Influence of Protein Binding and Competition on 7SK Ribonucleoprotein Complex Maintenance and Transition

by

Daniel Christopher Totten

B.A. in Biology, Hendrix College, 2010

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Molecular, Cellular, and Developmental Biology

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Daniel Totten

It was defended on

October 11, 2017

and approved by

Karen Arndt, PhD, Professor

William Saunders, PhD, Associate Professor

Paula Grabowski, PhD, Professor

John Woolford, PhD, Professor

Dissertation Advisor: Andrea Berman, PhD, Assistant Professor

INFLUENCE OF PROTEIN BINDING AND COMPETITION ON 7SK RIBONUCLEOPROTEIN COMPLEX MAINTENANCE AND TRANSITION

Daniel Totten, PhD

University of Pittsburgh, 2017

As the first step of gene expression, transcription is tightly regulated to ensure the proper growth, development, and homeostasis of an organism. In metazoans, an extra layer of control, known as promoter proximal pausing, is exerted approximately fifty nucleotides into elongation through the association of negative elongation factors with RNA polymerase II. The cyclin-dependent kinase positive transcription elongation factor b (P-TEFb) phosphorylates both the negative elongation factors and the C-terminal domain of the polymerase to release paused RNA polymerase II into productive elongation. Promoter proximal pausing is a highly pervasive control mechanism, as inhibition of P-TEFb abolishes global transcription. P-TEFb activity is predominantly controlled through sequestration and release from an inhibitory ribonucleoprotein complex (RNP) containing the non-coding small nuclear 7SK RNA (7SK-P-TEFb RNP). Release of P-TEFb from this RNP allows 7SK RNA to bind various heterogeneous nuclear ribonucleoproteins (hnRNPs) to form the 7SK-hnRNP RNPs. The transition between the 7SK-P-TEFb RNP and the 7SK-hnRNP RNPs controls active P-TEFb levels, and thus metazoan transcription. The functional consequence of assembly and transition between the 7SK RNPs is unknown. In this thesis, I investigated the effects of protein binding and competition in establishing, maintaining, and transitioning the respective 7SK RNPs in two studies. In the case of hnRNP A1 and serine-arginine splicing factor 2 (SRSF2) binding to Stem III of 7SK RNA, I found that the proteins differentially restructure the RNA element upon binding, helping to maintain exclusive RNP complex formation at high concentrations. However, the formation of

an intermediate complex with low concentrations of both factors helps mediate SRSF2 dissociation from the complex during RNP transition. In my second study, I found that direct phosphorylation of hnRNP K by P-TEFb modulates occupancy of competing proteins on Stem I of 7SK RNA. Furthermore, this post-translational modification may play a downstream role in mediating proper termination at select genes. Together, these studies detail the first biochemical examination of hnRNP-7SK RNA interactions and suggest that 7SK RNP maintenance and transition is controlled by transcription-dependent changes in the local concentrations of RNA-binding factors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This thesis is the culmination of years of hard work from a team of individuals, not just myself. I would like to thank the undergraduates that have helped me gather data – including all data not presented herein. Thank you Walter Wang, Megan Link, Ashley Neiswender, Rosalyn Marar, and Nick Mynarski for all the help, support, and tireless dedication to science and the lab. I can only hope that my mentorship can repay a fraction of what you guys gave.

I would also like to thank Dr. Andrea Berman for taking me into her lab and giving me free reign to pursue this incredibly difficult project. I'm sure most PIs would have given up a long time ago, and I thank you, from the bottom of my heart, for allowing me to persevere and find myself. As you've said countless times, "The PhD is just icing on the cake." I can't agree more. Thank you.

Finally, I would like to thank the Department of Biological Sciences for providing the resources, motivation, and incentives to accomplish my doctoral work. Thank you Cathy Barr for managing all administrative duties. Thank you to my committee for working with me to complete this project. I want to give special thanks to the Hatfull, Brodsky, Arndt, Saunders, Hendrix, and VanDemark labs for equipment, reagents, and advice.

I would like to dedicate this thesis to my boof. I love you.

