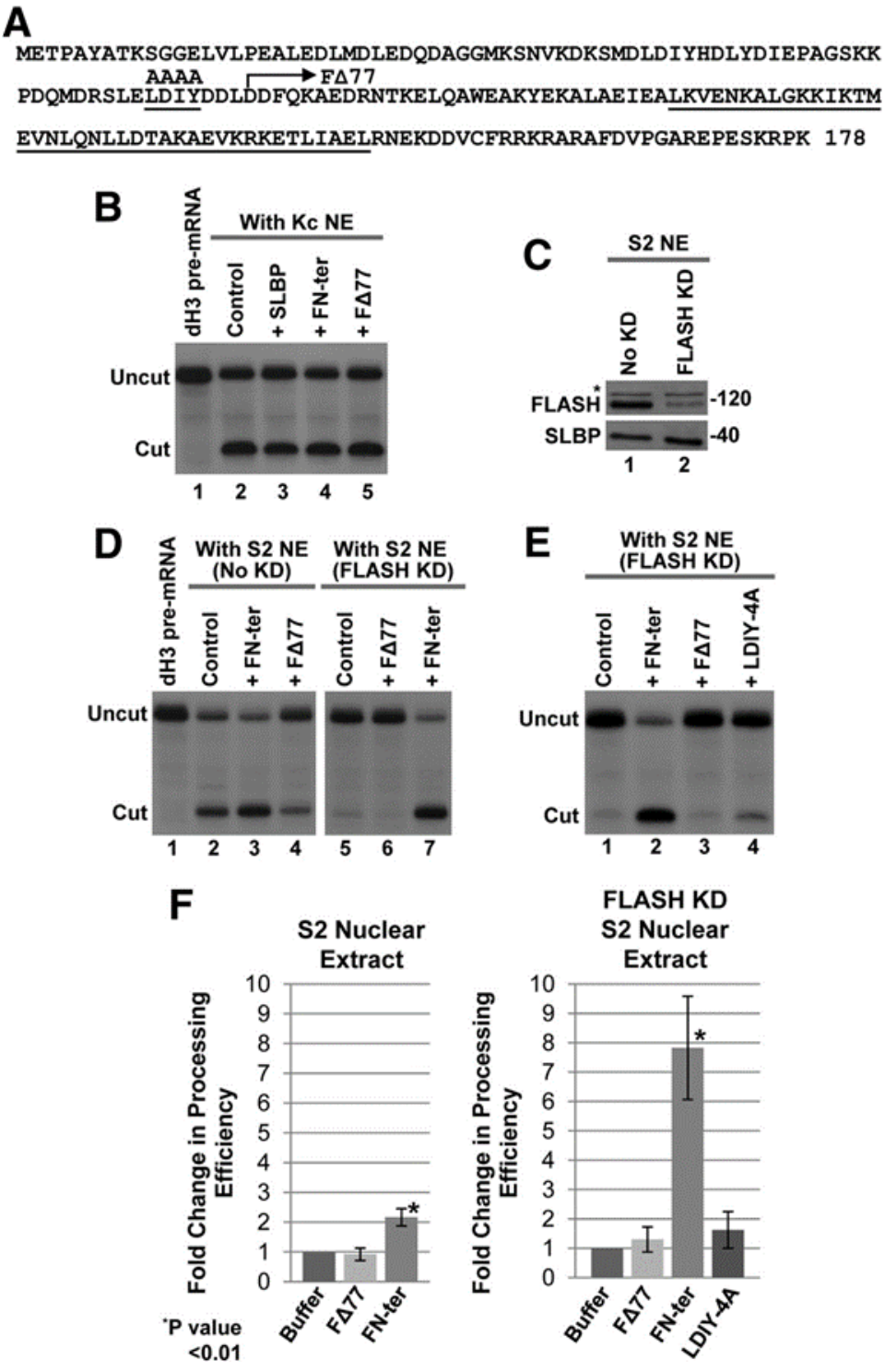

A
METPAYATKSGGELVLPEALEDLMDLEDQDAGGMKSNVKDKSMDLDIYHDLYDIEPAGSKK
AAAA
FΔ77
PDQMDRSLELDIYDDLDDFQKAEDRNTKELQAWEAKYEKALAEIEALKVENKALGKKIKTM
EVNLQNLLDTAKAEVKRKETLIAELRNEKDDVCFRRKRARAFDVPGAREPESKRPK 178
B
With Kc NE
dH3 pre-mRNA
Control
+ SLBP
+ FN-ter
+ FΔ77
Uncut
Cut
1 2 3 4 5
C
S2 NE
No KD
FLASH KD
FLASH*
SLBP
-120
-40
1 2
D
With S2 NE (No KD)
With S2 NE (FLASH KD)
dH3 pre-mRNA
Control
+ FN-ter
+ FΔ77
Control
+ FΔ77
+ FN-ter
Uncut
Cut
1 2 3 4 5 6 7
E
With S2 NE (FLASH KD)
Control
+ FN-ter
+ FΔ77
+ LDIY-4A
Uncut
Cut
1 2 3 4
F
S2 Nuclear Extract
Fold Change in Processing Efficiency
10
9
8
7
6
5
4
3
2
1
0
*
*P value <0.01
Buffer
FΔ77
FN-ter
FLASH KD S2 Nuclear Extract
Fold Change in Processing Efficiency
*
Buffer
FΔ77
FN-ter
LDIY-4A

Figure 14. Second conserved LDIY motif located between amino acids 45-48 has a minor role in processing efficiency *in vitro*. (A) The first 178 amino acids of *Drosophila* FLASH with two highly conserved LDIY motifs underlined. Both of these motifs were changed to alanines in a double FLASH mutant (LDIY-4A (x2)). (B) Western to identify dFLASH in nuclear extracts prepared from S2 small scale cell cultures either in the presence of dsRNA targeting dFLASH (lane 2) or without (lane 1). Arrow indicates a cross reacting band between the two nuclear extracts. (C) Time course for 3' end processing of labeled dH3 pre-mRNA in nuclear extracts prepared from untreated cells (lanes 2 and 3) or FLASH knockdown cells (lanes 4-11). Labeled dH3 in the absence of nuclear extract is in lane 1. Incubations were stopped at 90 or 180 minutes respectively. Approximately 100 ng of corresponding recombinant proteins or an equivalent volume of extract buffer were added to each reaction as indicated.

A LDIY-4A (x2)

AAAA

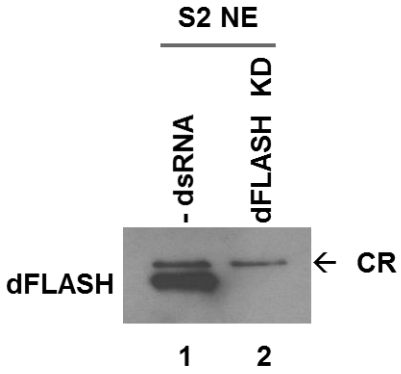
1 METPAYATKSGGELVLPEALEDLMDLEDQDAGGMKSNVKDKSMDLDIYHDLYDIEPAGSK

AAAA

KPDQMDRSLELDIYDDLDDFQKAEDRNTKELQAWEAKYEKALAEIEALKVENKALGKKIK

TMEVNLQNLLDTAKAEVKRKETLIAELRNEKDDVCFRRKR ARAFDVPGAREPESKRP 178

B



C

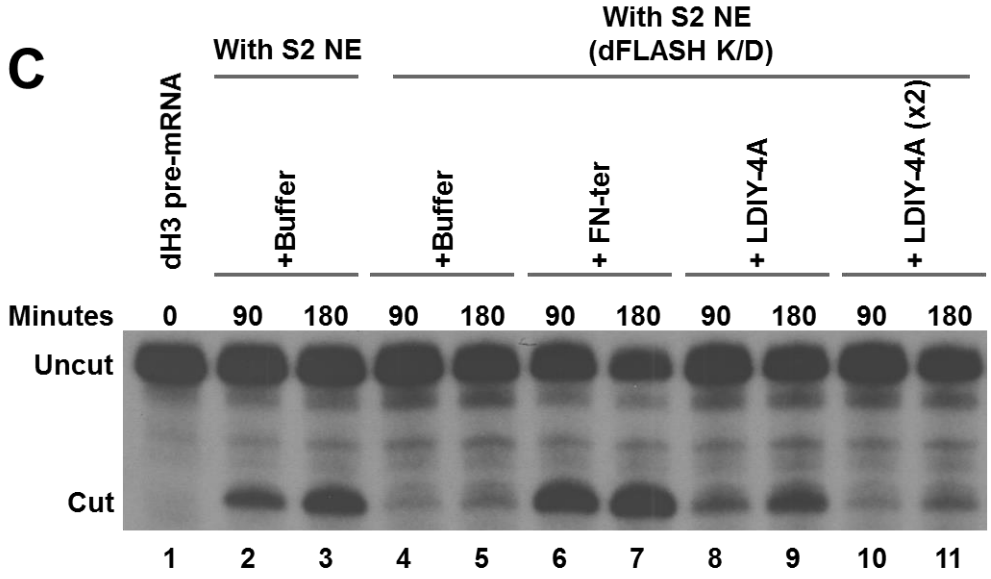


Figure 15. A complex of the N-terminal FLASH and Lsm11 binds *Drosophila* polyadenylation factors. (A–C) The interaction of indicated recombinant FLASH variants bound to the N-terminal Lsm11 with polyadenylation factors in S2 (A) or Kc (B,C) nuclear extracts. Nuclear proteins that interact with the Lsm11/FLASH complex were collected on glutathione (GSH) beads via the GST tag attached to FLASH and analyzed by Western blotting using specific antibodies. Material bound to GSH beads in the absence of recombinant proteins was analyzed in lanes 2 and 6 (A) and lane 2 (B,C). Input (15%) of the nuclear extract is shown in lane 1 of each panel. The arrows in C indicate a contaminant from the recombinant Lsm11 protein that cross-reacts with the antibody against CPSF73.

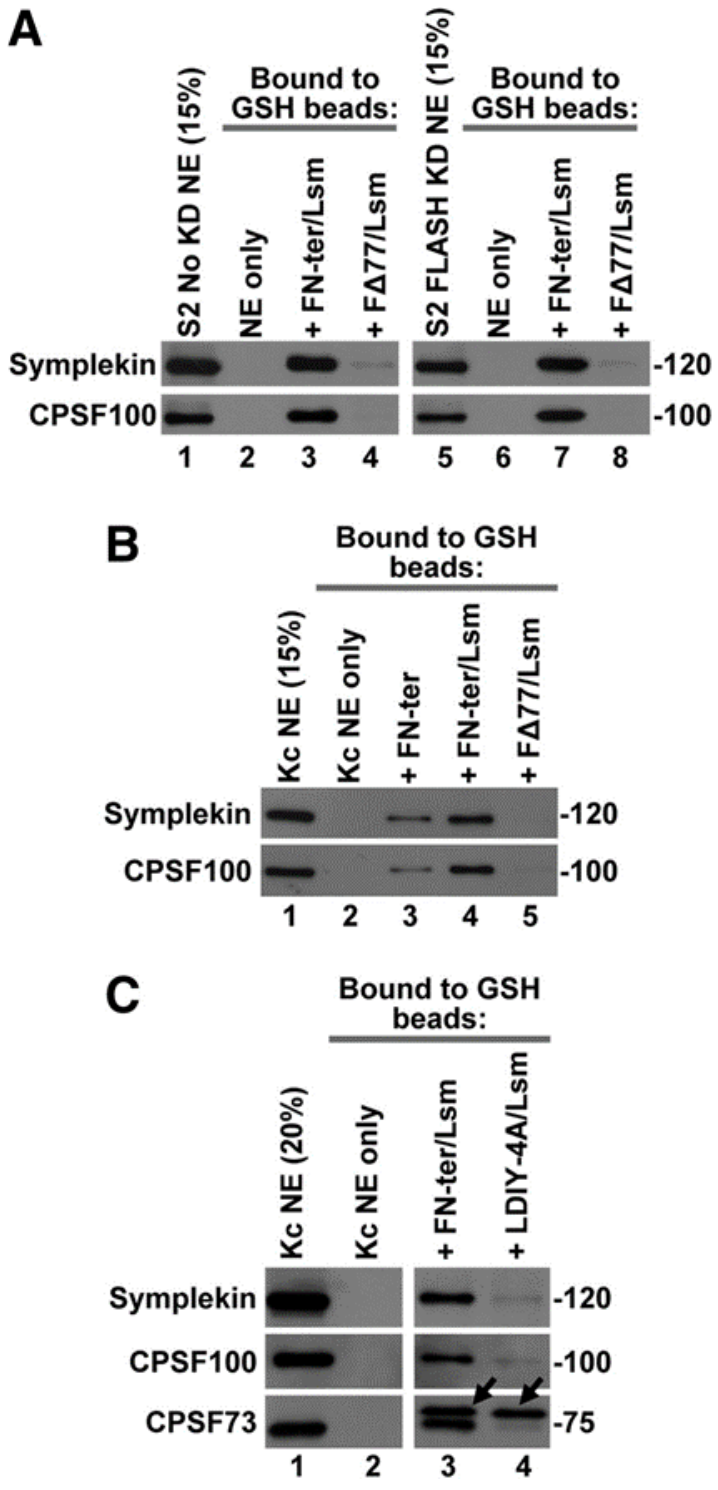

A
S2 No KD NE (15%)
Bound to GSH beads:
NE only
+ FN-ter/Lsm
+ FΔ77/Lsm
S2 FLASH KD NE (15%)
Bound to GSH beads:
NE only
+ FN-ter/Lsm
+ FΔ77/Lsm
Symplekin
CPSF100
-120
-100
1 2 3 4 5 6 7 8
B
Bound to GSH beads:
Kc NE (15%)
Kc NE only
+ FN-ter
+ FN-ter/Lsm
+ FΔ77/Lsm
Symplekin
CPSF100
-120
-100
1 2 3 4 5
C
Bound to GSH beads:
Kc NE (20%)
Kc NE only
+ FN-ter/Lsm
+ LDIY-4A/Lsm
Symplekin
CPSF100
CPSF73
-120
-100
-75
1 2 3 4

Figure 16. U7-dependednt degradation of *Drosophila* DCP generated during 3' end processing requires FLASH. *In vitro* degradation assays with 5' radio-labeled DCP in small scale *Drosophila* S2 nuclear extracts. Labeled DCP substrate at top of gel and degraded 5' monophosphate products at the bottom. Probe alone is in lane 1. Reactions with extracts from untreated S2 cells are shown in lanes 2-5. Lane 3 shows U7-dependence of degradation. Lanes 6-9 show reactions with extracts from cell culture treated by RNAi to deplete FLASH.

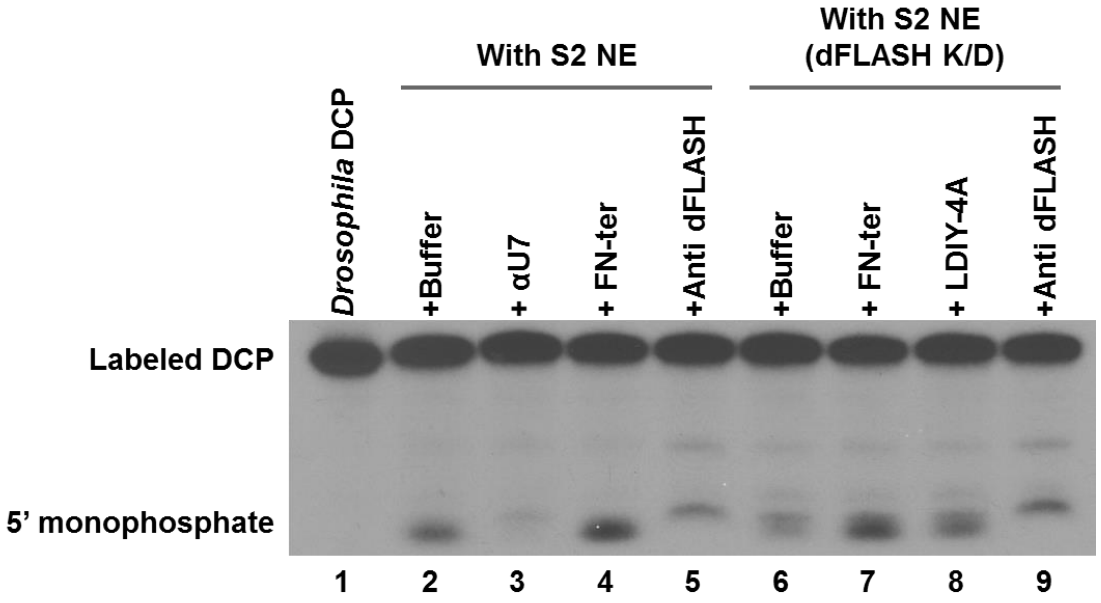

With S2 NE
With S2 NE
(dFLASH K/D)
Drosophila DCP
+Buffer
+ αU7
+ FN-ter
+Anti dFLASH
+Buffer
+ FN-ter
+ LDIY-4A
+Anti dFLASH
Labeled DCP
5’ monophosphate
1
2
3
4
5
6
7
8
9

Figure 17. Analyzing the involvement of *Drosophila* Ars2 in 3' end processing of histone pre-mRNAs. (A) Immunoprecipitation of Ars2 and FLASH from nuclear extracts prepared from freshly harvested Kc cells (lanes 1–5) or S2 cells grown in monolayers (MS, lanes 6–9). The antibodies used in the experiment are indicated at the top of each lane. The anti-Mock antibody is directed against 3′hExo, a human protein that has no clear homolog in *Drosophila*. The inputs are shown in lanes 1 and 6. (B) Proteins immunoprecipitated by the anti-Ars2 antibody from a Kc cell nuclear extract were detected by silver staining and analyzed by mass spectrometry. The major identified proteins are listed on the left. (C) Depletion of endogenous Ars2 in S2 cells by a specific dsRNA. S2 cells were treated for 5–6 d with dsRNA against various proteins and tested for depletion of Ars2 using the anti-Ars2 antibody. CR indicates a protein of ~70 kDa that cross-reacts with the antibody and served as a loading control. Mock1 (CG7769) and Mock2 (CG8443) are two proteins unrelated to 3' end processing of histone pre-mRNAs. (D) Northern blot analysis of endogenous histone H3 mRNA in S2 cells treated with dsRNAs shown in panel C. Species ending with the canonical stem–loop (dH3 SL) migrate as a single band. Polyadenylated species are longer and, due to the heterogeneity of the poly(A) tail, migrate as a smear. *Drosophila* 7SK RNA served as a loading control.