1.0 INTRODUCTION

This body of work is a record of the data gleaned from biochemical, molecular, and cellular techniques used to understand the consequences of protein binding to, and competition for, a non-coding RNA, and the subsequent effects on transcriptional regulation in metazoans. To understand the importance of the data, this chapter focuses on building the foundation of transcription regulation in eukaryotes. Specific attention is paid to elongation control through the phenomenon of promoter proximal pausing – what it is, what causes it, and what regulates it. I then detail the role of the abundant non-coding RNA, 7SK RNA, in controlling this process. Finally, I end the chapter with a brief outline of the two projects contained herein and the hypotheses that guided their progress.

1.1 TRANSCRIPTION OVERVIEW

Nearly every aspect of a cell is guided through the regulated expression of the information contained within DNA. Transcription is the first essential step of this process, whereby the stable hereditary material of DNA is transcribed into RNA. Contrary to the popular representation in the central dogma, we now appreciate RNA as being more than simply an intermediate molecule in protein synthesis. RNA has a plethora of roles throughout the cell, including, but not limited to, translation (e.g., messenger RNA, transfer RNA, the catalytic mechanism behind protein

synthesis [ribosomal RNA], and a major component of the signal recognition particle [7SL RNA]), catalytic roles in RNA processing (e.g., the spliceosome and tRNA maturation), chromatin architecture (e.g., XIST-mediated coating and silencing of an X chromosome), transcription (e.g., guide RNAs and 7SK RNA regulating promoter proximal pausing), and RNA silencing (e.g., miRNAs) [1].

Eukaryotes have three RNA polymerases (RNA polymerase I, II, and III) [2], and each polymerase is responsible for transcribing specific subsets of genes. For example, RNA polymerase I transcribes 28S, 18S, and 5.8S ribosomal RNAs [3]; RNA polymerase II produces messenger RNA, small nuclear RNA, and miRNA [4]; and RNA polymerase III generates 5S ribosomal RNA, transfer RNA, and non-coding regulatory RNAs [5]. While partially conserved, each RNA polymerase associates with specific transcription factors and has unique levels of regulation [2]. This introduction will focus on RNA polymerase II transcription and regulation.

1.1.1 RNA polymerase II and the C-terminal domain (CTD)

RNA polymerase II is an ~520 kDa protein complex composed of 12 subunits, RPB1-12 [6]. RNA polymerase II is the major transcribing unit for messenger RNAs, small nuclear RNAs, small nucleolar RNAs, and miRNAs [7]. Because RNA polymerase II directly controls expression of protein-coding genes, cells have evolved a series of regulatory events to govern its activity [8]. Efficient transcription relies on proper formation of the holo-enzyme composed of the core polymerase and the general transcription factors at promoter elements near the transcription start site (Section 1.1.2) [9].

Unique to RNA polymerase II, the largest subunit, RPB1, contains a C-terminal domain (CTD) composed of 52 repeats in humans with the consensus sequence of YSPTSPS [10]. The

CTD is essential, as full deletion of the domain is lethal [11, 12]. Each amino acid within the CTD has the potential to be post-translationally modified by chemical modifications like phosphorylation or by isomerization [10]. Meta-genomic profiles of these modifications have led to the observation that specific marks correlate with defined stages of transcription. The pattern of modifications, along with the timing of the events, has been termed the "CTD code" [13]. The CTD and the status of its post-translational modifications (PTMs) are critically important in recruiting and scaffolding the protein and RNA components necessary for proper, and efficient, transcription and post-transcriptional processing of the nascent transcript – including the messenger RNA capping complex, chromatin remodelers, histone modifying enzymes, elongation factors, splicing machinery, and termination factors [14]. Thus, regulation of RNA polymerase II requires precise control not only over recruitment of the holo-enzyme to appropriate promoters, but also of the factors that mediate the PTMs of the CTD.

1.1.2 The three stages of transcription

Transcription can be roughly divided into three stages: initiation, elongation, and termination. Each stage has unique regulations and control mechanisms that coordinate a vast network of signals to properly mediate gene expression. Because eukaryotic DNA is a protein-nucleic acid complex, termed chromatin, nucleosome architecture is a prevalent and potent barrier at every stage of transcription [15]. Indeed, initiation is no exception, and the stage can only begin upon clearing nucleosomes bound to promoter elements, allowing access to transcription factors [16].

RNA polymerase II has little intrinsic preference for promoter DNA and cannot efficiently initiate without the aid of transcription factors (TFs) [17]. Therefore, initiation first requires formation of the pre-initiation complex (PIC) – whereby RNA polymerase II is recruited

to the promoter and forms the complete holo-enzyme through the assemblage of general transcription factors (TFII B, D, E, F, and H) within the nucleosome-free region upstream of the transcription start site (Figure 1.1) [9]. The approximate order of binding events is illustrated in Figure 1.1. Of note, the promoter is not recognized by the polymerase itself, but rather from the initial binding events of TFIID [18, 19], TFIIB [20, 21], and, frequently, TFIIA [22]. Formation of the PIC is considered to be the rate-limiting step of transcription in yeast [23, 24]; however, growing evidence suggests that the rate-limiting step is gene-specific and may rely on either formation of the PIC or promoter proximal pausing (Section 1.2) in metazoans [25].

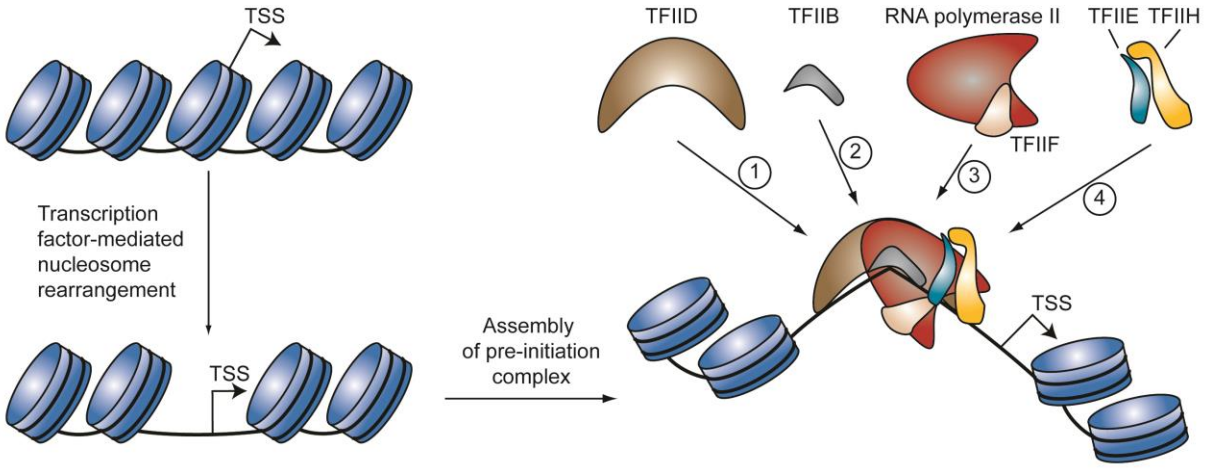

Figure 1.1. Formation of the pre-initiation complex. Adapted from [9]. Prior to initiation, promoter elements must be accessed via nucleosome remodeling events (left). Once exposed, the general transcription factors will recognize and bind the DNA, recruiting RNA polymerase II and the other general transcription factors necessary to initiate (right).

After formation of the PIC, additional signals from transcription factors bound to *cis*-acting elements (e.g., enhancers) are relayed to the PIC through the large protein complex known as Mediator [26]. Initiation begins upon phosphorylation of S5 within the CTD by the cyclin-

dependent kinase 7 subunit of TFIIH [27], while another TFIIH subunit hydrolyses ATP to induce DNA melting and formation of the transcription bubble [28]. Addition of the first two ribonucleotide triphosphates (NTPs) complementary to the template [29] and catalysis of the phosphodiester bond between them by the RBP2 subunit of RNA polymerase II follows [30]. Although highly processive in elongation, RNA polymerase II can undergo cycles of abortive transcription during initiation, during which it releases truncated RNA transcripts [31]. However, if the polymerase can successfully catalyze the addition of 14 nucleotides, it will successfully clear the promoter and enter the elongation phase of transcription [31].