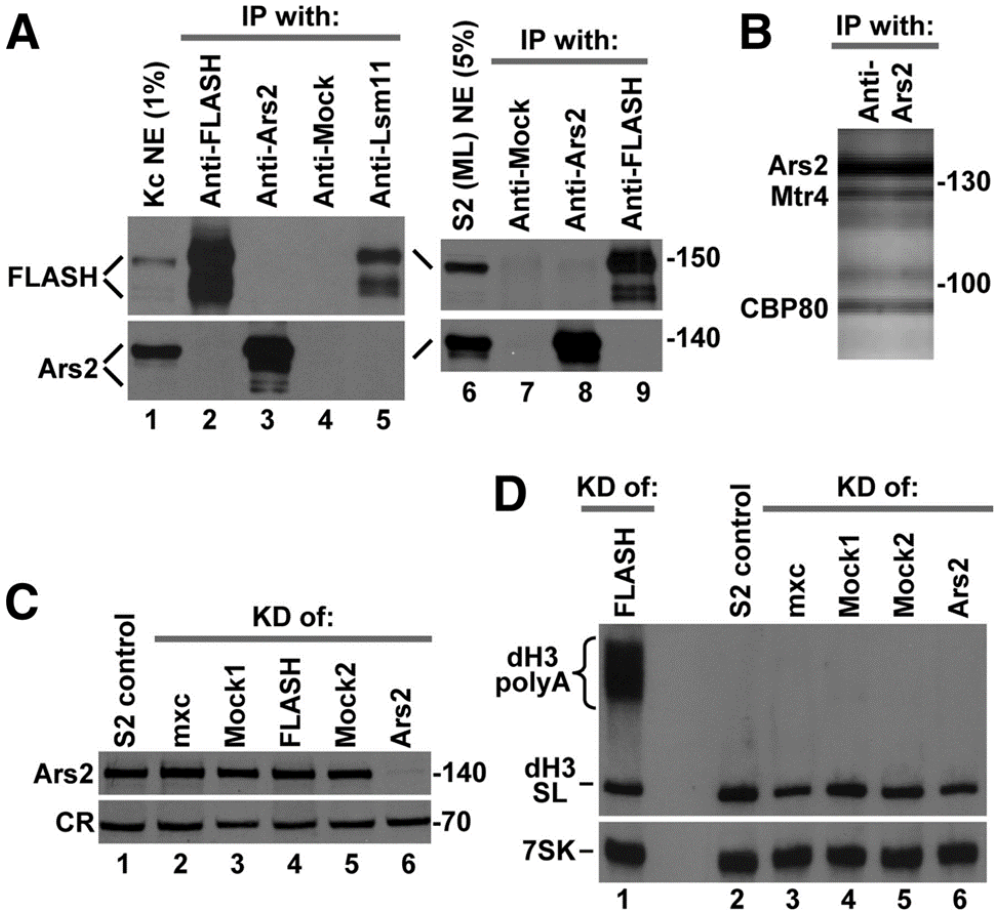
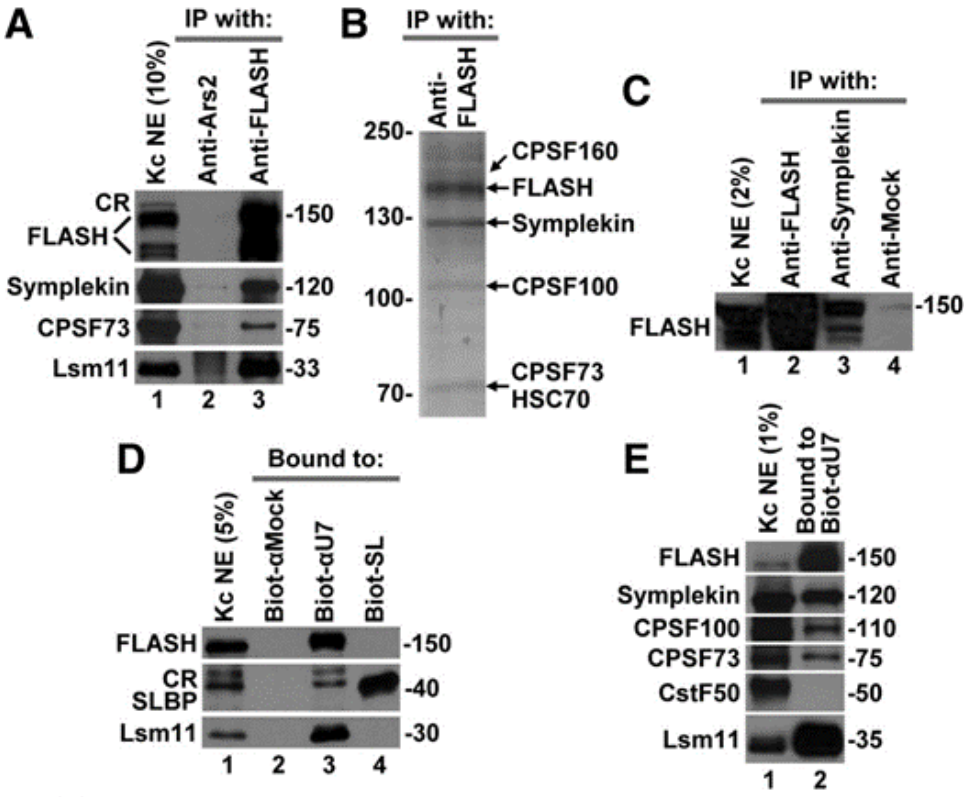
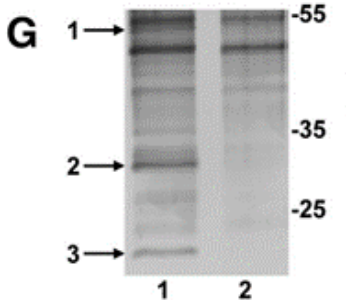

A
IP with:
Kc NE (1%)
Anti-FLASH
Anti-Ars2
Anti-Mock
Anti-Lsm11
S2 (ML) NE (5%)
IP with:
Anti-Mock
Anti-Ars2
Anti-FLASH
FLASH
Ars2
-150
-140
1 2 3 4 5 6 7 8 9
B
IP with:
Anti-Ars2
Ars2
Mtr4
CBP80
-130
-100
C
KD of:
S2 control
mxc
Mock1
FLASH
Mock2
Ars2
Ars2
CR
-140
-70
1 2 3 4 5 6
D
KD of:
FLASH
KD of:
S2 control
mxc
Mock1
Mock2
Ars2
dH3 polyA
dH3 SL
7SK
1 2 3 4 5 6

Figure 18. Composition of *Drosophila* U7 snRNP. (A) The material precipitated from a Kc nuclear extract with the anti-Ars2 (lane 2) or anti-FLASH (lane 3) antibodies was tested for the presence of indicated processing factors using specific antibodies. Lane 1 contains 10% of the input extract used for immunoprecipitation. CR indicates a protein cross-reacting with the anti-FLASH antibody. (B) Proteins precipitated with the anti-FLASH antibody from 2.0 mL of a Kc nuclear extracts were separated in an SDS/polyacrylamide gel, detected by silver staining, and identified by mass spectrometry. (C) Proteins immunoprecipitated from a Kc nuclear extract using indicated antibodies were analyzed for the presence of FLASH using the anti-FLASH antibody. Mock antibody (lane 4) was directed against a protein unrelated to 3' end processing of histone pre-mRNA (CG9958). Lane 1 contains 2% of the input extract used in immunoprecipitation. (D) Binding of processing factors to various RNAs containing biotin for subsequent purification on streptavidin Sepharose beads. Lane 1 contains 5% of the input extract used in each binding experiment. One of the blots was simultaneously probed with the anti-FLASH and anti-SLBP antibodies; CR indicates two FLASH degradation products that migrate immediately above full-length SLBP. SLBP is virtually undetectable in this lane and highly enriched in the material purified by the Biot-SL RNA. (E) The material bound to the Biot-αU7 oligonucleotide was analyzed by Western blotting for the presence of indicated proteins. Lane 1 contains 1% of the input extract used in the binding experiment. (F,G) Major proteins bound to either the Biot-αMock (lane 1) or Biot-αU7 (lane 2) oligonucleotides were separated in 8% (F) or 12% (G) SDS/polyacrylamide gels, stained with silver, identified by mass spectrometry, and are listed to the right.



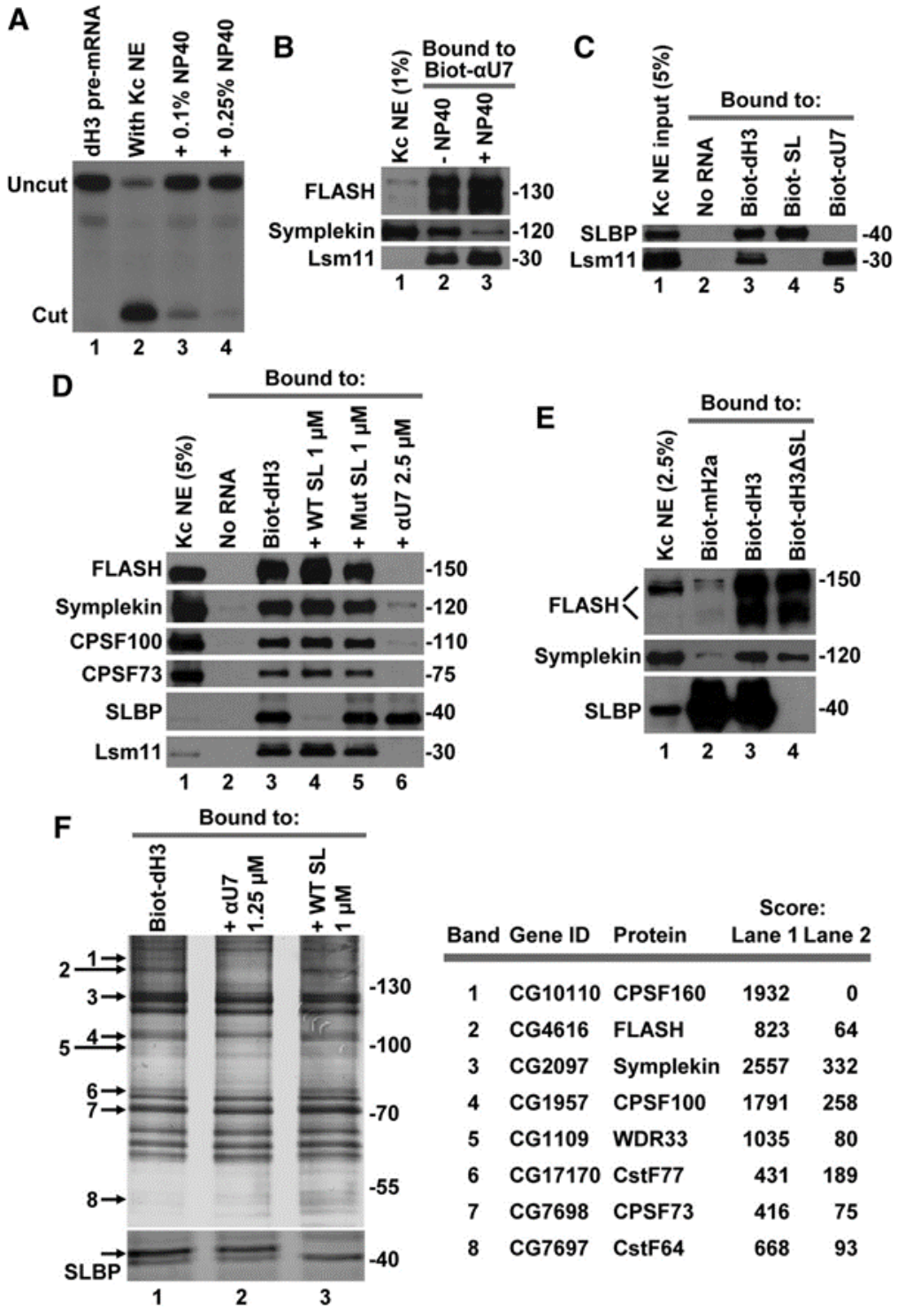
Band	Gene ID	Protein	Score: Lane 1	Lane 2
1	CG10120	CPSF160	199	0
2	CG4616	FLASH	687	0
3	CG2097	Symplekin	1187	0
4	CG1957	CPSF100	388	0
5	CG1109	WDR33	210	0
6	CG1691	IGF2BP1	1710	32
7	CG1691	IGF2BP1*	555	0
	CG7698	CPSF73	262	0
8	CG10279	Helicase p62	900	269
9	CG7697	CstF64	114	0



Band	Gene ID	Protein	Score: Lane 1	Lane 2
1	CG7697	CstF64	257	0
2	CG12924	Lsm11	575	0
3	CG5352	SmB1	127	39

Figure 19. Composition of processing complexes assembled on the dH3 pre-mRNA. (A) 3' end processing of the dH3 pre-mRNA in a Kc nuclear extract in the presence of indicated concentrations of NP40 (lanes 3, 4). The input alone and processing in the absence of NP-40 is shown in lanes 1 and 2, respectively. (B) The *Drosophila* U7 snRNP was affinity-purified from a Kc nuclear extract using the Biot-αU7 oligonucleotide either in the absence (lane 2) or presence (lane 3) of NP-40 and subsequently probed for the presence of indicated processing factors. Lane 1 contains 1% of the input extract used for the purification. Lsm11 is virtually undetectable in this lane and highly enriched in the material purified by the oligonucleotide. (C) Binding of SLBP and Lsm11 to indicated biotinylated RNAs. Lane 1 contains 5% of the input Kc extract used in each binding experiment. Lane 2 shows the background binding to streptavidin Sepharose beads in the absence of any biotinylated RNA. (D) Binding of processing factors from a Kc nuclear extract to the Biot-dH3 pre-mRNA (lanes 3–6), either in the absence of any RNA competitor (lane 3) or in the presence of indicated RNAs (lanes 4–6). Lane 1 contains 5% of the input Kc extract used in each binding experiment. SLBP and Lsm11 are virtually undetectable in this lane and highly enriched in the material purified by the Biot-dH3 pre-mRNA. Lane 2 shows the background binding to streptavidin Sepharose beads in the absence of the Biot-dH3 pre-mRNA. (E) Binding of FLASH, symplekin, and SLBP to indicated biotinylated RNAs. Lane 1 contains 2.5% of the Kc extract used in each binding experiment. (F) Protein factors from a Kc nuclear extract bound to the Biot-dH3 pre-mRNA in the absence of any RNA competitor (lane 1)

or in the presence of indicated RNAs (lanes 2, 3). The bound proteins were separated on a 12% SDS/polyacrylamide gel and stained with silver. The indicated bands and remaining portions of each lane were analyzed by mass spectrometry; the identified processing factors are listed to the right. The mass spectrometry results obtained for lane 3 were virtually identical to those obtained for lane 1



Band	Gene ID	Protein	Score: Lane 1	Lane 2
1	CG10110	CPSF160	1932	0
2	CG4616	FLASH	823	64
3	CG2097	Symplekin	2557	332
4	CG1957	CPSF100	1791	258
5	CG1109	WDR33	1035	80
6	CG17170	CstF77	431	189
7	CG7698	CPSF73	416	75
8	CG7697	CstF64	668	93

Figure 20. CstF64 is a component of the *Drosophila* U7 snRNP. (A) Western blot analysis of proteins bound to either the Biot-αU7 (lane 2) or Biot-αMock (lane 3) oligonucleotides using an antibody directed against full-length *Drosophila* CstF64. Lane 1 contains 2% of the input extract used in the binding experiment. (B) Immunoprecipitation of various polyadenylation factors from a Kc nuclear extract with the anti-CstF64 antibody alone (lane 3) or in the presence of the competing recombinant protein (lane 4). (*) Recombinant CstF64 bound to the antibody and collected on protein A agarose beads. Lane 1 contains 20% of the input extract used for immunoprecipitation, and lane 2 contains proteins of the nuclear extract bound to protein A agarose in the absence of any antibody. Immunoprecipitated proteins were identified by Western blotting using specific antibodies. (C) Immunoprecipitation of various polyadenylation factors from a Kc nuclear extract with the anti-CstF64 antibody alone (lane 3) or in the presence of the competing recombinant protein (lane 4). Lane 1 contains 10% of the input extract used for immunoprecipitation. (D) Proteins immunoprecipitated with the anti-CstF64 antibody (lane 3) or an anti-Mock antibody were separated on a 12% SDS/polyacrylamide gel, silver-stained, and analyzed by mass spectrometry. Lane 1 contains material bound to protein A agarose in the absence of any antibody. (E,F) Proteins precipitated with the anti-CstF64 antibody (lane 1 in both panels) or in the presence of the competing recombinant protein (lane 2 in panel E and lane 3 in panel F) were separated on a 10% (E) or 15% (F) SDS/polyacrylamide gel, silver-stained, and analyzed by mass spectrometry. Immunoprecipitation with the anti-CstF64 antibody using the nuclear extract pre-treated with RNase A is shown in lanes 3 (E) and 2 (F). (*) The CstF64 protein competitor bound to the antibody. The recombinant CstF64 used as a competitor is additionally shown in lane 4 of panel F. (G) Levels of *Drosophila* CF I_m68 in whole cell lysate from untreated S2 cells (lane 1) or cells treated with dsRNA against this factor (CF I_m68 KD,

lane 2), as determined by Western blotting using an antibody targeted to the N-terminal region of human CF I_m68. CR indicates two proteins cross-reacting with the antibody that served as a loading control. (H) Immunoprecipitation of *Drosophila* CF I_m68 from a Kc nuclear extract with the anti-CstF64 antibody (lane 4) or with a nonspecific antibody directed to a protein unrelated to 3' end processing (lane 3). The material bound to protein A agarose in the absence of any antibody is analyzed in lane 2. Lane 1 contains 10% of the input extract used for immunoprecipitation.

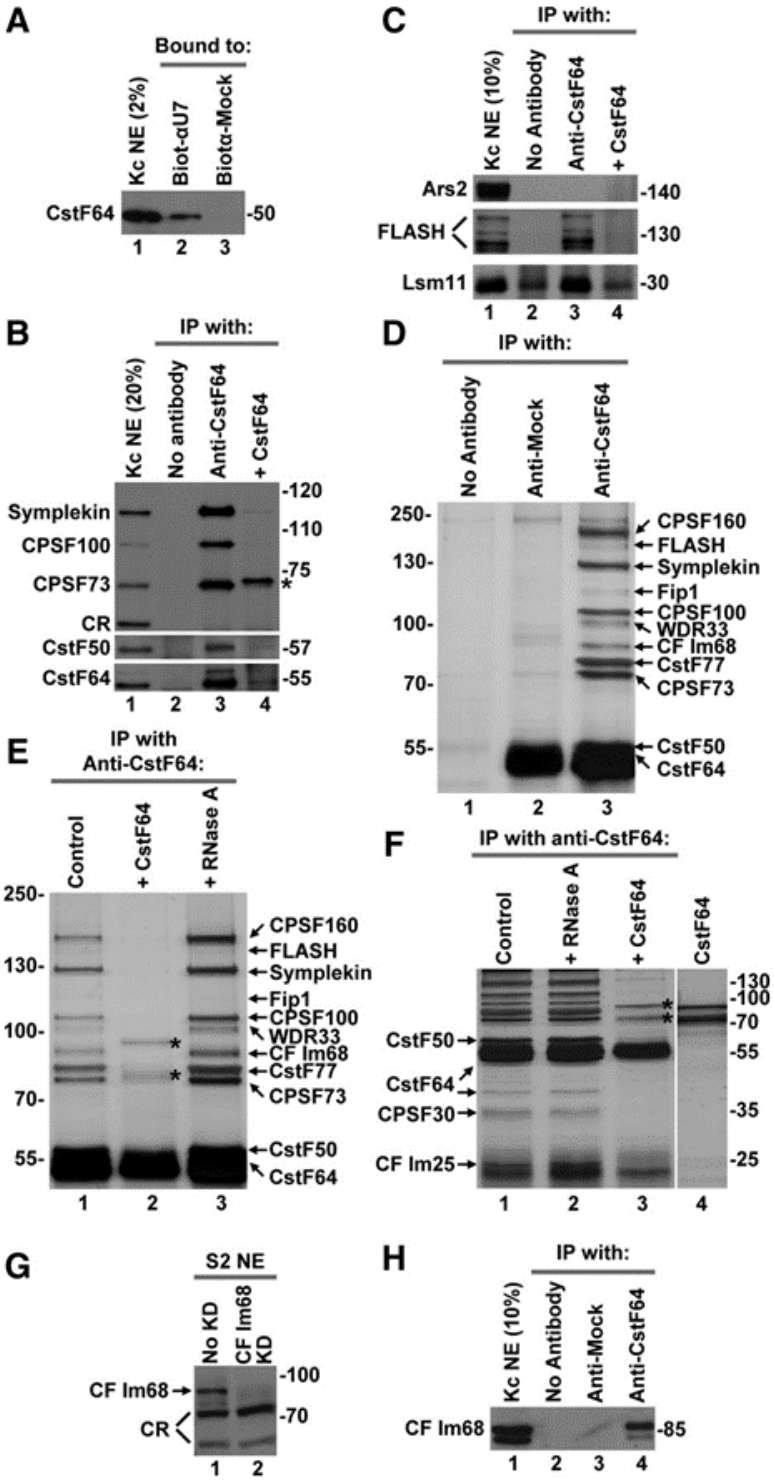
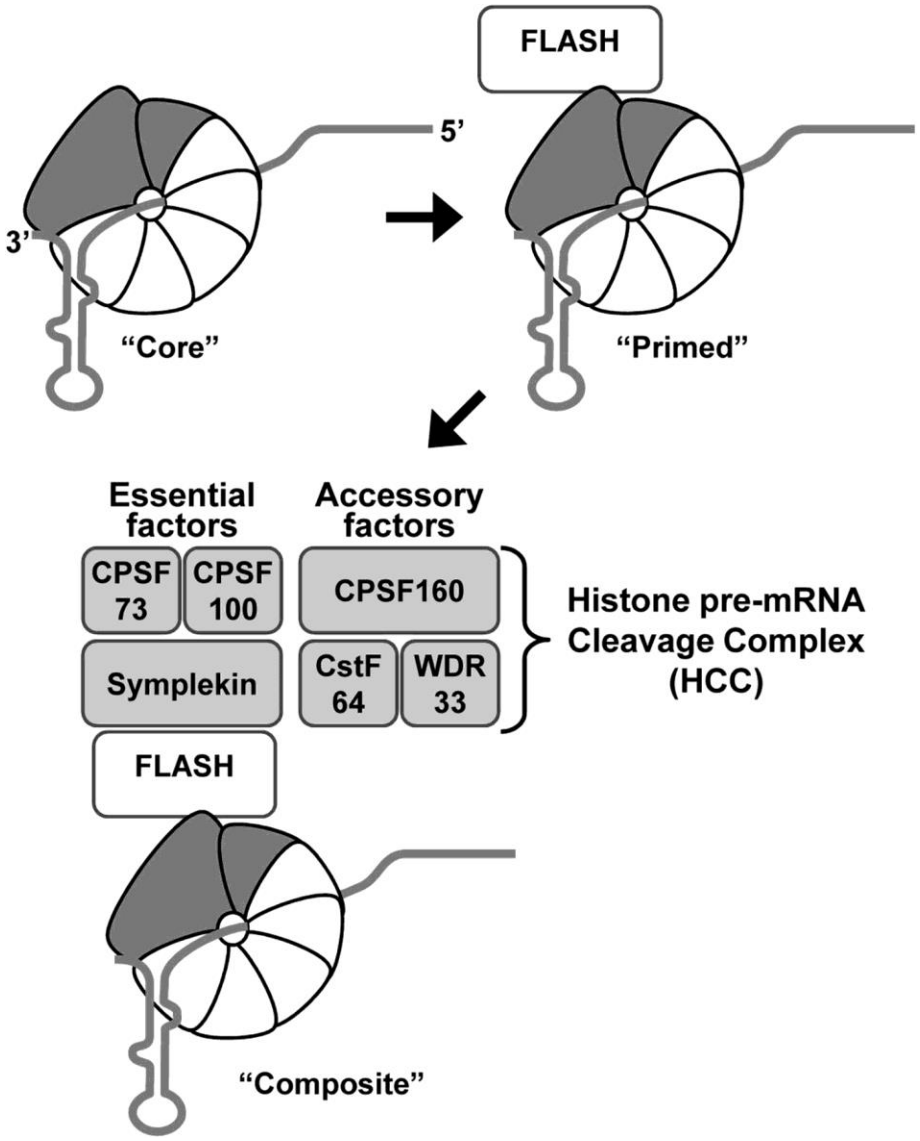

A
Bound to:
Kc NE (2%)
Biot-αU7
Biotα-Mock
CstF64
-50
1 2 3
B
IP with:
Kc NE (20%)
No antibody
Anti-CstF64
+ CstF64
Symplekin
CPSF100
CPSF73
CR
CstF50
CstF64
-120
-110
-75
-57
-55
1 2 3 4
C
IP with:
Kc NE (10%)
No Antibody
Anti-CstF64
+ CstF64
Ars2
FLASH
Lsm11
-140
-130
-30
1 2 3 4
D
IP with:
No Antibody
Anti-Mock
Anti-CstF64
250-
130-
100-
70-
55-
CPSF160
FLASH
Symplekin
Fip1
CPSF100
WDR33
CF Im68
CstF77
CPSF73
CstF50
CstF64
1 2 3
E
IP with Anti-CstF64:
Control
+ CstF64
+ RNase A
250-
130-
100-
70-
55-
CPSF160
FLASH
Symplekin
Fip1
CPSF100
WDR33
CF Im68
CstF77
CPSF73
CstF50
CstF64
1 2 3
F
IP with anti-CstF64:
Control
+ RNase A
+ CstF64
CstF64
CstF50
CstF64
CPSF30
CF Im25
-130
-100
-70
-55
-35
-25
1 2 3 4
G
S2 NE
No KD
CF Im68 KD
CF Im68
CR
-100
-70
1 2
H
IP with:
Kc NE (10%)
No Antibody
Anti-Mock
Anti-CstF64
CF Im68
-85
1 2 3 4

Figure 21. Major differences in 3' end processing of replication-dependent histone pre-mRNAs between mammals and *Drosophila.* Table identifies differences between 3' end processing of histone pre-mRNA in mammalian and insect cells.