Successful promoter clearance signifies the beginning of the elongation stage, during which RNA polymerase II transverses the entirety of the transcription unit until termination signals are reached. Metazoan elongation can be broadly divided into two sub-sections: promoter proximal pausing (discussed in detail in Section 1.2) and productive elongation (Figure 1.2) [32]. The transition between the two sub-stages is a key regulatory step in gene expression, and potentially adds a crucial quality-control checkpoint for the extensive processing events that occur throughout productive elongation.

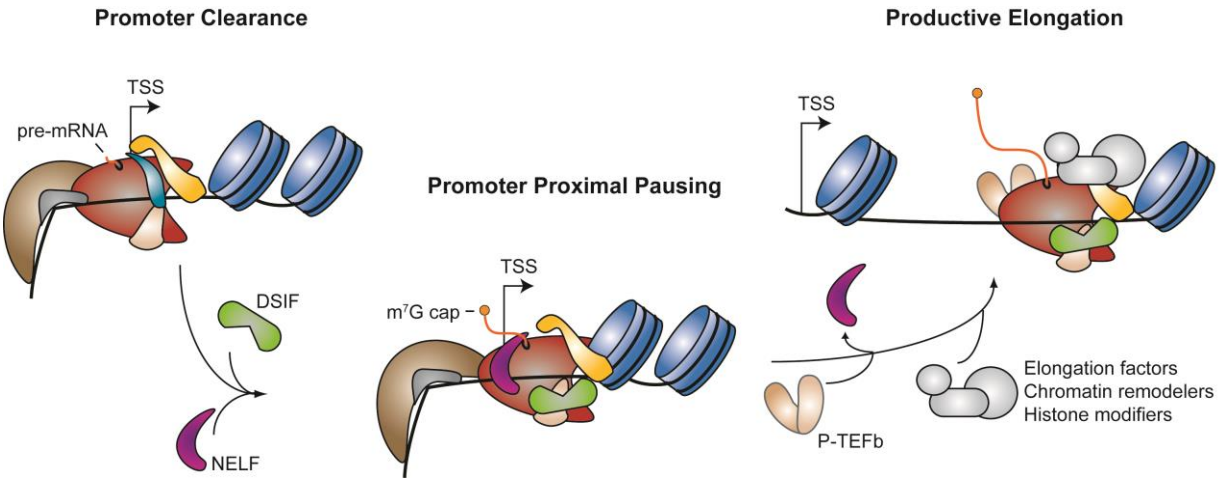

Figure 1.2. Transcription elongation has two phases. Adapted from [25]. After successful promoter clearance (left), negative elongation factors bind and hold the elongating polymerase near the promoter (promoter proximal pausing, middle). Upon stimulation by P-TEFb, RNA polymerase II is released and acquires a multitude of elongation factors as it enters productive elongation (right).

As the polymerase clears the promoter, initiation factors are exchanged for elongation factors (Figure 1.2) [32]. These factors aid in co-transcriptional RNA processing events such as addition of the 5' cap for messenger RNAs [33] and splicing [34], as well as successful navigation of chromatin [35]. Factor exchange and recruitment involves a dynamic interplay among elements such as the nascent RNA [36], histone post-translational modifications [37], and the status of the CTD code (S5 phosphorylation decreases as S2 phosphorylation increases) [38]. Successful elongation depends on the correct association of these factors throughout the progression of the gene. Indeed, RNA polymerase II has a low transcription rate (~0.5 kb/min) for the first several kilobases of elongation, presumably to ensure proper addition of elongation factors prior to full elongation (2-5 kb/min) [39]. However, elongation should not be considered as a "runaway train," as the rate of elongation varies dramatically not only throughout the body of a gene, but also between genes [40, 41]. Increasing evidence supports the idea that controlling

the speed of RNA polymerase II through pausing events is critical for proper transcription and genome stability [25].

Once the polymerase nears the end of a transcriptional unit, two critical processes must occur during termination, the final stage of transcription: 1) release of RNA from the elongating polymerase, and subsequent final RNA processing and 2) disengagement of the polymerase from DNA (Figure 1.3) [42]. Unlike initiation and elongation, termination is less-well studied, particularly owing to the difficulty in untangling the contribution of termination factors to the two processes. In metazoans, there appears to be three distinct mechanisms for terminating specific subsets of genes: one for messenger RNAs, one for small nuclear RNAs, and one for replication-dependent histone transcripts [42]. To add to the complexity, two different models exist to explain how RNA polymerase II disengages upon termination: the allosteric model and the torpedo model (Figure 1.3) [42].