Table 1. Differences in 3' end processing between mammals and *Drosophila*

Factor/Function	Mammalian	*Drosophila*
Position of cleavage site	Fixed position 5' of the HDE	Always 4 nts 3' of the stem-loop
Role of SLBP in processing	Stabilizes binding of the U7 snRNP. Dispensable if duplex between U7 snRNA 5' end and HDE is sufficient	Absolutely required
Composition of HCC	All six subunits of CPSF, symplekin, and CstF64	Four subunits of CPSF (Fip1 and CPSF30 missing), symplekin, and CstF64
Role of CstF64 in processing	Possibly essential (Ruepp et al. 2011)	Not essential
Ars2 and FLASH interaction	Yes	No
Role of FLASH in degradation of DCP	Not required	Essential for degradation

Figure 22. Potential forms of the U7 snRNP in *Drosophila*. The core particle consists of the U7 snRNA and the U7-specific Sm ring. This form is predominant in mammalian cells. In *Drosophila* cultured cells, most if not all U7 snRNP contains FLASH that interacts with Lsm11. As a result of this interaction, at least a fraction of the U7 snRNP assembles with a subset of polyadenylation factors (light gray) that form the Histone pre-mRNA Cleavage Complex (HCC). CPSF73, CPSF100, and symplekin are essential for 3' end processing of histone pre-mRNAs *in vivo*, whereas the remaining polyadenylation factors (CPSF160, WDR33, and CstF64) are dispensable and may play accessory/regulatory functions. The arrangement of polyadenylation factors within the HCC is arbitrary.


FLASH
5'
3'
"Core"
"Primed"
Essential factors
Accessory factors
CPSF 73
CPSF 100
CPSF160
Symplekin
CstF 64
WDR 33
Histone pre-mRNA Cleavage Complex (HCC)
FLASH
"Composite"

CHAPTER 4: INTERACTION OF C-TERMINAL FLASH WITH NPAT AND IDENTIFICATION OF MINIFLASH

INTRODUCTION

Nuclear Protein, Ataxia Telangiectaisa locus (NPAT) is a co-activator of histone gene transcription. CDK2/cyclinE phosphorylates five epitopes in the C-terminal half of NPAT at the G1 to S-phase transition. These modifications are essential for entry into S-phase and efficient histone gene transcription (Ma et al., 2000). The N-terminal half of NPAT contains a LisH domain at the N-terminus which contains residues critical to histone gene transcription (Wei et al., 2003) and a region that interacts with the histone specific transcription factor HINF-P (Miele et al., 2005).

Chapters 2 and 3 examined the critical regions of N-terminal FLASH in mammals and *Drosophila* and show it is required for histone pre-mRNA 3' end processing by interacting with Lsm11 and the HCC. Replication dependent histone genes are cell-cycle regulated with expression limited to S-phase, when there is a need to package newly synthesized DNA into chromatin. Upregulation of histone mRNA levels is achieved through a combination of transcriptional activation of histone genes coupled with histone pre-mRNA 3' end processing activity. Both NPAT and FLASH contribute to efficient histone gene expression and are stably associated with the histone locus body (HLB) (Barcaroli et al., 2006a; Barcaroli et al., 2006b). HLBs are nuclear bodies localized near histone gene clusters and enriched in factors critical for expression of histone genes. In human and mouse primary cells, four HLBs are visible during S-

phase and localized to the two histone gene loci on chromosomes 1 and 6 (Zhao et al., 2000). By concentrating factors essential to histone gene expression, HLBs unite two distinct processes required for histone biogenesis: transcriptional activation of the histone gene and post-transcriptional processing of the pre-mRNA 3' end. But the mechanism for localization of NPAT and FLASH to the HLB was unknown.

The C-terminal domain of FLASH (amino acids 1923-1976) is similar in sequence to that of an unrelated protein, YARP. YARP is an orthologue of *Drosophila* MUTE, which also localizes to the HLBs (White et al., 2011) and functions as a repressor of histone gene transcription (Bulchand et al., 2010). In this chapter, I demonstrate that C-terminal domains of FLASH and YARP bind C-terminal NPAT *in vitro*. When the C-terminal domains of FLASH or YARP are fused to an unrelated protein *in vivo*, the chimeric protein binds C-terminal NPAT. Furthermore, the interaction between NPAT with the C-terminal domains of FLASH or YARP is sufficient to localize the chimeric protein to HLBs. These results demonstrate a potential mechanism by which FLASH is recruited to HLBs for histone pre-mRNA 3' end processing. .

Additionally, I conducted a yeast two-hybrid screen using N-terminal Lsm11 (amino acids 1-168) with a human universal tissue library that isolated an alternatively spliced form of FLASH. This unique splice variant was not previously identified and encompasses the first N-terminal 138 amino acids shown to be critical for histone pre-mRNA 3' end processing fused to the terminal 52 amino acids shown to interact with NPAT. Named MiniFLASH, it is missing two massive internal exons that together account for approximately 1800 amino acids of the protein, including the region known to interact with Ars2. This interesting result demonstrates an instance where the two extreme ends of FLASH critical for 3' end processing and HLB localization are joined directly.

MATERIALS AND METHODS

Yeast two-hybrid system. The C-terminal region of FLASH (amino acids 1880-1982) was cloned into the pGBKT7 vector (Clontech) and used to screen a normalized Mate & Plate™ Library from HeLa S3 (Clontech, Cat. # 630479), as suggested by the manufacturer. Human MiniFLASH was isolated by screening normalized Universal Human Mate & Plate™ Library (Clontech, Cat. # 630480) against full length Lsm11 cloned in pGBKT7 (Yang et al., 2009a). Yeast diploid cells expressing proteins potentially interacting with the bait proteins were initially selected on plates lacking histidine and containing 5 mM 3-AT and subsequently tested on plates containing up to 100 mM 3-AT.

Antibodies. The following antibodies were use in this study: mouse monoclonal anti-HA 16B12 (Covance), mouse monoclonal anti-Myc 9E10 (Santa Cruz Biotechnology), rabit anti-CPSF73 A301-090A (Bethyl Labolatories), anti-SLBP (Wang et al., 1996), anti-FLASH (Yang et al., 2009a), sheep anti-mouse IgG NA931V (GE), donkey polyclonal anti-rabbit IgG Na934V (GE), goat anti-rabbit IgG labeled with green-fluorescent Alexa Fluor 488 (Life Technologies) and goat anti-mouse IgG labeled with red-fluorescent Alexa Fluor 594 (Life Technologies).

Transient expression of proteins in HeLa cells. The cDNA for human SLBP was extended at the C-terminus with a portion of human FLASH that encodes the last 103 amino acids of human FLASH (amino acids 1880-1982). The resultant insert was cloned into pcDNA3/HA downstream from the regions encoding three hemagglutinin (HA) tags. This construct (HA-SLBP/F) when expressed in HeLa cells yields a fusion protein consisting of three HA tags at the N-terminus followed by 234 amino acids of SLBP and the C-terminal portion of FLASH (amino acids 1880-1982). We also generated a construct consisting of the HA tag and SLBP but lacking the C-terminal FLASH (HA/SLBP). The C-terminal regions of human NPAT (amino acids 1297-1427)

and Drosophila mxc (amino acids 1669-1837) were cloned in the pcDNA3/Myc vector immediately downstream of a single Myc epitope. Approximately 106 HeLa cells grown in monolayers were co-transfected in the presence Lipofectamine 2000 (Invitrogen) with various combinations of constructs expressing the HA and Myc-tagged proteins and 24 hours later collected and lysed in NP40 buffer (0.5% NP40, 150 mM NaCl, 50 mM Tris-HCl pH 8, 10 mM sodium azide, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride). The cell lysates were supplemented with EDTA to a final concentration of 10 mM and incubated 30 min on ice with 50 ng of the stem-loop RNA containing biotin at the 5' end. Proteins associated with the RNA were collected on streptavidin sepharose beads (Sigma), separated in an SDS/polyacrylamide gel and analyzed by Western blotting, using anti-HA and anti-Myc antibodies.

Immunofluorescence. The HA-SLBP and HA-SLBP/F constructs were transiently expressed in HeLa cells and 24 hours later fixed with 3.7% formaldehyde and permeabilized with 0.5% Triton X-100. The intracellular localization of the tagged proteins was analyzed by confocal microscopy using an anti-HA antibody (Covance) and a Cy3-conjugated goat anti-mouse antibody. Histone Locus Bodies were detected with the DH3 monoclonal anti-NPAT antibody (kindly provided by E. Harlow) (Zhao et al., 2000) or rabbit polyclonal anti-FLASH antibody (Yang et al., 2009a).

***In vitro* protein expression and purification.** Recombinant FLASH, YARP and NPAT proteins, each N-terminally fused to glutathione S-transferase (GST), were expressed in bacteria from the pET-42a vector and purified on nickel beads (Qiagen) via the His tag present on each fusion protein. 35S-labeled proteins were cloned in a vector containing SP6 RNA promoter and expressed in rabbit reticulocytes using the Transcription and Translation coupled kit (TnT, Promega), according to the manufacturer's protocol.

Analysis of protein-protein interactions by GST pull down assays. Various combinations of two different proteins, one tagged with GST (~5 µg) and the other labeled with ^{35}S (10-15 µl of the TnT reaction, depending on the efficiency of labeling) were mixed in 100 µl of buffer D (100 mM KCl, 20 mM Hepes pH at 7.9, 20% glycerol, 0.2 mM EDTA at pH 8, 0.5 mM DTT) and incubated on ice for 30 min. The volume was subsequently increased to 600 µl with buffer D and each binding reaction was rotated with 30 µl of glutathione agarose beads (Sigma) at 4oC for 1 hour. The beads were rinsed several times with buffer D, rotated with the same buffer for additional 30 min, transferred to a new tube, re-suspended in a SDS sample buffer and bound proteins analyzed on an SDS/polyacrylamide gel. Each gel was initially stained with Coomassie Blue to monitor the amount of the GST tagged proteins and subsequently dried and used for autoradiography to detect ^{35}S-labeled proteins.

RESULTS

The C-terminal domains of FLASH and YARP interact with the C-terminus of NPAT. Human FLASH is a 220 kDa protein consisting of 1982 amino acids. Its N-terminal 135 amino acids play a critical role in 3' end processing of histone pre-mRNAs in vitro by forming a platform with Lsm11 that recruits the HCC to the U7 snRNP and hence to histone pre-mRNA (Yang et al., 2013). This is the most highly conserved region of FLASH, sharing a weak but recognizable homology with its *Drosophila* orthologue (Yang et al., 2009a; Burch et al., 2011). The C-terminal region of FLASH is highly conserved in all vertebrates but displays no similarity to the C-terminus of *Drosophila* FLASH. Strikingly, a sequence of 55 amino acids highly similar to the C-terminal region of FLASH is also present at the end of an unrelated vertebrate protein known as YY1 Associated Protein Related Protein 1 (YARP) or Gon4l (Fig. 23B)

(Ohtomo et al., 2007; Lu et al., 2010). As part of our interest in understanding the functional organization of FLASH, we investigated the role of its C-terminus and potential links between FLASH and YARP.

FLASH is known to interact through the N-terminal domain with itself (Kiriyama et al., 2009) and Lsm11 (Yang et al., 2009a) and through a short central motif (amino acids 931-943) with Ars2 (Kiriyama et al., 2009). To identify potential binding partners of the highly conserved C-terminal common to FLASH and YARP, Lalitha Kunduru cloned the last 108 amino acids of FLASH (Fig. 23C) in the yeast pGBKT7 bait vector (Clontech) and screened this fragment against a 2-hybrid library from HeLa cells (Clontech) in the presence of 5mM 3-aminotriazol. Among over 10 million diploids screened, fewer than 100 colonies rapidly grew under these relatively stringent conditions, with seven colonies containing the same insert encoding only the last 16 amino acids of NPAT (Nuclear Protein, Ataxia-Telangiectasia Locus), a known activator of histone gene transcription and a stable resident of Histone Locus Bodies (Ma et al., 2000; Zhao et al., 1998; Ye et al., 2003; Zhao et al., 2000). The second most frequently identified positive clone (4 independent isolates) contained the C-terminal half of Proliferating Cell Nuclear Antigen (PCNA), a component of the DNA replication machinery. Judging from the growth rate in the presence of increasing concentrations of 3-AT, the C-terminal region of FLASH interacts stronger with NPAT than with PCNA (Fig. 1D).

Since NPAT and FLASH are stable residents of HLBs, their interaction detected in yeast cells was likely biologically relevant and raised the possibility that YARP may also interact with NPAT and could be a new HLB component with a role in histone gene expression. We used a GST pull down assay to determine whether the C-terminal regions of FLASH and YARP directly interact in vitro with the C-terminus of NPAT. The terminal 16-amino acid region of NPAT was

cloned downstream of GST in pET42-a (Fig. 1E) and expressed as the GST-Np16C fusion protein in bacteria. The C-terminal regions of FLASH (Fl103C, amino acids 1880-1982) and YARP (Ya97C, amino acids 2117-2213) were expressed and labeled with 35S-methionine in rabbit reticulocyte lysate using the TnT kit (Promega). In the pull down assay, consistently more than 20% of 35S-labeled Fl103C used in the experiment bound to the GST-Np16C fusion protein and retained on glutathione agarose beads (Fig. 1G, top two panels, lane 3). The corresponding domain of YARP also bound to GST-Np16C, although the interaction was less efficient, with on average 15% input recovered from glutathione beads (Fig. 1G, top middle panels, lane 2). In contrast, the N-terminal domain of FLASH (GST-F138N, amino acids 1-138) that is involved in 3' end processing of histone pre-mRNAs does not detectably interact with GST-NP16C (Fig. 1G, top bottom panels, lane 2).

To confirm the interaction of the C-terminal NPAT with the C-terminal regions of FLASH or YARP, we carried out a reciprocal experiment. The C-terminal amino acids of FLASH (1880-1982) or YARP (2117-2213) were cloned downstream of an N-terminal GST tag and expressed in bacteria (GST-Fl103C and GST-Ya97C, respectively). The C-terminal region of NPAT encompassing the last 131 amino acids of NPAT (Fig. 2A) was expressed rabbit reticulocyte lysate, yielding 35S labeled NPAT-131C. In the pull down assay, as much as 20% of the input NPAT-131C was consistently bound by GST-Fl103C or GST-Ya97C (Fig. 2B, top panel). To determine whether the interaction involves only the last 16 amino acids of NPAT or extends to a more upstream sequence, we removed these amino acids from the NPAT-131C protein, creating NPAT-Δ16C. This deletion significantly reduced the interaction with GST-Fl103C and nearly abolished the interaction with GST-Ya97C (Fig. 2B, middle panel, lanes 3 and 4, respectively). Deleting additional 9 amino acids from the end (NpΔ25C) was sufficient to

eliminate the interaction with GST-Fl103C (Fig. 2B, bottom panel, lane 3). We concluded that the C-terminal domains of FLASH and YARP interact with a larger than 16 amino acids portion of NPAT, and the interaction likely extends to as many as its 25 last amino acids. In addition, these results indicate that the C-terminal domain of NPAT interacts more strongly with FLASH than with YARP.

To prove that the interaction of NPAT with FLASH or YARP is mediated by the shared 55-amino acid domain rather than by unique sequences located outside the conserved region, we expressed a GST fusion protein linked to the domain of FLASH (GST-Fl54C). In the GST-pull downs assay Np131C interacted with this domain as efficiently as with the longer C-terminal FLASH fragment consisting of 103 amino acids (Fig. 2C, lanes 3 and 4). We synthesized a peptide containing the last 25 amino acids of NPAT (Np25C) and tested its ability to compete with GST-Np16C in binding 35S-labeled Fl103C (Fig. 2D, lanes 4 and 5). The interaction of GST-Np16C with Fl103C was completely eliminated in the presence of 1 µg and 10 µg of Np25C (2 and 20 fold excess of the peptide relative to GST-Np16C, respectively), confirming that the 25-amino acid C-terminal region of NPAT alone is capable of strongly interacting with the C-terminal domain of FLASH.

We also tested the in vitro interaction between the C-terminal regions of FLASH and YARP with PCNA. PCNA is a cofactor of the DNA polymerase delta and plays a key role in enhancing the efficiency of DNA synthesis of the leading strand during S-phase (Jonsson and Hubscher, 1997). Its interaction with FLASH could facilitate coupling of histone gene expression with DNA replication and other key S-phase events. The GST pull down assay confirmed that full length PCNA (261 amino acids) interacts with the C-terminal FLASH in vitro (Fig. 2E, lane 3), although this interaction is significantly weaker than between NPAT and

FLASH. This result is in a good agreement with the data from the yeast two-hybrid screen (Fig. 1D). The 54-amino acid domain of FLASH is sufficient for the interaction. The same domain of YARP did not interact with PCNA, indicating that FLASH and YARP in spite of sharing significant homology at the C-terminus may interact with both distinct and overlapping partners. PCNA is known to bind numerous proteins in vivo (De Biasio and Blanco, 2013; Maga and Hubscher, 2003) and further studies are required to determine whether its interaction with FLASH has biological significance.

The interaction between the C-terminal domains of FLASH and NPAT in HeLa cells. To determine whether the C-terminal regions of FLASH and NPAT interact in cells, we constructed the HA-SLBP/F clone encoding human SLBP extended at the C-terminus with the last 103 amino acids of FLASH and at the N-terminus with a 3xHA tag (Fig. 3A). We also constructed a similar clone that lacked the C-terminal FLASH extension, consisting solely of the 3xHA tag at the N-terminus followed by full length SLBP (HA-SLBP).

Since the endogenous NPAT is present in limiting quantities and generally difficult to detect by Western blotting, we constructed a clone that expresses the last 131 amino acids of human NPAT (amino acids 1297-1427) as a fusion with an N-terminal Myc epitope (Myc-Np131C). As a negative control, we used a clone encoding the last 169 amino acids of mxc (amino acids 1669-1837, Myc-mxc169C), the Drosophila orthologue of NPAT, which localizes to HLBs in Drosophila (White et al., 2011) but shares no recognizable homology with the C-terminus of human NPAT (Fig. 3A). Each HA tagged version of SLBP was co-expressed in HeLa cells with either Myc-Np131C or Myc-mxc169C. Approximately 24 hours following transfection, HeLa cells were collected, lysed and the SLBP from each lysate was affinity-purified on streptavidin agarose beads uisng a short RNA containing the stem-loop structure

from histone mRNA and biotin on the 5' end (Biot-SL). This method of affinity purifying SLBP was used by us previously (Dominski et al., 2003) and is efficient and specific.

The Biot-SL RNA purifies HA-SLBP and HA-SLBP/F, with the efficiency varying between experiments from 10 to 40% of the input (Fig. 3C and not shown). The precipitate also contains the endogenous SLBP, which does not react with the anti-HA antibody and can only be detected by an anti-SLBP antibody (not shown). We used anti-Myc antibodies and Western blotting to analyze the material purified by the stem-loop RNA. Importantly, HA-SLBP/F but not HA-SLBP co-purified with Myc-Np131C, demonstrating that the C-terminal regions of FLASH and NPAT form a complex in vivo. As expected, the material did not contain detectable amounts of CPSF73 and other proteins of the whole cell lysate, confirming that the RNA-mediated purification is very selective. The specificity of the method was additionally confirmed by expressing an HA-tagged protein that lacks SLBP and fails to interact with the SL RNA (see Fig. 6F, lane 4).

To test whether the interaction between the C-terminal regions of human FLASH and NPAT is specific, we co-expressed HA-SLBP/F with Myc-mxc169C, containing the C-terminal amino acids of Drosophila mxc (amino acids 1669-1837). Whole cell lysates were prepared from HeLa cells transiently expressing HA-SLBP/F and Myc-mxc169C and SLBP was purified from each lysate by the Biot-SL RNA. As determined by Western blotting with anti-HA and Myc antibodies, the recovered material contained HA-SLBP/F but lacked any detectable Myc-mxc169C (Fig. 3D, lane 2), demonstrating that these two proteins do not interact.

We also tested the interaction between HA-SLBP/F and Myc-Np131C using immunoprecipiation. The same lysates were incubated with an anti-SLBP antibody and the precipitated material analyzed by Western blotting, using anti-HA and anti-Myc antibodies.

Again, the HA-tagged SLBP fused with the C-terminal region of FLASH (HA-SLBP/F) co-precipitated with Myc-Np131C (Fig. 3E, compare lanes 2 and 4). Altogether, the in vitro pull down experiments and studies in HeLa cells established that the C-terminal portions of FLASH and human NPAT make direct contacts with each other. Previous studies demonstrated that the endogenous FLASH and NPAT co-precipitate, but failed to determine whether they interact directly or exist in a common complex containing additional proteins (Barcaroli et al., 2006a; Kiriyama et al., 2009).

The C-terminal domain of FLASH targets a heterologous protein to HLBs. SLBP is a 3' end processing factor that tightly binds the highly conserved stem-loop structure in histone pre-mRNA located upstream of the cleavage site and remains associated with the mature histone mRNA following 3' end processing (Marzluff et al., 2008). The complex of mature histone mRNA and SLBP is ultimately exported from the nucleus for translation. As a result, in mammalian cells SLBP is only transiently associated with histone gene loci and is detected in the S-phase cells throughout the nucleoplasm and cytoplasm, displaying no significant enrichment in Histone Locus Bodies (Erkmann et al., 2005) (see Fig. 4B). This is in a sharp contrast to the U7 snRNP, including its integral components FLASH and Lsm11, which persists in HLBs after each round of processing (Liu et al., 2006; Pillai et al., 2001b).

To investigate the biological role of the interaction between the C-terminal regions of FLASH and NPAT, we tested whether SLBP fused with the last 103 amino acids of FLASH is concentrated in HLBs. We transiently expressed HA-SLBP and HA-SLBP/F in HeLa cells and used fluorescent confocal microscopy to determine their intracellular localization 24 hours post transfection. Since NPAT is a limiting factor, it was possible that overexpression of HA-SLBP/F

would result in its diffused localization throughout the nucleoplasm, hence preventing the detection of HLBs. To ensure expression at a physiological level, HeLa cells were transfected with small amounts of plasmid DNA. Indeed, the steady-state levels of both HA-SLBP and HA-SLBP/F were comparable to the level of endogenous SLBP (Fig. 4A, lanes 2 and 3).

HeLa HLBs were stained with an antibody against the N-terminal portion of FLASH and the intracellular localization of the HA-tagged SLBP proteins expressed from transfected plasmids was analyzed by an anti-HA antibody. HA-SLBP, differing from the endogenous SLBP only by the presence of the HA-tag, was localized predominantly to the nucleoplasm, showing a uniform distribution and no detectable foci that would co-localize with HLBs (Fig. 4B). In contrast, the HA-SLBP/F fusion protein localized to distinct nuclear structures and this punctuate pattern of staining was coincident with HLBs detected by the anti-FLASH antibody (Fig. 4C). We conclude that the C-terminal domain of FLASH that interacts with NPAT functions as a signal for localizing/retaining proteins in HLBs.

Intracellular localization of YARP. The presence of the FLASH-like domain at the C-terminus of YARP and its ability to interact with NPAT in vitro suggested that this protein may be a stable component of HLBs. We first tested whether YARP and NPAT interact in vivo. An HA-tagged version of SLBP extended at the C-terminus by amino acids 2117-2213 of YARP (HA-SLBP/Y) was transiently co-expressed with Myc-Np131C in HeLa cells and purified from whole cell lysate 24 hours post transfection by the Biot-SL RNA. The material collected on streptavidin beads in the presence of this RNA was highly enriched in HA-SLBP/Y and contained Myc-Np131C (Fig. 5A, lane 6), indicating that these two proteins interact in vivo. As controls, we again tested the interaction of Myc-Np131C with the two other SLBP fusion proteins: HA-SLBP

and HA-SLBP/F (Fig. 5A, lane 6). Consistent, with the data presented above, only HA-SLBP/F co-purified with Myc-Np131C. No background amounts of Myc-Np131C were detected in the material purified from cells expressing HA-SLBP, confirming that the interaction requires the C-terminal region of either FLASH or YARP.

Xiao Yang transfected HeLa cells with low doses of DNA encoding HA-SLBP/Y and used confocal fluorescent microscopy to monitor the localization of the transiently expressed HA-SLBP/Y in HeLa cells. As determined by an anti-SLBP primary antibody and a goat anti-rabbit secondary antibody conjugated to green-fluorescent Alexa Fluor 488, SLBP in addition to showing a uniform nucleoplasmic distribution was targeted to a number of much brighter nuclear foci. These foci were coincident with NPAT detected by an anti-NPAT monoclonal antibody (Zhao et al., 2000) and stained red with a secondary, anti-mouse antibody conjugated to Alexa Fluor 594 (Fig. 5B). Thus, the C-terminal portion of YARP functions in the same manner as the homologous C-terminal portion of FLASH and when appended to a heterologous protein, promotes its retention in HLBs.

We next analyzed the intra-cellular localization of endogenous YARP and FLASH in HeLa cells using the cross-adsorbed and affinity-purified anti-YARP(C) or anti-FLASH(C) primary antibodies and a goat anti-rabbit secondary antibody conjugated to green-fluorescent Alexa Fluor 488. HLBs were detected in the same cells using a mouse monoclonal antibody against NPAT (Zhao et al., 2000) and stained red with a secondary anti-mouse antibody conjugated to Alexa Fluor 594. The anti-FLASH(C) antibody detected a number of foci in HeLa nuclei that perfectly co-localized with the NPAT-stained HLBs (Fig. 5D). This result is in agreement with our previous localization experiments conducted with an antibody directed against the N-terminal region of FLASH (Yang et al., 2009a) (Fig. 5E). Importantly, the anti-

YARP(C) antibody also stained a number of foci that always co-localized with NPAT. Based on these results we conclude that YARP is a newly identified component of HLBs in mammalian cells and its localization is likely mediated by the C-terminal domain that interacts with NPAT. In addition to detecting HLBs, the anti-YARP(C) antibody stained the nucleoplasm, excluding the nucleolus. Thus, YARP while enriched in the vicinity of histone gene loci may also bind to other regions of chromatin, consistent with its proposed role in controlling expression of various genes (see below).

Mini-FLASH, a short form of FLASH is localized to HLBs. In FLASH, amino acids 100-135 interact with Lsm11, the largest component of the Sm ring in U7 snRNP (Yang et al., 2009a; Yang et al., 2011). Together they form a platform that recruits Histone pre-mRNA Cleavage Complex (HCC), which contains the CPSF73 endonuclease and other polyadenylation factors to the U7 snRNP (Yang et al., 2013). The short N-terminal fragment of FLASH containing the binding sites for Lsm11 and the HCC strongly stimulates the activity of nuclear extracts in 3' end processing of histone pre-mRNAs (Yang et al., 2009a; Yang et al., 2011), indicating that these two regions mediates all essential processing functions of FLASH in nuclear extracts and that the remaining 95% of the protein is dispensable in vitro. Here, we have determined that the C-terminal domain of FLASH (amino acids 1895-1982) targets FLASH to HLBs. Thus, out of nearly 2,000 amino acids of FLASH, fewer than 200 residues derived from the opposing N- and C-terminal ends may be sufficient to support full processing activity of this protein in vivo.

As part of our continued effort to identify proteins interacting with Lsm11, we screened several independent cDNA libraries using the yeast two-hybrid system. In most screens, we consistently isolated clones encoding various N-terminal portions of FLASH that included the

Lsm11 binding site (amino acids 100-135) (Yang et al., 2009a). Strikingly, a universal cDNA library obtained from a collection of adult human tissues representing a broad range of expressed genes and both male and female donors yielded a FLASH variant in which the N- and C-terminal regions were directly linked and nearly 1800 amino acids of central part of the protein was missing. Of the 190-amino acid total length of this variant, the first 138 residues and the last 52 residues originated from the N- and C-terminus of full length FLASH, respectively (Fig. 6A). We refer to this protein as MiniFLASH (MiniF). MiniF is encoded by an alternatively splice FLASH mRNA that lacks two internal exons, consisting of over 2,000 and 3,000 nucleotides (Fig. 6B). These exons vastly exceed 150 nucleotides for an average exon and are among the largest internal exons in the genome, (Hawkins, 1988; Bolisetty and Beemon, 2012).

The N-terminal region of MiniF (amino acids 1-138) contains all sequences necessary for 3' end processing and is active in vitro. We first tested whether MiniF is capable of supporting in vitro processing of a model 5'-labeled histone pre-mRNA derived from the mouse histone H2a-614 gene (Yang et al., 2009a; Yang et al., 2011). This 86-nucleotide substrate contains all necessary processing sequences and upon cleavage is converted to a 48-nucelotide mature product terminated with the terminal stem-loop (Fig. 6C). The GST-tagged MiniF was expressed in bacteria and added to a highly diluted mouse nuclear extract (equivalent of 0.5 µl of undiluted extract) that is virtually inactive in processing due to limiting amounts of FLASH (Fig. 6D, lane 2). Consistent with our previous reports (Yang et al., 2009a; Yang et al., 2011), the GST tagged N-terminal region of FLASH (GST-F138N) stimulates 3' end processing of a model mouse H2a pre-mRNA increasing processing from a nearly undetectable level to 10% (Fig. 6D, lane 5). As expected, the NPAT-interacting FLASH region encompassing amino acids 1880-1982 (GST-Fl103C) has no effect on processing (Fig. 6D, lane 5). Importantly, GST-MiniF that

contains the same N-terminal region followed by the last 52 C-terminal amino acids of FLASH was functional (Fig. 6D, lane 4). Thus, adding 52 FLASH C-terminal amino acids to the N-terminal region does not interfere with its ability to function in processing in vitro.

The last 52 amino acids of MiniF correspond to a truncated NPAT-binding domain that lacks the 9-amino acid GEIIILWTR sequence at the N-terminus (Fig. 6A and B). We tested whether this incomplete C-terminal domain is sufficient to interact with the C-terminal NPAT and can support proper localization of MiniF to HLBs in vivo. In the GST pull down assay, MiniF compared to Fl103C showed reduced ability to interact with the C-terminus of NPAT (Fig. 6E, top panel, lanes 3 and 4), indicating that the missing 9-amino acid region is important for the optimal interaction. Indeed, eight residues of this region are conserved in all known vertebrate FLASH and YARP orthologue and represent the N-terminal boundary of the sequence similarity between FLASH and YARP (Fig. 1A). Shortening the C-terminal region of NPAT by removing the last 16 amino acids eliminated the interaction with MiniF (Fig. 6E, bottom panel, lane 4). Consistent with the data shown in Fig. 2, the interaction of NPATΔ16 with the entire C-terminal domain of FLASH was significantly reduced (Fig. 6E, bottom panel, lane 3). Altogether, these data indicate that the incomplete C-terminal FLASH domain in MiniF retains partial ability to interact with NPAT.

The fact that human MiniF is generated as a result of alternative splicing and this isoform was not present in HeLa cell libraries suggests that it could be expressed in a tissue- or developmentally-specific manner and functional in vivo. To determine whether the incomplete C-terminal domain is sufficient to deliver MiniF to HLBs, we transiently expressed a 3xHA tagged version of MiniF in HeLa cells and monitored its intracellular localization. MiniF expressed at levels readily detectable by Western blotting and migrated at the expected size of 30

kDa (Fig. 6F, lane 3). HLBs were detected by antibodies against FLASH (which also detects MiniF) and stained green, whereas MiniF was detected by anti-HA antibodies and stained red (Fig. 6G). These experiments showed that the HA-tagged MiniF is concentrated in HLBs, indicating that when expressed, it may participate in generating mature histone mRNAs in vivo.

DISCUSSION

Human FLASH, a protein of nearly 2000 amino acids, is essential for U7-dependent 3' end processing of histone pre-mRNAs and resides in Histone Locus Bodies (HLBs) (Yang et al., 2009a; Ghule et al., 2008; Barcaroli et al., 2006a; Barcaroli et al., 2006b). Its N-terminal region of about 135 amino acids interacts with Lsm11, a component of the U7 snRNP, and together these two proteins recruit a subset of polyadenylation factors, including the endonuclease, to histone pre-mRNA for 3' end cleavage (Yang et al., 2009a; Yang et al., 2011). The N-terminal region is also involved in FLASH self-association, whereas a short motif located in the center mediates the interaction of FLASH with Ars2, a stable component of the Cap Binding Complex (Kiriyama et al., 2009). The structural and functional organization of FLASH outside these regions is poorly characterized. The 55-amino acid domain located near the C-terminus is among the most highly conserved regions in vertebrate FLASH. Strikingly, a highly similar sequence with about 50% identical residues is also present at the end of an unrelated protein known as YARP (Yin Yang 1 Associated Protein-Related Protein) or Gon4-like (Gon4l). Studies on YARP suggested that it functions as a transcriptional regulator involved in controlling critical developmental decisions but provided no information on the role of its C-terminal domain (Lu et al., 2010; Lu et al., 2011).

The C-terminal domains of FLASH and YARP interact with the extreme C-terminus of NPAT. Using yeast two-hybrid screens, we identified NPAT as a factor that binds the C-terminal domain of FLASH. This finding was validated by in vitro GST pull down assays and in vivo studies in HeLa cells. The homologous C-terminal domain of YARP also interacts with NPAT. Strikingly, of nearly 1500 amino acids of NPAT, the region binding the C-terminal domains of FLASH and YARP is limited to the last 25-30 amino acids, with as few as 16 extreme C-terminal NPAT residues being sufficient to mediate a strong interaction in vitro.

The structures of the C-terminal domains of FLASH and YARP have been determined by structural genomic projects and both proteins adopt the same overall fold, consisting of three α-helices. This fold structurally resembles the SANT and Myb domains and is referred to as the SANT/Myb-like domain. The Myb domain functions as a DNA recognition module in Myb-related transcriptional activators (Ogata et al., 1994). The SANT domain is an essential element of many chromatin-remodeling proteins and was named after four proteins that share this domain: SWI3, ADA2, N-CoR and TFIIIB (Boyer et al., 2004). Sequence alignments, secondary structure predictions and X-ray crystallography suggest that in spite of structural similarities with the Myb domain, the SANT domain is unable to bind DNA and its function remains unknown (Horton et al., 2007). Interestingly, it has been proposed that the SANT domain may function as a universal module that interacts with histone tails (Boyer et al., 2004). Our studies demonstrated that the related C-terminal SANT/Myb-like domain in FLASH and YARP functions in vertebrates as an NPAT-interacting module.

Strikingly, compared to the FLASH orthologue, all vertebrate YARP orthologues contain an insertion of a single amino acid, a glycine that is located at the C-terminal end of helix 2. The absolute conservation of this amino acid suggests that it functionally important for YARP and

may provide a critical feature that discriminates this domain from the C-terminal domain of FLASH. The interaction between NPAT and FLASH is significantly stronger than that between NPAT and YARP, indicating the FLASH and YARP C-terminal domains in spite of overall similarities are not fully equivalent.

The extreme C-terminal region of NPAT is among the most highly conserved regions of NPAT in vertebrates, although based on the sequence analysis and secondary structure prediction it does not resemble any known domain. A similar sequence can be also found at the end of the uncharacterized sea urchin Strongylocentrotus purpuratus LOC582916 protein, which based on the presence of the Lissencephaly type-1-like homology (Lis-H) motif at the N-terminus likely represents the orthologue of NPAT. The genome of *S. purpuratus* also encodes the orthologue of YARP that ends with a SANT/Myb-like domain. Thus, the interaction between the C-terminal region of NPAT and the SANT/Myb-like domain likely arose early in metazoan evolution.

The similarity between vertebrate and *Drosophila* FLASH is limited to the N-terminal region, which interacts with Lsm11 and polyadenylation factors (Yang et al., 2009a; 2013) and the C-terminal regions are unrelated (Burch et al., 2011). The NPAT orthologue in *Drosophila* has been recently identified as *mxc* (White et al., 2011) and its C-terminus also shares no recognizable sequence similarity with the vertebrate NPAT. Intriguingly, the localization of FLASH to HLBs in *Drosophila* depends on its C-terminal region and the C-terminal region of *mxc* (White et al., 2011; Rajendra et al., 2010). Thus, the interaction between the C-terminal regions of FLASH and NPAT may be a universal feature of all animals, although in *Drosophila* (protostomes) and vertebrates/sea urchins (deuterostomes) it is mediated by unrelated sequences.

This conserved arrangement may be mechanistically essential and critical for the proper function of FLASH and NPAT in histone gene expression and HLB biogenesis.

The SANT/Myb-like domain of FLASH and YARP functions as an HLB localization signal.

What is the role of the interaction between the SANT/Myb-like domain of FLASH and NPAT? We showed that the addition of the C-terminal region of FLASH to SLBP converts SLBP from a shuttling protein that displays diffused pattern of localization throughout the nucleoplasm and cytoplasm into a stable resident of HLBs. Thus, at least one purpose of this interaction is to efficiently accumulate proteins in HLBs. Theoretically, FLASH might localize to HLBs through its interaction with the U7 snRNP, which binds the histone downstream element (HDE) in histone pre-mRNA. However, our studies with a mutant *Drosophila* FLASH containing the Lsm11-interacting N-terminal domain but lacking the C-terminal region indicate that this way of localizing FLASH to HLBs in Drosophila cells (Burch et al., 2011)and flies (D. Tatomer, R.J. Duronio, Z.D., W.F.M., submitted) is inefficient. Truncated FLASH has a largely diffused distribution in the nucleoplasm of Drosophila cells, explaining the requirement for an alternative targeting mechanism (Burch et al., 2011). The interaction between FLASH and NPAT (or *mxc* in *Drosophila*) may additionally serve to mobilize the U7 snRNP to HLBs before generation of the nascent histone pre-mRNAs and to integrate the transcriptional co-activation function of NPAT with the 3'end processing activity of FLASH. The accumulation of processing factors in distinct nuclear structures, likely HLBs, in a manner independent of binding to histone pre-mRNA was observed in *Xenopus* oocytes (Abbott et al., 1999; Wu and Gall, 1993) and *Drosophila* embryos (White et al., 2011; Salzler et al., 2013). The U7 snRNP is limiting and creating its high local concentration in the direct vicinity of histone genes prior to the onset of

their transcription may be important to increase the efficiency and fidelity of the correct 3' end processing, preventing undesired processing by cleavage and polyadenylation (Wagner et al., 2007).

YARP as the orthologue of *Drosophila* Mute. Consistent with the presence of the NPAT-interacting domain, Xiao Yang demonstrates that endogenous YARP localizes to HLBs in HeLa cells. YARP throughout most of its length shares significant similarity with *Drosophila* Mute, a component of HLBs in *Drosophila* cells (Bulchand et al., 2010; White et al., 2011), suggesting that these two proteins may be functional counterparts. It has been suggested that the C-terminal region of Mute also contains the SANT/Myb-like domain (Bulchand et al., 2010), although it displays only a weak similarity to the C-terminal regions of vertebrate YARP and FLASH. It is unknown whether this degenerate SANT/Myb-like domain is responsible for targeting Mute to HLBs and whether it directly interacts with *mxc*. Clearly, the C-terminal region of *Drosophila* FLASH does not resemble the Mute SANT/Myb-like domain and likely operates as an HLB-targeting signal trough a different mechanism.