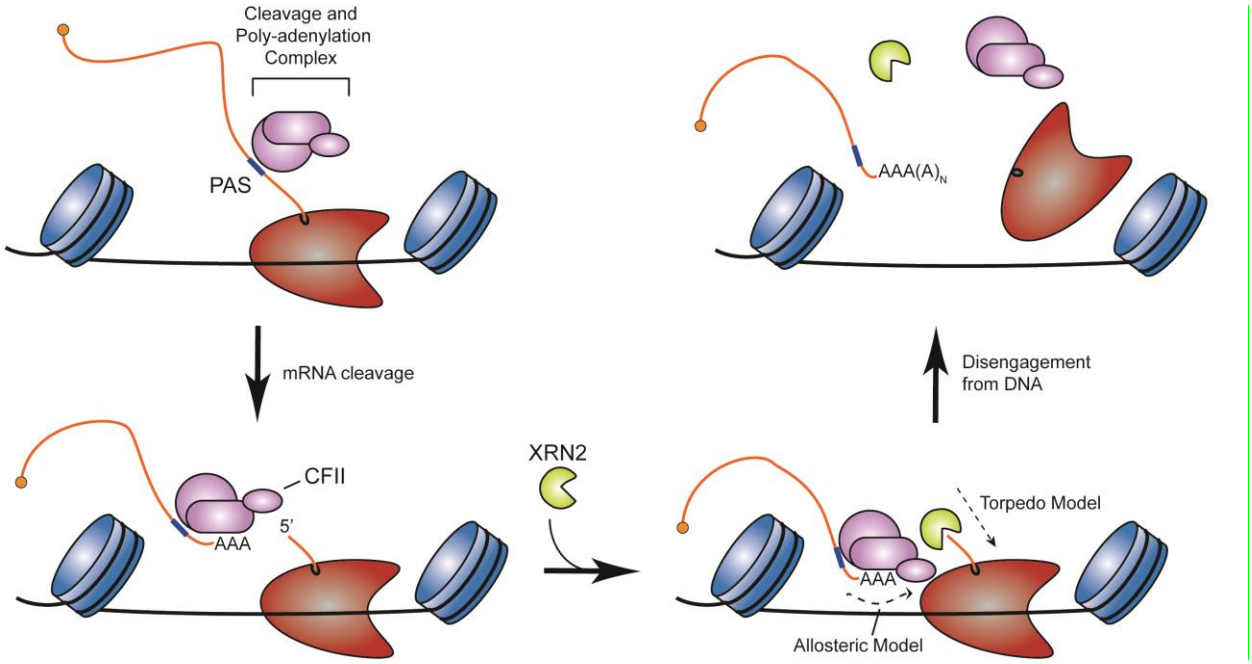

Figure 1.3. Transcription termination requires mRNA and polymerase disengagement. Adapted from [42]. After transcribing the poly-adenylation signal (PAS), the cleavage and poly-adenylation complex is recruited to a paused polymerase to cleave the RNA downstream of the PAS. In the torpedo model, exonuclease XRN2 is recruited to the polymerase, whereupon digestion of the nascent transcript releases the polymerase from DNA [43]. In the allosteric model, interactions mediated in part through cleavage factor II (CFII) and RNA polymerase II lead to disengagement [44, 45]. Afterwards, the messenger RNA transcript undergoes final quality control processing and export to the cytoplasm, while RNA polymerase II initiates a new transcript.