Deficiency of Mute affects *Drosophila* muscle development, hence Mute is also referred to as "Muscle wasted" (Bulchand et al., 2010). Mutations in Mute result at the molecular level in an increased accumulation of histone mRNAs in fly embryos (Bulchand et al., 2010). This observation supports the possibility that one of the functions of Mute is to repress transcription of histone genes (Bulchand et al., 2010). It has been shown that YARP forms a complex consisting of the multifunctional transcriptional factor Ying Yang 1 (YY1), the transcriptional co-repressor Sin3a and histone deacetylase 1 (HDAC1), which functions as a transcriptional repressor that commits B cell progenitors to B lymphopoiesis (Lu et al., 2011). Interestingly,

virtually all metazoan replication-dependent histone genes contain elements that bind YY1 (Last et al., 1999; Eliassen et al., 1998; Palko et al. 2004). Thus, YARP through the interaction with YY1 may recruit HDAC1 to these genes to negatively regulate their transcription, as postulated for Mute.

Both YARP and Mute are likely versatile transcriptional factors and in addition to regulating histone genes may affect expression of multiple genes, including those essential for development (Bulchand et al., 2010; Kiriyama et al., 2009; Liu et al., 2007). Some of these other functions of YARP and Mute may be executed by their various splice variants and paralogues. The human genome contains for example a gene encoding a YARP homologue, YY1 Associated Protein (YY1AP) that was generated as a result of partial duplication of the YARP gene during evolution of anthropoids (Ohtomo et al., 2007; Kuryshev et al., 2006). YY1AP corresponds to the central part of YARP (amino acids 601-1338) and the two regions share nearly 100% sequence identity, consistent with the evolutionarily recent duplication event. YY1AP also binds YY1 (Wang et al., 2004)but lacks the SANT/Myb-like domain and is therefore unlikely to localize to HLBs.

MiniF is a splicing variant of FLASH. FLASH is encoded by multiple exons with a potential to generate multiple splice variants. However, sequencing data based on RNA from cultured cells or human tissues provides no indication of extensive alternative splicing of FLASH pre-mRNA. To our surprise, by screening a cDNA library constructed on mRNA isolated from a mixture of human tissues we identified a number of independent clones encoding a short form of FLASH that we refer to as MiniFLASH (MiniF). MiniF consists of only 190 amino acids and results from skipping two extremely large internal exons and an in-frame fusion of exons encoding the N- and C-terminal regions. The first 138 amino acids of MiniF derive from the N-

terminus of FLASH and we previously demonstrated that this region contains all elements necessary for 3' end processing of histone pre-mRNAs in vitro (Tang et al., 2009a; 2013; 2011). The remaining 52 amino acids of MiniF correspond to the C-terminus of FLASH and encompass almost the entire SANT/Myb-like domain. This incomplete domain retains some ability to interact with NPAT and is sufficient to localize MiniF to HLBs, the site of histone pre-mRNA processing in vivo.

The composition of MiniF, combining the two external FLASH domains required for processing and HLB localization, suggests that it may at least partially substitute for full length FLASH in generating mature histone mRNAs in vivo. This interpretation is supported by our recent studies in *Drosophila*. The processing and HLB localization domains in Drosophila FLASH are also located in the N- and C-terminal regions, respectively. We engineered an artificial Drosophila FLASH variant by directly linking the terminal domains and deleting most of the protein body of nearly 550 amino acids. This artificial Drosophila MiniF restores virtually all phenotypic and molecular defects resulting from the deficiency of the endogenous full length FLASH in Drosophila tissue culture cells (Burch et al., 2011) and whole animals (D. Tatomer, R.J. Duronio, Z.D., W.F.M., submitted).

Exons, in contrast to introns, are typically short, encoding on average less than 50 amino acids (Hawkins, 1988). This limited length facilitates the mechanism of exon definition by promoting direct interaction between factors bound to the flanking splice sites (Eliassen et al., 1998; Robberson et al., 1990). Long exons tend to be skipped and their recognition frequently requires an active mechanism, involving a subset of specific RNA binding proteins that interact with exonic splicing enhancers (Bolisetty and Beemon, 2012). FLASH pre-mRNA contains two of the longest exons in the mammalian geneome. It is unknown whether their skipping and

generation of MiniF is an aberrant process or a result of a highly regulated alternative splicing specific for some tissues or developmental stages. The central region of mammalian FLASH deleted from MiniF due to exon skipping consists of nearly 1,800 amino acids and plays a largely unknown role. Parts of this region may control apoptosis (Imai et al.) and transcription of selected genes (Alm-Kristiansen et al., 2008), two processes previously linked to FLASH. A short 13-amino acid motif located between amino acids 931-943 that is missing in MiniF interacts with Ars2 (Kiriyama et al., 2009). Ars2 exists in a tight complex with Cap Binding Proteins of 20 and 80 kDa and its interaction with FLASH may co-ordinate transcription of histone gens with maturation of the 5' and 3' ends of histone mRNAs. Indeed, a mutant FLASH lacking the Ars2-intercating motif is inefficient in supporting proliferation and cell cycle progression of human pharyngeal carcinoma-derived KB cells, suggesting that MiniFLASH may only be partially proficient in vivo (Kiriyama et al., 2009). Clearly, the absence of this motif and other functional domains may also liberate MiniF from tight regulatory constrains, potentially resulting in important consequences for cell metabolism and phenotype.

Figure 23. 1. The C-terminal regions of FLASH and YARP interact with NPAT. A. A diagram of human FLASH consisting of 1982 amino acids. Positions of the conserved N- and C-terminal domains are indicated with gray boxes. **B.** Comparison of the amino acid sequences of the C-terminal domains of FLASH (top, amino acids 1895-1976) and YARP (bottom, amino acids 2123-2203). Conserved amino acids are shown in the middle. **C.** Sequence of the last 103 amino acids of human FLASH, with the C-terminal domain shared with YARP being underlined. **D.** The yeast two-hybrid screen identifies NPAT and PCNA as potential interactions of the FLASH C-terminal region that encompasses the last 103 amino acids. Growth of yeast cells expressing the C-terminal region of FLASH and either the last 16 amino acids of NPAT (left vertical row) or full length PCNA (right vertical row) in the presence of increasing concentration of 3-amino triazol (3-AT). **E.** Sequence of the last 16 amino acids of human NPAT fused to the N-terminal GST (GST-N16). **F.** Sequence of 97 amino acids of human YARP that are located near the C-terminal end and contain the C-terminal domain shared with FLASH (underlined). **G.** GST pull down assay to analyze *in vitro* interaction of ^{35}S-labeled C-terminal regions of FLASH (FL103C) or YARP (YA97C) with the last 16 amino acids of NPAT fused to GST (GST-N16). ^{35}S-labeled N-terminal region of FLASH (F138N, amino acids 1-138) and GST alone were used as negative controls. The amount of each GST protein purified on glutathione beads was monitored by staining the gel with Coomassie Blue (bottom panels). Note that GST alone migrates higher than GST-N16 due to containing random amino acids at the C-terminus from translating the multiple cloning site present in the pET42a vector.

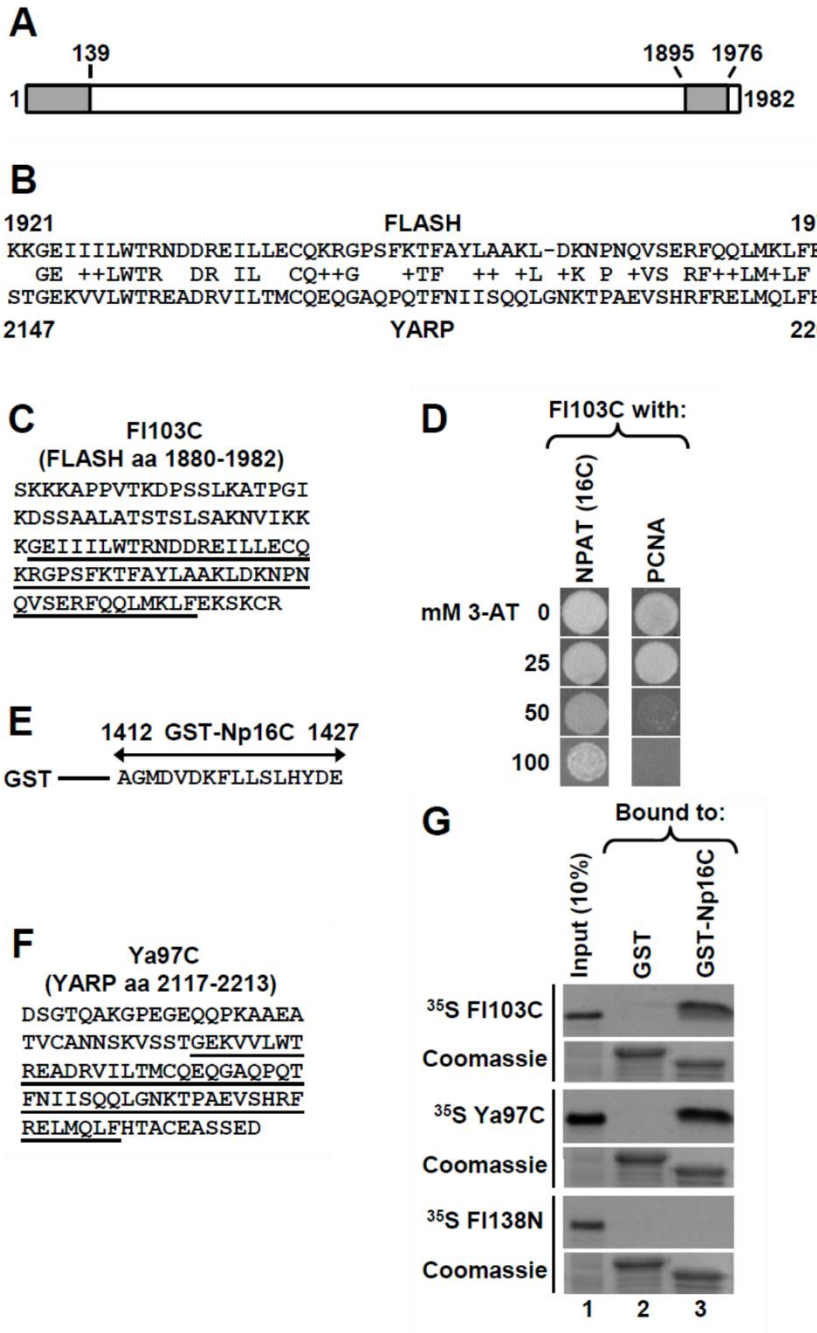

A
139
1895
1976
1
1982
B
1921
FLASH
1978
KKGEIIILWTRNDDREILLECQKRGPSFKTFAYLAAKL-DKNPNQVSERFQQLMKLFEK
GE ++LWTR DR IL CQ++G +TF ++ +L +K P +VS RF++LM+LF
STGEKVVLWTREADRVILTMCQEQGAQPQTFNIISQQLGNKTPAEVSHRFRELMQLFHT
2147
YARP
2205
C
Fl103C
(FLASH aa 1880-1982)
SKKKAPPVTKDPSSLKATPGI
KDSSAALATSTSLSAKNVIKK
KGEIIILWTRNDDREILLECQ
KRGPSFKTFAYLAAKLDKNPN
QVSERFQQLMKLFEKSKCR
D
Fl103C with:
NPAT (16C)
PCNA
mM 3-AT 0
25
50
100
E
1412 GST-Np16C 1427
GST
AGMDVDKFLLSLHYDE
F
Ya97C
(YARP aa 2117-2213)
DSGTQAKGPEGEQQPKAAEA
TVCANNSKVSSTGEKVVLWT
READRVILTMCQEQGAQPQT
FNIISQQLGNKTPAEVSHRF
RELMQLFHTACEASSED
G
Bound to:
Input (10%)
GST
GST-Np16C
35S Fl103C
Coomassie
35S Ya97C
Coomassie
35S Fl138N
Coomassie
1
2
3

Figure 24. 2. The C-terminal domain shared by FLASH and YARP is sufficient for the interaction with NPAT. A. Sequence of the last 131 amino acids of human NPAT (Np131C). The 16-amino acid extreme C-terminal region cloned in the yeast two-hybrid screen and capable of interacting with C-terminal regions of FLASH and YARP is underlined. The last amino acids of the truncated NpΔ16C and NpΔ25C proteins are indicated with vertical arrows. **B.** GST pull down assay to analyze *in vitro* interaction of ^{35}S-labeled Np131C or its truncated versions (NpΔ16C and NpΔ25C) with the C-terminal regions of FLASH (GST-Fl103C) or YARP (GST-Ya97C), each fused to the N-terminally located GST. GST alone (lane 2) was used as a negative control. The amount of each GST protein purified on glutathione beads was monitored by staining gels with Coomassie Blue and was equal in each lane (not shown). **C.** GST pull down assay to analyze the interaction of ^{35}S-labeled Np131C with GST-Fl103C or its shorter version containing only the 54-amino acid region of homology with YARP (GST-Fl54C). The amount of each GST protein purified on glutathione beads was monitored by staining the gel with Coomassie Blue (bottom panel). **D.** GST pull down assay to analyze competition between GST-Np16 (final concentration 2 μM) and two different amounts, as indicated, of the Np25C oligopeptide in binding ^{35}S-labeled Fl103C. GST alone (lane 2) was used as negative control. The amounts of the GST-Np16 and GST proteins purified on glutathione beads were monitored by staining the gel with Coomassie Blue (bottom panel). **E and F.** GST pull down assay to analyze *in vitro* interaction of ^{35}S-labeled full length PCNA with GST-Fl103C, GST-Ya97C or GST-Fl54C, as indicated at the top of each lane. GST alone (lane 2 in both panels) was used as a negative control. The amount of each protein purified on glutathione beads was monitored by staining the gel with Coomassie Blue (bottom panel in F and not shown).

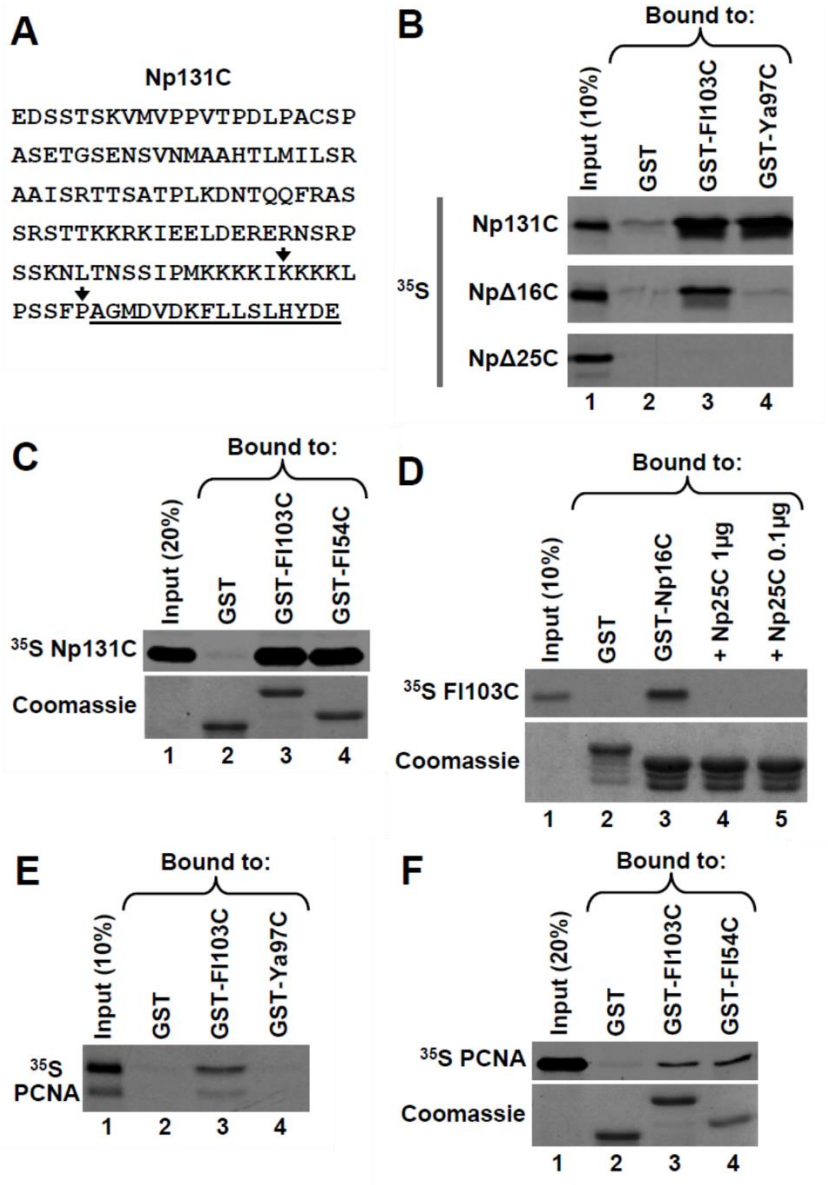

A
Np131C
EDSSTSKVMVPPVTPDLPACSP
ASETGSENSVNMAAHTLMILSR
AAISRTTSATPLKDNTQQFRAS
SRSTTKKRKIEELDERERNSRP
SSKNLTNSSIPMKKKKIKKKKL
PSSFPAGMDVDKFLLSLHYDE
B
Bound to:
Input (10%)
GST
GST-Fl103C
GST-Ya97C
35S
Np131C
NpΔ16C
NpΔ25C
1 2 3 4
C
Bound to:
Input (20%)
GST
GST-Fl103C
GST-Fl54C
35S Np131C
Coomassie
1 2 3 4
D
Bound to:
Input (10%)
GST
GST-Np16C
+ Np25C 1µg
+ Np25C 0.1µg
35S Fl103C
Coomassie
1 2 3 4 5
E
Bound to:
Input (10%)
GST
GST-Fl103C
GST-Ya97C
35S
PCNA
1 2 3 4
F
Bound to:
Input (20%)
GST
GST-Fl103C
GST-Fl54C
35S PCNA
Coomassie
1 2 3 4

Figure 25. 3. The C-terminal regions of FLASH and NPAT interact in HeLa cells. A. A schematic representation of FLASH and NPAT clones transiently expressed in HeLa cells. **B.** Nucleotide sequence of the Biot-SL RNA used to purify SLBP. **C.** Myc-Np131C was co-expressed in HeLa cells with either HA-SLBP (lanes 1-2) or HA-SLBP/F (lanes 3-4). SLBP was affinity-purified by the Biot-SL RNA from whole cell lysates prepared from transfected HeLa cells and the precipitated material tested for the presence of the transiently expressed HA- and Myc-tagged proteins (lanes 2 and 4). Lanes 1 and 3 contain 30% of the material used for purification. Myc-Np131C co-purified by virtue of interacting with HA-SLBP/F is indicated with an arrow. A protein cross-reacting with the anti-Myc antibody (indicated with an asterisk) served together with CPSF73 as a control to measure the extent of nonspecific background present in the purified material. **D.** Myc-mxc169C was co-expressed in HeLa cells with Ha-SLBP/F and the interaction between these two proteins tested as described in panel C. A protein cross-reacting with the anti-Myc antibody is indicated with an asterisk. **E.** Myc-Np131 was co-expressed in HeLa cells with either HA-SLBP (lanes 1-2) or Ha-SLBP/F (lanes 3-4). SLBP was precipitated by an anti-SLBP antibody and the precipitated material tested for the presences of the transiently expressed HA- and Myc-tagged proteins (lanes 2 and 4). A protein cross-reacting with the anti-Myc antibody is indicated with an asterisk

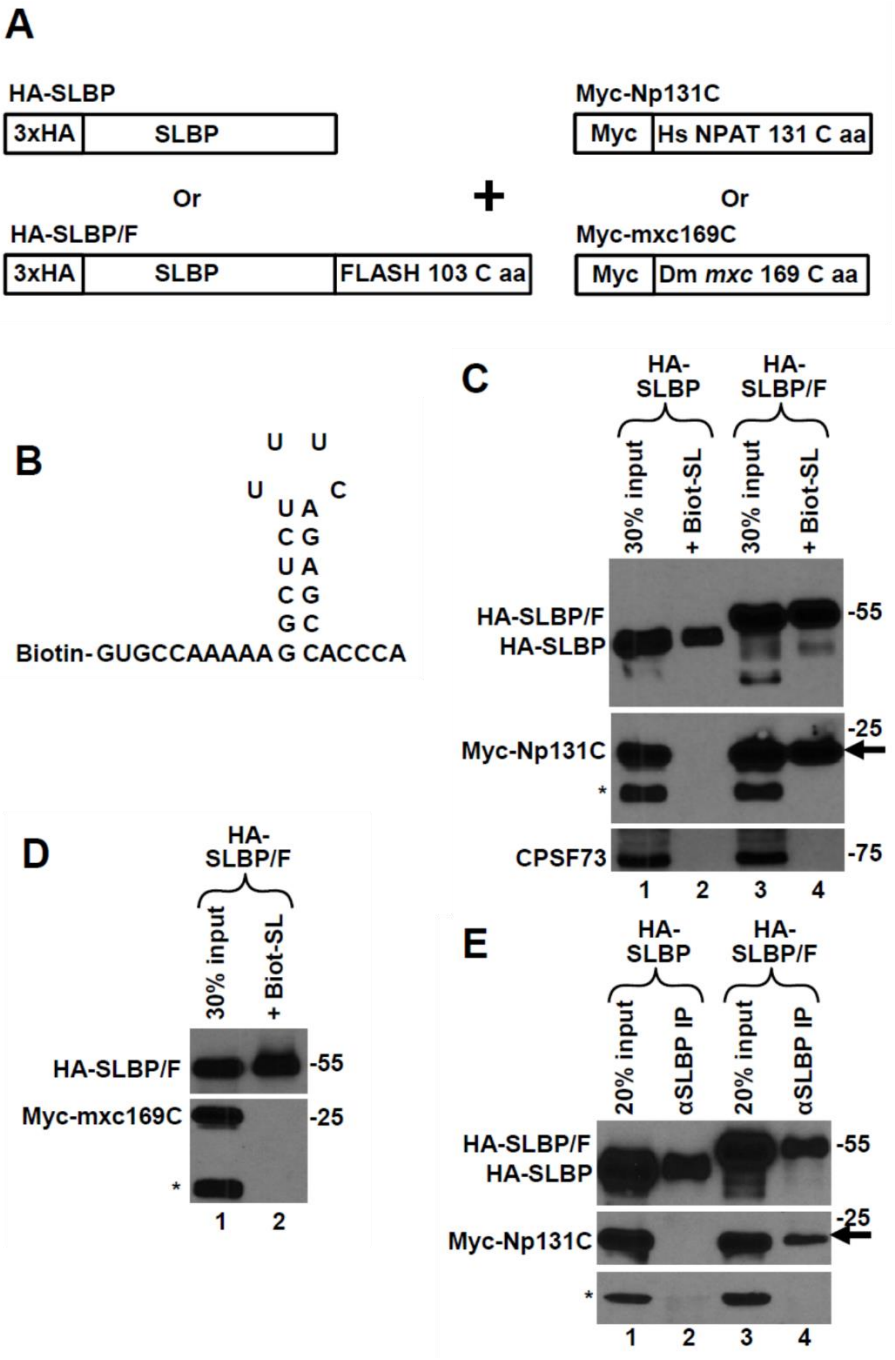

A
HA-SLBP
3xHA
SLBP
Or
HA-SLBP/F
3xHA
SLBP
FLASH 103 C aa
+
Myc-Np131C
Myc
Hs NPAT 131 C aa
Or
Myc-mxc169C
Myc
Dm mxc 169 C aa
B
Biotin-GUGCCAAAAA G CACCCA
C
HA-SLBP
HA-SLBP/F
30% input
+ Biot-SL
30% input
+ Biot-SL
HA-SLBP/F
HA-SLBP
-55
Myc-Np131C
-25
*
CPSF73
-75
1 2 3 4
D
HA-SLBP/F
30% input
+ Biot-SL
HA-SLBP/F
-55
Myc-mxc169C
-25
*
1 2
E
HA-SLBP
HA-SLBP/F
20% input
αSLBP IP
20% input
αSLBP IP
HA-SLBP/F
HA-SLBP
-55
Myc-Np131C
-25
*
1 2 3 4

Figure 26. 4. The C-terminal regions of FLASH targets SLBP to HLBs in HeLa cells. A. The HA-tagged versions of SLBP (HA-SLBP) or SLBP extended at the C-terminus by the last 103 amino acids of FLASH (HA-SLBP/F) were transiently expressed in HeLa cells and their levels compared relative to the endogenous SLBP using an anti-SLBP antibody. **B.** Immunofluorescence of HeLa cells transiently expressing HA-SLBP. HLBs were detected using a rabbit antibody against the N-terminal FLASH and stained green. HA-SLBP was detected using a mouse anti-HA antibody and stained red. Nuclei were visualized with DAPI. The top image includes three cells of which only two were transfected and express HA-SLBP. A magnified image of an independent transfected HeLa cells is shown at the bottom. **C.** Immunofluorescence of HeLa cells transiently expressing HA-SLBP/F (see above for details).

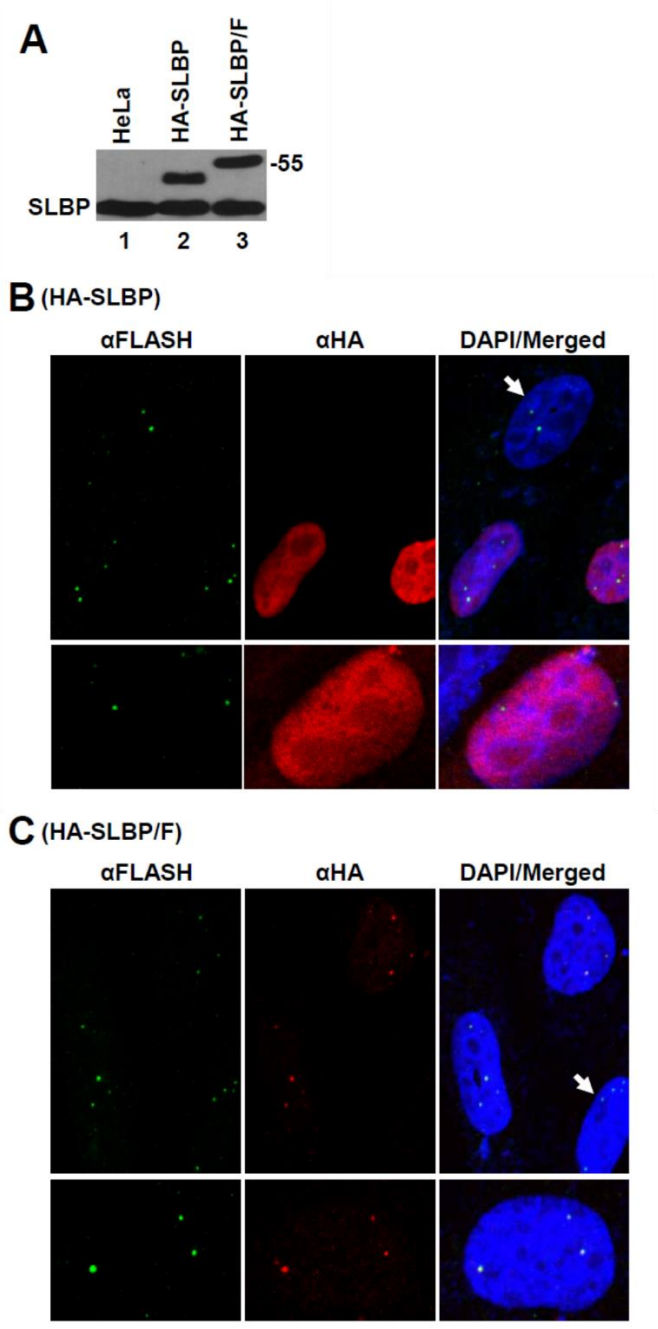

A
HeLa
HA-SLBP
HA-SLBP/F
-55
SLBP
1
2
3
B (HA-SLBP)
αFLASH
αHA
DAPI/Merged
C (HA-SLBP/F)
αFLASH
αHA
DAPI/Merged

Figure 27. 5. YARP is a component of HLBs in HeLa cells. A. Myc-Np131C was co-expressed in HeLa cells with either HA-SLBP (lanes 1-2), HA-SLBP/F (lanes 3-4) or HA-SLBP/Y (lanes 5-6). SLBP was affinity-purified by the Biot-SL RNA from whole cell lysates prepared from transfected HeLa cells and the precipitated material tested for the presence of the transiently expressed HA- and Myc-tagged proteins (lanes 2, 4 and 6). Lanes 1, 3 and 5 contain 15% of the material used for purification. Proteins cross-reacting with the anti-Myc antibody are indicated with asterisks. **C.** A fraction of SLBP in HeLa cells transiently expressing HA-SLBP/Y localizes to HLBs. SLBP was detected by an anti-SLBP antibody and stained green. HLBs were detected using a mouse monoclonal antibody against NPAT and stained red

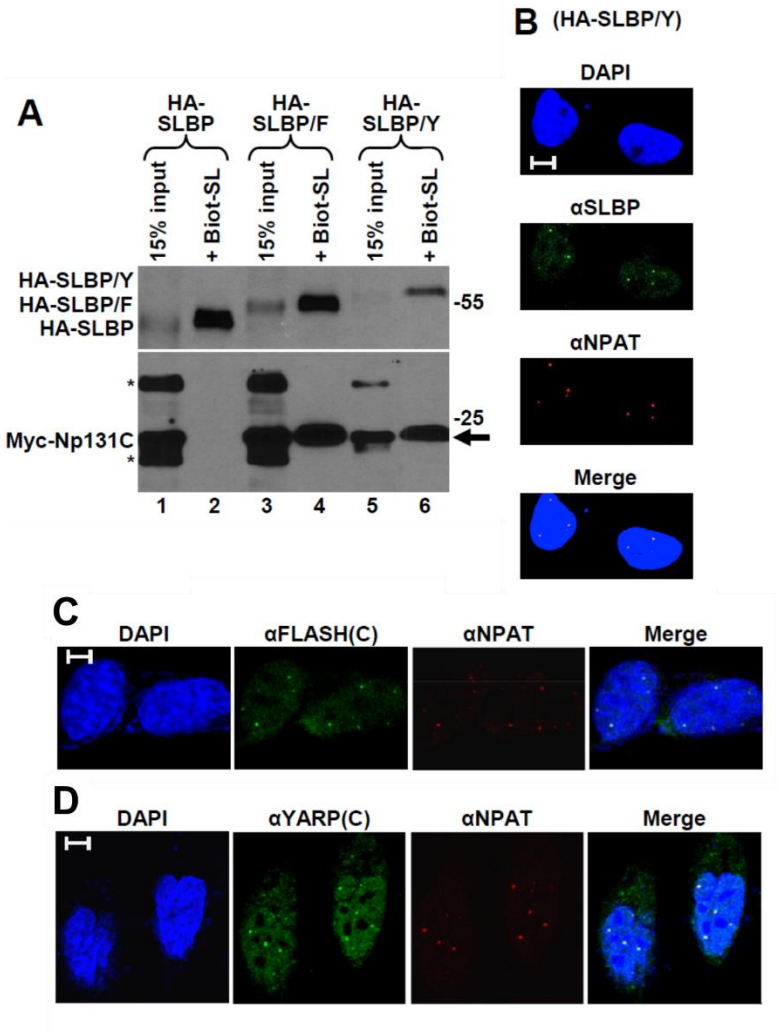
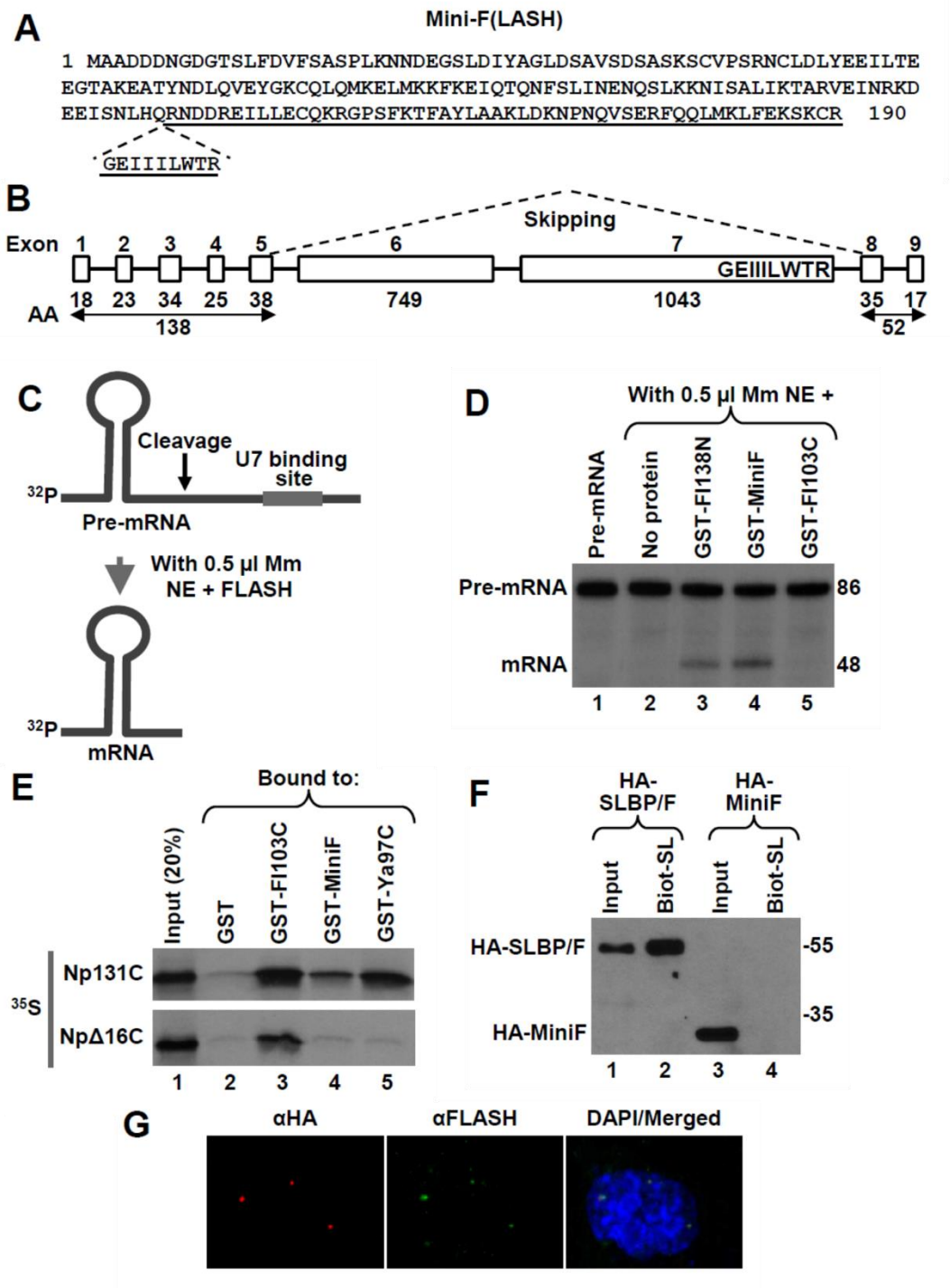

A
HA-SLBP
HA-SLBP/F
HA-SLBP/Y
15% input
+ Biot-SL
HA-SLBP/Y
HA-SLBP/F
HA-SLBP
-55
-25
Myc-Np131C
1 2 3 4 5 6
B (HA-SLBP/Y)
DAPI
αSLBP
αNPAT
Merge
C
DAPI
αFLASH(C)
αNPAT
Merge
D
DAPI
αYARP(C)
αNPAT
Merge

Figure 28. 6. Identification of human MiniFlash (MiniF). A. The amino acid sequence of 190-amino acid human MiniF. The incomplete C-terminal domain is underlined and the missing GEIIILWTR sequence is shown below. **B.** The exon/intron organization of the Open Reading Frame of human FLASH. The number of amino acids (AA) encoded by individual exons 1-9 of FLASH and those spliced together in MiniF are shown at the bottom. The MiniF cDNA results from skipping of the two large exons 6 and 7, as depicted by dashed lines, and in-frame linking of the N- and C-terminal exons. The highly conserved GEIIILWTR sequence of the C-terminal domain of FLASH that is encoded by the end of exon 7 and hence missing in MiniF is indicated. **C.** GST-pull down assay to analyzed the interaction between MiniF and the C-terminal region of NPAT. Proteins shown at the top of lanes 2-5 were bacterially expressed as fusion with GST, incubated with ^{35}S-labeled Np131C or NpΔ16, and the amount of each radioactive protein absorbed on glutathione beads was determined by autoradiography. As determined by Coomassie Blue staining, the amount of GST proteins in each lane was comparable (not shown). Lane 1 contains 20% of each radioactive protein used in the experiment. **D.** Transient expression of HA-MiniF in HeLa cells was compared to expression of HA-SLBP/F. As expected, only HA-SLBP/C can be purified by binding to the stem-loop RNA containing biotin at the 5' end (Biot-SL). **F.** Immunofluorescence of HeLa cells transiently expressing HA-MiniF. Localization of HA-MiniF was detected by a mouse anti-HA antibody and stained red. HLBs were detected by an antibody against the N-terminal FLASH and stained green. Nuclei were visualized by staining with DAPI.


A
Mini-F(LASH)
1 MAADDDNGDGTSLFDVFSASPLKNNDEGSLDIYAGLDSAVSDSASKSCVPSRNCLDLYEEILTE
EGTAKEATYNDLQVEYGKCQLQMKELMKKFKEIQTQNFSLINENQSLKKNISALIKTARVEINRKD
EEISNLHQRNDDREILLECQKRGPSFKTFAYLAAKLDKNPNQVSERFQQLMKLFEKSKCR 190
GEIIILWTR
B
Exon 1 2 3 4 5 6 7 8 9
Skipping
GEIIILWTR
AA 18 23 34 25 38 749 1043 35 17
138
52
C
Cleavage
U7 binding site
32P
Pre-mRNA
With 0.5 µl Mm NE + FLASH
32P
mRNA
D
With 0.5 µl Mm NE +
Pre-mRNA
No protein
GST-FI138N
GST-MiniF
GST-FI103C
Pre-mRNA
mRNA
86
48
1 2 3 4 5
E
Bound to:
Input (20%)
GST
GST-FI103C
GST-MiniF
GST-Ya97C
35S
Np131C
NpΔ16C
1 2 3 4 5
F
HA-SLBP/F
HA-MiniF
Input
Biot-SL
Input
Biot-SL
HA-SLBP/F
HA-MiniF
-55
-35
1 2 3 4
G
αHA
αFLASH
DAPI/Merged

CHAPTER 5: SUMMARY

In metazoans, 3’ end formation of RD histone mRNA is a critical step in the upregulation of histone gene expression in S-phase. This step accounts for a significant portion of the cell-cycle dependent increases in mRNA levels. Histone pre-mRNA 3’ end processing is U7-dependent and occurs by a single endonucleolytic cleavage between two sequence elements in the 3’-UTR, a conserved stem-loop structure and a histone downstream element (HDE) 3’ of the cleavage site. SLBP binds the stem-loop and the HDE recruits the U7 snRNP via partial base pairing with the U7 snRNA 5’ end. Binding of SLBP to the stem-loop is somewhat dispensable in mammals if the duplex formed between the HDE and U7 snRNP is sufficiently strong, but absolutely essential for 3’ end processing in *Drosophila*. In mammals, cleavage occurs 4-5 nts 3’ of the stem-loop while this distance is always 4 nts in *Drosophila*.

Chapter 2 shows that a unique subset of polyadenylation factors, including the endonuclease, stably associates with FLASH and the U7 snRNP in mammalian extracts. This complex is named the histone pre-mRNA cleavage complex (HCC) and is recruited to histone pre-mRNA in a U7 dependent fashion. Chapter 2 also identifies the region of human Lsm11 that binds FLASH and examines the elements of Lsm11 required to isolate the HCC *in vitro*. Chapter 3 characterizes the HCC in *Drosophila* and demonstrates the functional conservation of elements in *Drosophila* FLASH required for processing. Additionally, I examine the role of FLASH in *Drosophila* downstream cleavage product (DCP) degradation. Chapter 4 shows an interaction between the C-terminal regions of FLASH and NPAT is sufficient to localize FLASH to HLBs

in mammalian cells. I also describe a splice variant of FLASH that contains the N-terminal region necessary for 3' end processing fused directly to the extreme C-terminus required for binding NPAT.

Biochemical isolation and identification of the HCC. Histone pre-mRNA 3' end processing is U7-dependent. Although CPSF73 is known to be the endonuclease required for cleavage, precisely how it is recruited to histone pre-mRNA was unknown. Based on available literature, it was hypothesized the unique 170 N-terminal amino acids of Lsm11 were essential for recruitment of additional 3' end processing factors. N-terminal Lsm11 forms a tight complex with FLASH, a processing factor identified in our lab and whose N-terminal region is critical for histone pre-mRNA 3' end processing. Both of these proteins are unique to U7-dependent processing. By forming a stable Lsm11/FLASH complex, we hoped to biochemically isolate additional processing factors that depend on this interaction *in vitro*.

Dr. Dominski devised a single step biochemical purification strategy using the critical regions of N-terminally tagged recombinant Lsm11 (amino acids 1-168) and FLASH (amino acids 29-139) to form a stable complex and identify factors that associate with this platform in mammalian nuclear extracts. This approach successfully isolated the known histone pre-mRNA processing factors CPSF73, CPSF100, and symplekin in addition to a number of polyadenylation factors not previously associated with histone pre-mRNA 3' end processing. Western and mass spectrometry analysis of the material that co-purified with the Lsm11/FLASH complex revealed a distinct subset of polyadenylation factors. The newly identified complex is unique to histone 3' end processing and is collectively named the histone pre-mRNA cleavage complex (HCC). The HCC contains all six CPSF subunits, symplekin and CstF64. Conspicuously missing are the other subunits of CstF, CstF50 and CstF77. Mutational analysis showed the interaction between

the Lsm11/FLASH complex and HCC depends on a highly conserved LDLY motif (amino acids 55-58) in FLASH. The HCC resembles the HLF complex purified and described by Nikolay Kolev and Joan Steitz with the exception of CstF77. But this was the first time a biochemical purification protocol relying on factors unique to histone 3' end processing was able to isolate the HCC.

I mapped the regions of human N-terminal Lsm11 necessary for interacting with FLASH and determined amino acids 19 to 55 of Lsm11 are required to form a tight complex. While N-terminal Lsm11 (amino acids 1-168) or N-terminal FLASH (amino acids 28-139) are individually capable of interacting weakly with components of the HCC, together they form a stable platform for efficient binding of the HCC. In FLASH, the essential residues are in the highly conserved LDLY motif, while in Lsm11 the region required for cooperative binding of the HCC is not as clearly defined but lies between amino acids 66 and 130. These results indicate elements of both FLASH and Lsm11 are required for recruitment of the HCC and it is only upon forming a Lsm11/FLASH complex that a repositioning of critical domains occurs and a robust interaction with the HCC is possible. One of the significant questions remaining to be answered is what factor(s) make direct contact with the conserved LDLY motif.

The association of the HCC and FLASH with the endogenous U7 snRNP was shown in a single step purification of the snRNP from nuclear extracts using a biotinylated oligonucleotide complimentary to the 5' end of U7 snRNA. These studies revealed the composite nature of the U7 snRNP. Adding recombinant FLASH to the mixture modestly enhanced the ability of the U7 snRNP to interact with the HCC in HeLa extracts, suggesting FLASH may be limiting in these extracts. Using a biotinylated H2a pre-mRNA, I was able to demonstrate the entire composite U7 snRNP is recruited to histone pre-mRNA in a U7-dependent manner. The techniques developed

in these investigations were useful in my subsequent efforts to characterize the U7 snRNP in *Drosophila*.

***Drosophila* HCC contains a different set of polyadenylation factors.** I identified a *Drosophila* HCC that is similar in composition to the mammalian one. The *Drosophila* HCC contains symplekin, CstF64, and four subunits of CPSF (CPSF73, CPSF100, CPSF160, and WDR33). Unlike the mammalian HCC, the *Drosophila* complex does not include CPSF30 or Fip1. As in vertebrates, the HCC stably associates with the U7 snRNP and the larger complex can be isolated from nuclear extracts as a composite structure in a single step purification using an anti-U7 oligonucleotide. CPSF73, CPSF100, and symplekin are the only members of the HCC essential for 3' end processing. The exact role of the remaining members of the HCC in processing is not clear.

While cleavage of *Drosophila* histone pre-mRNA depends on binding of SLBP, recruitment of the composite U7 snRNP to the pre-mRNA does not require SLBP binding nor the stem-loop structure itself. This was demonstrated *in vitro* using a SLBP stem-loop oligo competitor to sequester SLBP in the presence of biotinylated *Drosohphila* H3 pre-mRNA or a biotinylated H3 pre-mRNA in which the stem-loop had been removed. Under both conditions, the composite U7 snRNP containing the HCC was recruited to the pre-mRNA. These results suggest SLBP binding of the stem-loop serves an additional role in 3' end processing not related to recruitment of the composite U7 snRNP. Additional investigations into SLBPs processing role could better elucidate the mechanisms behind activation of the endonuclease.

***Drosophila* HCC differs from a larger super complex specific to cleavage and polyadenylation.** Curiously, CstF64 is the only subunit from CstF found in the HCC. CstF50 and CstF77 do not co-purify with the complex. Using an antibody raised against CstF64, I

conducted immunoprecipitation experiments with *Drosophila* Kc extracts. A large macromolecular complex co-precipitated with CstF64. This complex contained the six subunits of CPSF, both subunits of CF I_m, symplekin, and all three subunits of CstF, including CstF50 and Cst77. CF II_m was the only cleavage and polyadenylation factor not detected, although it is possible that it co-purified in undetectable, sub-stoichiometric amounts. Treating the extracts with RNase A did not disrupt the complex, indicating protein-protein interactions are primarily responsible for its formation.

This was the first time that a large complex containing this set of cleavage and polyadenylation factors was isolated. The HCC contrasts sharply with the super complex of polyadenylation factors. The absence of CstF50 and CstF77 in the HCC, but inclusion in the polyadenylation complex, suggests a complexity in the connectivity between subunits. This could partially be explained by the mutually exclusive binding of symplekin by CstF64 and CstF77.

Functional elements of FLASH required for 3' end processing are conserved in *Drosophila*. *Drosophila* Lsm11 and FLASH also form a tight complex. In mammalian FLASH, a LDLY motif located at amino acids 55-58 is critical for histone pre-mRNA 3' processing and recruitment of the HCC. I examined the corresponding LDIY motif (amino acids 71-74) in *Drosophila* FLASH and showed a similar dependence. To examine the role of the LDIY motif in processing, I prepared small scale nuclear extracts from *Drosophila* S2 cells depleted of FLASH by RNA interference. *In vitro* 3' end processing assays with 5' radio-labeled H3 pre-mRNA showed the depleted extracts had minimal processing activity compared to extracts prepared from untreated cells. Adding bacterially expressed *Drosophila* FLASH (amino acids 1-178) to depleted extracts not only restored processing, but stimulated it to levels exceeding that of non-

depleted extracts. In contrast, an N-terminal fragment of FLASH with the conserved LDIY motif (amino acids 71-74) mutated to alanines severely limited the ability of the recombinant protein to stimulate processing *in vitro*. The stimulatory effect of the mutant FLASH was approximately 5-10% of that seen with the wild type fragment in depleted extracts. Unpublished data from Deirdre Tatomer indicated another conserved LDIY motif (amino acids 45-48) is sufficient for viability when combined with the LDIY to AAAA mutation at amino acids 71-74. While the LDIY motif at amino acids 71-74 is primarily responsible for 3' end processing *in vitro*, Deirdre's result demonstrated a redundant role for the minor LDIY motif *in vivo*. My subsequent experiments showed mutating both conserved LDIY motifs (amino acids 45-48 and 71-74) to alanines completely abolished the ability of recombinant FLASH to restore processing in depleted extracts.

I also showed the LDIY motif (amino acids 71-74) has a role in recruiting polyadenylation factors known to participate in histone pre-mRNA 3' end processing. In Kc nuclear extracts, bacterially expressed *Drosophila* Lsm11 (amino acids 1-153) bound to FLASH (amino acids 1-178) interacts with a complex containing the endonuclease CPSF73, CPSF100, and symplekin. These factors co-purify with a Lsm11/FLASH complex in a LDIY (amino acids 71-74) dependent manner. Western analysis of the purified material showed changing these residues to alanines severely hinders the ability of the Lsm11/FLASH complex to interact with these essential processing factors. Despite the LDIY mutation, there are trace amounts of CPSF73, CPSF100, and symplekin that co-purify with the Lsm11/FLASH LDIY71AAA mutant. The minor LDIY motif (amino acids 45-48) could play a redundant role in binding these factors though with far less efficiency. Additional pull down experiments using a Lsm11/FLASH

complex containing the double mutant of FLASH could help clarify whether the minor LDIY (amino acids 45-48) has a contribution to recruiting processing factors

FLASH is required for DCP degradation in *Drosophila*. Degradation of the downstream cleavage product (DCP) in mammals is U7-dependent and does not require SLBP. The DCP is degraded in the 5' to 3' direction exonucleolitically by CPSF73, the endonuclease in histone and canonical pre-mRNA 3' end processing (Yang et al., 2009a). *In vitro* degradation assays using mouse nuclear extracts with a 5' radio-labeled DCP show FLASH is not required for degradation. The addition of purified anti-FLASH antibody does not disrupt normal degradation nor does adding recombinant N-terminal FLASH stimulate the reactions. In contrast, I demonstrated an essential role for FLASH in *Drosophila* DCP degradation.

To check whether *Drosophila* FLASH participates in the degradation reaction, I conducted *in vitro* degradation assays comparing nuclear extracts from untreated S2 cells and FLASH depleted S2 cells. In extracts that were not depleted, degradation proceeded in a U7-dependent fashion. Blocking the interaction between the HDE and 5' end of the U7 snRNA inhibited the reaction. Adding N-terminal recombinant FLASH to the extract resulted in a 3-4 fold increase in degradation of the DCP. Adding purified anti-FLASH antibody to the extracts completely inhibited activity.

FLASH depleted extracts showed only residual degradation activity, substantially lower than their untreated counterparts. Degradation activity in the depleted extracts was restored to levels surpassing extracts containing endogenous FLASH upon addition of N-terminal recombinant FLASH. Mutating the conserved LDIY motif (amino acids 71-74) resulted in a reduction of degradation rescue by approximately 50% compared to the addition of wild type N-terminal FLASH. This loss of activity is less severe than that seen in 3' end processing assays

where changing the LDIY causes a 90% reduction of FLASH activity. It is possible the minor LDIY motif (amino acids 45-48) is more important for DCP degradation than it is for 3' end processing but this needs to be tested.

The different requirements for FLASH in DCP degradation between mammals and insects could be related to the ability of mammalian Lsm11 to weakly interact with the HCC on its own, without binding FLASH. This weak interaction may be sufficient to recruit CPSF73 to the DCP for exonucleolytic activity. In *Drosophila*, retention of HCC after 3' end processing likely requires the continued interaction between Lsm11 and FLASH. The added contribution of FLASH may stabilize the interaction of the U7 snRNP with the HCC in a manner that promotes the switch from endo- to exonucleolytic mode for DCP degradation.

An interaction between C-terminal FLASH and NPAT localizes FLASH to HLBs. A 55 amino acid domain located near the C-terminus of human FLASH (amino acids 1923-1975) is highly conserved in vertebrates and shares 50% sequence identity with a C-terminal domain of an unrelated protein, YARP. The shared region is predicted to contain three α-helices with an overall fold that resembles SANT and Myb domains, hence named the SANT/Myb-like domain. YARP is an orthologue of the *Drosophila* protein Mute. In *Drosophila*, Mute acts as a regulator of histone gene transcription and localizes to HLBs (Bulchand et al. 2010).

Lalitha Kunduru conducted a yeast two-hybrid screen using the C-terminal 103 amino acids of FLASH, which includes the SANT/Myb-like domain, with a HeLa S3 library and identified NPAT as a binding partner of FLASH. NPAT is a constitutive member of HLBs and co-activator of histone gene transcription. It is required for the G1 to S-phase transition (Ma et al., 2000) and is essential for histone gene transcription (Zhao et al., 2000). Using deletion analysis and peptide competitors, I showed the SANT/Myb-like domain near the C-terminus of

FLASH interacts with the C-terminal 25 amino acids of NPAT *in vitro*. The SANT/Myb-like domain is conserved in YARP but the interaction with NPAT is weaker..

SLBP remains associated with histone mRNA after 3' end processing and thus immunofluorescent staining for endogenous SLBP in HeLa cells shows a diffuse pattern throughout the nucleus and cytoplasm. Appending the last 103 amino acids of FLASH to the C-terminus of SLBP *in vivo* results in localization of the chimeric protein to HLBs. A similar pattern was observed when SLBP was extended with the C-terminal region of YARP (amino acids 2117-2213).

MiniFLASH is a splice variant of FLASH that joins two regions essential for histone biogenesis. Mammalian FLASH is a large protein of 1982 amino acids that is encoded by 9 exons. Yet there is no published data to indicate the existence of alternatively spliced FLASH variants. I conducted a yeast two-hybrid screen using N-terminal Lsm11 with a human universal tissue library and identified a FLASH splice variant containing the N-terminus fused directly and in frame to the C-terminus. Comprising only 190 amino acids, this variant was named MiniFLASH. Surprisingly, it includes the first 138 amino acids shown to be essential for histone pre-mRNA 3' end processing *in vitro* and the C-terminal 52 amino acids required for binding NPAT and localizing FLASH to HLBs. The splice variant consists of exons 1-5 spliced to exons 8 and 9. The two largest exons, 6 and 7, have been skipped.

Fusing the extreme N- and C-terminal domains of FLASH into MiniFLASH may have biological relevance *in vivo*. In *Drosophila*, FLASH is 844 amino acids and essential for viability (Burch et al., 2011). In a mutant fly deficient for endogenous FLASH, a *Drosophila* MiniFLASH composed of the N- and C-termini rescues viability and molecular defects, including mis-processing of histone pre-mRNA, despite lacking 550 internal amino acids (Tatomer et al.

Submitted). This suggests MiniFLASH can at a minimum compensate for full length FLASH's role in the biogenesis of histone mRNA in *Drosophila*. The regulatory consequences of removing such a large internal portion of the protein are not entirely clear. In mammalian cells, cell-cycle progression and cellular proliferation were shown to be dependent on an interaction between the internal region of FLASH and Ars2. But human MiniFLASH does not include the region of FLASH that binds Ars2. Interestingly, I showed *Drosophila* FLASH and Ars2 do not interact or associate in a complex together, suggesting this interaction is not critical in *Drosophila*. Identification of MiniFLASH has provoked new ideas regarding FLASH as a 3' end processing factor and, in our lab, has led to fresh investigations regarding the potential biological significance of this splice variant.

REFERENCES

Abbott, J., Marzluff, W.F., and Gall, J.G. (1999). The stem loop binding protein (SLBP1) is present in coiled bodies of the Xenopus germinal vesicle. Mol.Biol.Cell *10*, 487–499.

Alm-Kristiansen, A.H., Saether, T., Matre, V., Gilfillan, S., Dahle, O., and Gabrielsen, O.S. (2008). FLASH acts as a co-activator of the transcription factor c-Myb and localizes to active RNA polymerase II foci. Oncogene *27*, 4644–4656.

Aravind, L. (1999). An evolutionary classification of the metallo-beta-lactamase fold proteins. In Silico.Biol. *1*, 69–91.

Azzouz, T.N., Gruber, A., and Schumperli, D. (2005). U7 snRNP-specific Lsm11 protein: dual binding contacts with the 100 kDa zinc finger processing factor (ZFP100) and a ZFP100-independent function in histone RNA 3' end processing. Nucleic Acids Res. *33*, 2106–2117.

Bai, Y., Auperin, T.C., Chou, C.Y., Chang, G.G., Manley, J.L., and Tong, L. Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. Mol.Cell *25*, 863–875.

Barabino, S.M.L., Ohnacker, M., and Keller, M. Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. EMBO J. *19*, 3778–3787.

Barcaroli, D., Bongiorno-Borbone, L., Terrinoni, A., Hofmann, T.G., Rossi, M., Knight, R.A., Matera, A.G., Melino, G., and De V, L. FLASH is required for histone transcription and S-phase progression. Proc.Natl.Acad.Sci.U.S.A *103*, 14808–14812.

Barcaroli, D., Dinsdale, D., Neale, M.H., Bongiorno-Borbone, L., Ranalli, M., Munarriz, E., Sayan, A.E., McWilliam, J.M., Smith, T.M., Fava, E., et al. FLASH is an essential component of Cajal bodies. Proc.Natl.Acad.Sci.U.S.A *103*, 14802–14807.

Barcaroli, D., Dinsdale, D., Neale, M.H., Bongiorno-Borbone, L., Ranalli, M., Munarriz, E., Sayan, a E., McWilliam, J.M., Smith, T.M., Fava, E., et al. (2006). FLASH is an essential component of Cajal bodies. Proc. Natl. Acad. Sci. U. S. A. *103*, 14802–14807.

De Biasio, A., and Blanco, F.J. (2013). Proliferating cell nuclear antigen structure and interactions: Too many partners for one dancer? Adv. Protein Chem. Struct. Biol. *91*, 1–36.

Billings, P.B., Allen, R.W., Jensen, F.C., and Hoch, S.O. (1982). Anti-RNP monoclonal antibodies derived from a mouse strain with lupus-like autoimmunity. J. Immunol. *128*, 1176–1180.

Birchmeier, C., Grosschedl, R., and Birnstiel, M.L. (1982). Generation of authentic 3' termini of an H2A mRNA in vivo is dependent on a short inverted DNA repeat and on spacer sequences. Cell *28*, 739–745.

Birnstiel, M.L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3' processing: the end is in site! Cell *41*, 349–359.

Bolisetty, M.T., and Beemon, K.L. (2012). Splicing of internal large exons is defined by novel cis-acting sequence elements. Nucleic Acids Res. *40*, 9244–9254.

Bond, U.M., Yario, T.A., and Steitz, J.A. (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.

Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? Nat. Rev. Mol. Cell Biol. *5*, 158–163.

Brown, K.M., and Gilmartin, G.M. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol.Cell *12*, 1467–1476.

Bulchand, S., Menon, S.D., George, S.E., and Chia, W. Muscle wasted: a novel component of the Drosophila histone locus body required for muscle integrity. J. Cell Sci. *123*, 2697–2707.

Burch, B.D., Godfrey, A.C., Gasdaska, P.Y., Salzler, H.R., Duronio, R.J., Marzluff, W.F., and Dominski, Z. (2011). Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila . RNA *17*, 1132–1147.

Callebaut, I., Moshous, D., Mornon, J.-P., and de Villartay, J.-P. (2002). Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. Nucleic Acids Res. *30*, 3592–3601.

Choi, Y.H., Kim, K.B., Kim, H.H., Hong, G.S., Kwon, Y.K., Chung, C.W., Park, Y.M., Shen, Z.J., Kim, B.J., Lee, S.Y., et al. FLASH coordinates NF-kappa B activity via TRAF2. J. Biol. Chem. *276*, 25073–25077.

Cohn, R.H., Lowry, J.C., and Kedes, L.H. (1976). Histone genes of the sea urchin (S. purpuratus) cloned in E. coli; order, polarity and strandedness of the five histone coding and spacer regions. Cell *9*, 147–161.

Dominski, Z. Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. Crit Rev.Biochem.Mol.Biol. *42*, 67–93.

Dominski, Z. (2010). The hunt for the 3' endonuclease. Wiley Interdiscip. Rev. RNA *1*, 325–340.

Dominski, Z., Sumerel, J., Hanson, R.J., and Marzluff, W.F. (1995). The polyribosomal protein bound to the 3' end of histone mRNA can function in histone pre-mRNA processing. RNA *1*, 915–923.

Dominski, Z., Zheng, L.-X., Sanchez, R., and Marzluff, W.F. (1999). The stem-loop binding protein facilitates 3' end formation by stabilizing U7 snRNP binding to the histone pre-mRNA. Mol. Cell. Biol. *19*, 3561–3570.

Dominski, Z., Yang, X., Raska, C.S., Santiago, C.S., Borchers, C.H., Duronio, R.J., and Marzluff, W.F. (2002). 3' end processing of Drosophila histone pre-mRNAs: Requirement for phosphorylated dSLBP and co-evolution of the histone pre-mRNA processing system. Mol. Cell. Biol. *22*, 6648–6660.

Dominski, Z., Yang, X., Kaygun, H., and Marzluff, W.F. (2003). A 3' exonuclease that specifically interacts with the 3' end of histone mRNA. Mol. Cell *12*, 295–305.

Dominski, Z., Yang, X.C., Purdy, M., Wagner, E.J., and Marzluff, W.F. (2005a). A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. Mol. Cell. Biol. *25*, 1489–1500.

Dominski, Z., Yang, X., Purdy, M., and Marzluff, W.F. (2005b). Differences and similarities between Drosophila and mammalian 3' end processing of histone pre-mRNAs. RNA *11*, 1835–1847.

Dominski, Z., Yang, X.C., and Marzluff, W.F. (2005c). The polyadenylation factor CPSF-73 is involved in histone pre-mRNA processing. Cell *123*, 37–48.

Eliassen, K.A., Baldwin, A., Sikorski, E.M., and Hurt, M.M. (1998). Role for a YY1 binding site in replication-dependent mouse histone gene expression. Mol. Cell. Biol. *18*, 7106–7118.

Erkmann, J.A., Wagner, E.J., Dong, J., Zhang, Y.P., Kutay, U., and Marzluff, W.F. (2005). Nuclear import of the stem-loop binding protein and localization during the cell cycle. Mol.Biol.Cell *16*, 2960–2971.

Frey, M.R., and Matera, A.G. Coiled bodies contain U7 small nuclear RNA and associate with specific DNA sequences in interphase human cells [published erratum appears in Proc Natl Acad Sci U S A 1995 Aug 29;92(18):8532]. Proc. Natl. Acad. Sci. U. S. A. *92*, 5915–5919.

Ghule, P.N., Dominski, Z., Yang, X.C., Marzluff, W.F., Becker, K.A., Harper, J.W., Lian, J.B., Stein, J.L., Van Wijnen, A.J., and Stein, G.S. Staged assembly of histone gene expression machinery at subnuclear foci in the abbreviated cell cycle of human embryonic stem cells. Proc.Natl.Acad.Sci.U.S.A *44*, 16964–16969–.

Gick, O., Krämer, A., Keller, W., and Birnstiel, M.L. (1986). Generation of histone mRNA 3' ends by endonucleolytic cleavage of the pre-mRNA in a snRNP-dependent in vitro reaction. EMBO J. *5*, 1319–1326.

Gick, O., Krämer, a, Vasserot, a, and Birnstiel, M.L. (1987). Heat-labile regulatory factor is required for 3' processing of histone precursor mRNAs. Proc. Natl. Acad. Sci. U. S. A. *84*, 8937–8940.

Gilmartin, G.M. Eukaryotic mRNA 3’ processing: a common means to different ends. Genes Dev. *19*, 2517–2521.

Gilmartin, G.M., Schaufele, F., Schaffner, G., and Birnstiel, M.L. (1988). Functional analysis of the sea urchin U7 small nuclear RNA. Mol. Cell. Biol. *8*, 1076–1084.

Godfrey, A.C., Kupsco, J.M., Burch, B.D., Zimmerman, R.M., Dominski, Z., Marzluff, W.F., and Duronio, R.J. (2006). U7 snRNA mutations in Drosophila block histone pre-mRNA processing and block oogenesis. RNA *12*, 396–409.

Gorgoni, B., Andrews, S., Schaller, A., Schumperli, D., Gray, N.K., and Muller, B. The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. RNA *11*, 1030–1042.

Grimm, C., Stefanovic, B., and Schümperli, D. (1993). The low abundance of U7 snRNA is partly determined by its Sm binding site. EMBO J. *12*, 1229–1238.

Gruber, A.R., Martin, G., Keller, W., and Zavolan, M. (2014). Means to an end: Mechanisms of alternative polyadenylation of messenger RNA precursors. Wiley Interdiscip. Rev. RNA *5*, 183–196.

Gruber, J.J., Zatechka, D.S., Sabin, L.R., Yong, J., Lum, J.J., Kong, M., Zong, W.X., Zhang, Z., Lau, C.K., Rawlings, J., et al. Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. Cell *138*, 328–339.

Gruber, J.J., Olejniczak, S.H., Yong, J., La, R.G., Dreyfuss, G., and Thompson, C.B. Ars2 promotes proper replication-dependent histone mRNA 3’ end formation. Mol Cell *45*, 87–98.

Hatton, L.S., Eloranta, J.J., Figueiredo, L.M., Takagaki, Y., Manley, J.L., and O’Hare, K. The Drosophila homologue of the 64 kDa subunit of cleavage stimulation factor interacts with the 77 kDa subunit encoded by the suppressor of forked gene. Nucleic Acids Res. *28*, 520–526.

Hawkins, J.D. (1988). A survey on intron and exon lengths. Nucleic Acids Res. *16*, 9893–9908.

Helmling, S., Zhelkovsky, A., and Moore, C.L. (2001). Fip1 regulates the activity of Poly(A) polymerase through multiple interactions. Mol. Cell. Biol. *21*, 2026–2037.

Horton, J.E., Elgar, S.J., Khan, S.I., Zhang, X., Wade, P.A., and Cheng, X. (2007). Structure of the SANT domain from the Xenopus chromatin remodeling factor ISWI. Proteins Struct. Funct. Genet. *67*, 1198–1202.

Imai, Y., Kimura, T., Murakami, A., Yajima, N., Sakamaki, K., and Yonehara, S. The CED-4-homologous protein FLASH is involved in Fas-mediated activation of caspase-8 during apoptosis. Nature(London) *398*, 777–785.

Jonsson, Z.O., and Hubscher, U. (1997). Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. Bioessays *19*, 967–75.

Jurica, M.S., and Moore, M.J. Pre-mRNA splicing: awash in a sea of proteins. Mol.Cell *12*, 5–14.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO J. *23*, 616–626.

Keon, B.H., Schafer, S., Kuhn, C., Grund, C., and Franke, W.W. Symplekin, a novel type of tight junction plaque protein. J. Cell Biol. *134*, 1003–1018.

Khusial, P., Plaag, R., and Zieve, G.W. (2005). LSm proteins form heptameric rings that bind to RNA via repeating motifs. Trends Biochem. Sci. *30*, 522–528.

Kiriyama, M., Kobayashi, Y., Saito, M., Ishikawa, F., and Yonehara, S. Interaction of FLASH with arsenite resistance protein 2 is involved in cell cycle progression at S phase. Mol.Cell Biol. *29*, 4729–4741.

Kiss, A.M., Jady, B.E., Darzacq, X., Verheggen, C., Bertrand, E., and Kiss, T. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. Nucleic Acids Res. *30*, 4643–4649.

Kolev, N.G., and Steitz, J. a (2005). Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. Genes Dev. *19*, 2583–2592.

Kolev, N.G., and Steitz, J.A. Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. Genes Dev. *19*, 2583–2592.

Kolev, N.G., Yario, T. a, Benson, E., and Steitz, J. a (2008). Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. EMBO Rep. *9*, 1013–1018.

Krieg, P.A., and Melton, D.A. (1984). Formation of the 3' end of histone mRNA by post-transcriptional processing. Nature(London) *308*, 203–206.

Kuryshev, V.Y., Vorobyov, E., Zink, D., Schmitz, J., Rozhdestvensky, T.S., M??nstermann, E., Ernst, U., Wellenreuther, R., Moosmayer, P., Bechtel, S., et al. (2006). An anthropoid-specific segmental duplication on human chromosome 1q22. Genomics *88*, 143–151.

Last, T.J., Van Wijnen, A.J., Birnbaum, M.J., Stein, G.S., and Stein, J.L. (1999). Multiple interactions of the transcription factor YY1 with human histone H4 gene regulatory elements. J. Cell. Biochem. *72*, 507–516.

Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Rätsch, G., and Weigel, D. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. *105*, 8795–8800.

Lerner, M.R., and Steitz, J.A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. Proc. Natl. Acad. Sci. U. S. A. *76*, 5495–5499.

Lifton, R.P., Goldberg, M.L., Karp, R.W., and Hogness, D.S. (1978). The organization of the histone genes in drosophila melanogaster: functional and evolutionary implications. Cold Spring Harb. Symp. Quant. Biol. *42*, 1047–1051.

Liu, J.L., Wu, Z., Nizami, Z., Deryusheva, S., Rajendra, T.K., Beumer, K.J., Gao, H., Matera, A.G., Carroll, D., and Gall, J.G. Coilin is essential for Cajal body organization in Drosophila melanogaster. Mol Biol Cell *20*, 1661–1670.

Liu, J.L., Murphy, C., Buszczak, M., Clatterbuck, S., Goodman, R., and Gall, J.G. The Drosophila melanogaster Cajal body. J. Cell Biol. *172*, 875–884.

Liu, J.-L., Murphy, C., Buszczak, M., Clatterbuck, S., Goodman, R., and Gall, J.G. (2006). The Drosophila melanogaster Cajal body. J. Cell Biol. *172*, 875–884.

Liu, Y., Du, L., Osato, M., Eng, H.T., Qian, F., Jin, H., Zhen, F., Xu, J., Guo, L., Huang, H., et al. (2007). The zebrafish udu gene encodes a novel nuclear factor and is essential for primitive erythroid cell development. Blood *110*, 99–106.

Lobbes, D., Rallapalli, G., Schmidt, D.D., Martin, C., and Clarke, J. (2006). SERRATE: a new player on the plant microRNA scene. EMBO Rep. *7*, 1052–1058.

Lu, P., Hankel, I.L., Knisz, J., Marquardt, A., Chiang, M.-Y., Grosse, J., Constien, R., Meyer, T., Schroeder, A., Zeitlmann, L., et al. (2010). The Justy mutation identifies Gon4-like as a gene that is essential for B lymphopoiesis. J. Exp. Med. *207*, 1359–1367.

Lu, P., Hankel, I.L., Hostager, B.S., Swartzendruber, J.A., Friedman, A.D., Brenton, J.L., Rothman, P.B., and Colgan, J.D. (2011). The developmental regulator protein Gon4l associates with protein YY1, co-repressor Sin3a, and histone deacetylase 1 and mediates transcriptional repression. J. Biol. Chem. *286*, 18311–18319.

Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. Interaction profiling identifies the human nuclear exosome targeting complex. Mol.Cell *43*, 624–637.

Ma, T., Van Tine, B.A., Wei, Y., Garrett, M.D., Nelson, D., Adams, P.D., Wang, J., Qin, J., Chow, L.T., and Harper, J.W. (2000). Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. Genes Dev. *14*, 2298–2313.

Maga, G., and Hubscher, U. (2003). Proliferating cell nuclear antigen (PCNA): a dancer with many partners. J. Cell Sci. *116*, 3051–3060.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature(London) *444*, 953–956.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3???-end processing. Cell. Mol. Life Sci. *65*, 1099–1122.

Martincic, K., Campbell, R., Edwalds-Gilbert, G., Souan, L., Lotze, M.T., and Milcarek, C. (1998). Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G0 to S phase transition. Proc. Natl. Acad. Sci. U. S. A. *95*, 11095–11100.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. *9*, 843–854.

Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J.P., and Maltais, L. (2002). The human and mouse replication-dependent histone genes. Genomics *80*, 487–498.

Matera, A.G., Terns, R.M., and Terns, M.P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. *8*, 209–220.

McCall, C.M., Miliani de Marval, P.L., Chastain, P.D., Jackson, S.C., He, Y.J., Kotake, Y., Cook, J.G., and Xiong, Y. Human immunodeficiency virus type 1 Vpr-binding protein VprBP, a WD40 protein associated with the DDB1-CUL4 E3 ubiquitin ligase, is essential for DNA replication and embryonic development. Mol Cell Biol *28*, 5621–5633.

Miele, A., Braastad, C.D., Holmes, W.F., Mitra, P., Medina, R., Xie, R., Zaidi, S.K., Ye, X., Wei, Y., Harper, J.W., et al. HiNF-P directly links the cyclin E/CDK2/p220NPAT pathway to histone H4 gene regulation at the G1/S phase cell cycle transition. Mol Cell Biol *25*, 6140–6153.

Milovic-Holm, K., Krieghoff, E., Jensen, K., Will, H., and Hofmann, T.G. (2007). FLASH links the CD95 signaling pathway to the cell nucleus and nuclear bodies. EMBO J. *26*, 391–401.

Mowry, K.L., and Steitz, J.A. (1987). Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNA's. Science (80-.). *238*, 1682–1687.

Nilsen, T.W. (2003). The spliceosome: The most complex macromolecular machine in the cell? BioEssays *25*, 1147–1149.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell *79*, 639–648.

Ohtomo, T., Horii, T., Nomizu, M., Suga, T., and Yamada, J. (2007). Molecular cloning of a structural homolog of YY1AP, a coactivator of the multifunctional transcription factor YY1. Amino Acids *33*, 645–652.

Palko, L., Bass, H.W., Beyrouthy, M.J., and Hurt, M.M. (2004) The Yin Yang-1 (YY1) protein undergoes a DNA-replication-associated switch in localization from the cytoplasm to the nucleus at the onset of S phase. J. Cell Sci. *117*, 465–476.

Parsons, C.J., Stefanovic, B., Seki, E., Aoyama, T., Latour, A.M., Marzluff, W.F., Rippe, R.A., and Brenner, D.A. Mutation of the 5' untranslated region stem-loop structure inhibits {alpha}1(i) collagen expression in vivo. J. Biol. Chem. *286*, 8609–8619.

Pillai, R.S., Will, C.L., Lührmann, R., Schümperli, D., and Müller, B. (2001a). 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.

Pillai, R.S., Will, C.L., Lehrmann, R., Schumperli, D., and Muller, B. (2001b). 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.

Pillai, R.S., Grimmler, M., Meister, G., Will, C.L., Lührmann, R., Fischer, U., and Schümperli, D. (2003). Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev. *17*, 2321–2333.

Rajendra, T.K., Praveen, K., and Matera, A.G. (2010). Genetic analysis of nuclear bodies: from nondeterministic chaos to deterministic order. Cold Spring Harb.Symp.Quant.Biol *75*, 365–374.

Robberson, B.L., Cote, G.J., and Berget, S.M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. *10*, 84–94.

Rüegsegger, U., Beyer, K., and Keller, W. Purification and characterization of human cleavage factor Im, involved in the 3' end processing of messenger RNA precursors. J. Biol. Chem. *271*, 6107–6113.

Rüegsegger, U., Beyer, K., and Keller, W. (1996). Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. J. Biol. Chem. *271*, 6107–6113.

Rüegsegger, U., Blank, D., and Keller, W. (1998). Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Mol. Cell *1*, 243–253.

Ruepp, M.D., Schweingruber, C., Kleinschmidt, N., and Schumperli, D. Interactions of CstF-64, CstF-77, and symplekin: implications on localisation and function. Mol Biol Cell *22*, 91–104.

Ruepp, M.D., Vivarelli, S., Pillai, R.S., Kleinschmidt, N., Azzouz, T.N., Barabino, S.M., and Schumperli, D. 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. -.

Sabin, L.R., Zhou, R., Gruber, J.J., Lukinova, N., Bambina, S., Berman, A., Lau, C.K., Thompson, C.B., and Cherry, S. Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. Cell *138*, 340–351.

Sadowski, M., Dichtl, B., Hubner, W., and Keller, W. Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. EMBO J. *22*, 2167–2177.

Salzler, H.R., Tatomer, D.C., Malek, P.Y., McDaniel, S.L., Orlando, A.N., Marzluff, W.F., and Duronio, R.J. (2013). A sequence in the Drosophila H3-H4 promoter triggers histone locus body assembly and biosynthesis of replication-coupled histone mRNAs . Dev. Cell *24*, 623–634.

Sanchez, R., and Marzluff, W.F. (2002). The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. Mol. Cell. Biol. *22*, 7093–7104.

Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. Nat. Rev. Mol. Cell Biol. *6*, 139–149.

Scharl, E.C., and Steitz, J.A. (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.

Scharl, E.C., and Steitz, J.A. (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.

Schaufele, F., Gilmartin, G.M., Bannwarth, W., and Birnstiel, M.L. (1986). Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA. Nature(London) *323*, 777–781.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates III, J.R., Frank, J., and Manley, J.L. Molecular architecture of the human pre-mRNA 3' processing complex. Mol Cell *33*, 365–376.

Smith, H.O., Tabiti, K., Schaffner, G., Soldati, D., Albrecht, U., and Birnstiel, M.L. (1991). Two-step affinity purification of U7 small nuclear ribonucleoprotein particles using complementary biotinylated 2'-O-methyl oligoribonucleotides. Proc. Natl. Acad. Sci. U. S. A. *88*, 9784–9788.

Stefanovic, B., Wittop Koning, T.H., and Schümperli, D. (1995a). A synthetic histone pre-mRNA-U7 small nuclear RNA chimera undergoing cis cleavage in the cytoplasm of Xenopus oocytes. Nucleic Acids Res. *23*, 3152–3160.

Stefanovic, B., Hackl, W., Lührmann, R., and Schümperli, D. (1995b). Assembly, nuclear import and function of U7 snRNPs studied by microinjection of synthetic U7 RNA into Xenopus oocytes. Nucleic Acids Res. *23*, 3141–3151.

Strahl, B.D., and Allis, C.D. The language of covalent histone modifications. Nature(London) *403*, 41–45.

Streit, A., Koning, T.W., Soldati, D., Melin, L., and Schümperli, D. (1993). Variable effects of the conserved RNA hairpin element upon 3' end processing of histone pre-mRNA in vitro. Nucleic Acids Res. *21*, 1569–1575.

Strub, K., and Birnstiel, M.L. (1986). Genetic complementation in the Xenopus oocyte: co-expression of sea urchin histone and U7 RNAs restores 3' processing of H3 pre-mRNA in the oocyte. EMBO J. *5*, 1675–1682.

Strub, K., Galli, G., Busslinger, M., and Birnstiel, M.L. (1984). The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing. EMBO J. *3*, 2801–2807.

Sullivan, E., Santiago, C., Parker, E.D., Dominski, Z., Yang, X., Lanzotti, D.J., Ingledue, T.C., Marzluff, W.F., and Duronio, R.J. (2001). Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. Genes Dev. *15*, 173–187.

Sullivan, K.D., Mullen, T.E., Marzluff, W.F., and Wagner, E.J. (2009a). Knockdown of SLBP results in nuclear retention of histone mRNA. RNA *15*, 459–472.

Sullivan, K.D., Steiniger, M., and Marzluff, W.F. (2009b). A core complex of CPSF73, CPSF100 and symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. Mol. Cell *34*, 322–332.

Takagaki, Y., and Manley, J.L. Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol.Cell Biol. *20*, 1515–1525.

Takagaki, Y., and Manley, J.L. (1998). Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. Mol. Cell *2*, 761–771.

Tan, D., Marzluff, W.F., Dominski, Z., and Tong, L. Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hExo ternary complex. Science (80-.). *339*, 318–321.

Venkataraman, K., Brown, K.M., and Gilmartin, G.M. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. *19*, 1315–1327.

Veraldi, K.L., Edwalds-Gilbert, G., MacDonald, C.C., Wallace, A.M., and Milcarek, C. (2000). Isolation and characterization of polyadenylation complexes assembled in vitro. RNA *6*, 768–777.

De Vries, H., Rüegsegger, U., Hübner, W., Friedlein, A., Langen, H., and Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. *19*, 5895–5904.

Wagner, E.J., Burch, B.D., Godfrey, A.C., Salzer, H.R., Duronio, R.J., and Marzluff, W.F. (2007). A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing. Mol. Cell *28*, 692–699.

Walther, T.N., Wittop, K.T.H., Schümperli, D., and Muller, B. A 5'-3' exonuclease activity involved in forming the 3' products of histone pre-mRNA processing in vitro. RNA *4*, 1034–1046.

Wang, C.-Y., Liang, Y.-J., Lin, Y.-S., Shih, H.-M., Jou, Y.-S., and Yu, W.C.Y. (2004). YY1AP, a novel co-activator of YY1. J. Biol. Chem. *279*, 17750–17755.

Wang, Z.-F., Whitfield, M.L., Ingledue, T.I., Dominski, Z., and Marzluff, W.F. (1996). The protein which binds the 3' end of histone mRNA: a novel RNA- binding protein required for histone pre-mRNA processing. Genes Dev. *10*, 3028–3040.

Wei, N., Serino, G., and Deng, X.W. (2008). The COP9 signalosome: more than a protease. Trends Biochem. Sci. *33*, 592–600.

Wei, Y., Jin, J., and Harper, J.W. (2003). The cyclin E/Cdk2 substrate and Cajal body component p220(NPAT) activates histone transcription through a novel LisH-like domain. Mol. Cell. Biol. *23*, 3669–3680.

White, A.E., Burch, B.D., Yang, X.C., Gasdaska, P.Y., Dominski, Z., Marzluff, W.F., and Duronio, R.J. (2011). Drosophila histone locus bodies form by hierarchical recruitment of components. J. Cell Biol. *193*, 677–694.

Whitfield, M.L., Zheng, L.-X., Baldwin, A., Ohta, T., Hurt, M.M., and Marzluff, W.F. (2000). Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms. Mol. Cell. Biol. *20*, 4188–4198.

Wu, C.-H.H., and Gall, J.G. (1993). U7 small nuclear RNA in C snurposomes of the Xenopus germinal vesicle. Proc. Natl. Acad. Sci. U. S. A. *90*, 6257–6259.

Yang, X., Xu, B., Sabath, I., Kunduru, L., Burch, B.D., Marzluff, W.F., and 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.

Yang, X.C., Sullivan, K.D., Marzluff, W.F., and Dominski, Z. Studies on the 5’ exonuclease and endonuclease activity of CPSF-73 in histone pre-mRNA processing. Mol.Cell Biol. *29*, 21–32.

Yang, X.C., Xu, B., Sabath, I., Kunduru, L., Burch, B.D., Marzluff, W.F., and Dominski, Z. 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.

Yang, X.C., Burch, B.D., Yan, Y., Marzluff, W.F., and Dominski, Z. (2009a). FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3’ end processing of histone pre-mRNAs. Mol. Cell *36*, 267–278.

Yang, X.C., Torres, M.P., Marzluff, W.F., and Dominski, Z. (2009b). Three proteins of the U7-specific Sm ring function as the molecular ruler to determine the site of 3’ end processing in mammalian histone pre-mRNA. Mol. Cell. Biol. *29*, 4045–4056.

Yang, X.-C., Sabath, I., Dębski, J., Kaus-Drobek, M., Dadlez, M., Marzluff, W.F., and Dominski, Z. (2013). A complex containing the CPSF73 endonuclease and other polyadenylation factors associates with U7 snRNP and is recruited to histone pre-mRNA for 3’-end processing. Mol. Cell. Biol. *33*, 28–37.

Ye, X., Wei, Y., Nalepa, G., and Harper, J.W. The cyclin E/Cdk2 substrate p220(NPAT) is required for S-phase entry, histone gene expression, and Cajal body maintenance in human somatic cells. Mol Cell Biol *23*, 8586–8600.

Zhao, J., Kennedy, B.K., Lawrence, B.D., Barbie, D.A., Matera, A.G., Fletcher, J.A., and Harlow, E. (2000). NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. Genes Dev. *14*, 2283–2297.

Zhao, J.Y., Dynlacht, B., Imai, T., Hori, T., and Harlow, E. Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry. Genes Dev. *12*, 456–461.

Zheng, L.-X., Dominski, Z., Yang, X., Elms, P., Raska, C.S., Borchers, C.H., and Marzluff, W.F. (2003). Phosphorylation of SLBP on two threonines triggers degradation of SLBP, the sole cell-cycle regulated factor required for regulation of histone mRNA processing, at the end of S-phase. Mol.Cell Biol. *23*, 1590–1601.