For genes encoding messenger RNAs, termination begins once the polymerase transcribes the poly-adenylation signal (PAS) [46, 47]. This signal recruits members of the cleavage and poly-adenylation complex, which work together to cleave the nascent transcript and add the poly-adenylate tail that aids in messenger RNA translation and stability [48, 49]. The polymerase is then either disengaged through allosteric interactions via a cleavage factor II component (the allosteric model) [44, 45], via displacement by the 5'-3' exonuclease activity of XRN2 on the remaining nascent transcript (the torpedo model) [43], or via a conservative model

that integrates both (Figure 1.3) [50]. Regardless of the mechanism, efficient termination (and maintenance of transcriptional unit boundaries [51, 52]) is aided in part through pausing of the polymerase at the end of the transcriptional unit [53-55], although the exact contribution(s) of this event are still debated.

Together, these three stages encompass the entire process of transcription. As the first step in gene expression, transcription is highly regulated and critically important; however, it is by no means the end of the story. The cell employs a plethora of post-transcriptional, translational, and post-translational mechanisms to adeptly process and convey the information stored within DNA necessary for homeostasis and survival. In the following sections, I focus on the metazoan-specific phenomenon of promoter proximal pausing and provide detailed analysis of the major components that influence its regulation.

1.2 PROMOTER PROXIMAL PAUSING

In *Saccharomyces cerevisiae*, a major model organism for studying eukaryotic transcription, formation of the PIC is the rate-limiting step at many genes [23, 24]. However, when investigators examined the transcription of inducible genes, such as *MYC* [56], in human cell lines, there appeared to be a transcriptional block past initiation but prior to elongation. In agreement with this observation, a series of studies in *Drosophila melanogaster* discovered that RNA polymerase II accumulated near the promoter of *HSP70* [57, 58]. These polymerases were transcriptionally engaged, since stimulation upon temperature shift elicited rapid production of transcripts [58]. These studies were corroborated with others that utilized ATP analogs such as 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB). DRB halted transcription through an

unknown mechanism that did not directly inhibit RNA polymerase II [59, 60]. Instead of observing a complete loss of RNA production, treatment with DRB gave rise to short, capped RNAs [61]. Together, these pioneer studies suggested that metazoan RNA polymerase II efficiently initiates but does not continue into elongation. However, these stalled promoter proximal polymerases enter into productive elongation upon stimulation, and are therefore considered to be paused rather than arrested. These studies laid the groundwork for the metazoan-specific phenomenon of promoter proximal pausing.

1.2.1 Mechanisms governing promoter proximal pausing and release

Unlike transcriptional arrest, paused polymerases do not disengage the nascent RNA from the active site and can resume elongation upon stimulation. Transcriptional pausing is an intrinsic property of polymerases [62], including RNA polymerase II. Weak pause sites appear at least once in every 100 base-pairs throughout transcribed units [32]. Strategic positioning of strong pause sites throughout genes [63, 64] may play a role in mediating co-transcriptional folding and processing, e.g., at exon boundaries and poly-adenylation sites [25]. In addition, transcriptional pausing may aid in preserving genomic integrity, as a number of studies have shown that increasing the elongation rate results in abnormal chromatin maintenance. This leads to nucleosome repositioning across important regulatory elements that eventually results in an overall decrease in transcription [65-67].

Pausing has such a large effect on elongation rate that the measured rate *in vitro* without additional factors (25-50 bases/minute) [68] is over 100-fold lower than the rate measured *in vivo* (~4 kilobases/minute) [39, 41]. This implies that additional elements are required for efficient elongation. Indeed, there are not only a host of positive elongation factors that decrease

polymerase pausing and enhance the elongation rate, but also negative elongation factors that work to elicit pausing events. Clearly, not one individual factor works alone, and therefore elongation is the culmination of a myriad of factors working together to achieve the observed rate of ~4 kilobases/minute [39, 41].

While splicing and termination events have strong, defined encoded pause sites [25], no consensus sequence has been described for promoter proximal pausing. Instead, it is thought that promoter proximal pausing is a unique regulatory mechanism brought about through recruitment of negative elongation factors via specific transcription factors acting in *trans*. After promoter clearance, RNA polymerase II interacts with the negative elongation factors DRB-sensitivity inducing factor (DSIF), a homolog of yeast Spt4 and Spt5 [69], and the multi-subunit negative elongation factor (NELF) [70]. Binding to both the polymerase [71, 72] and the nascent RNA [70, 73, 74] approximately 50 nucleotides past the transcription start site, these factors work together to prompt promoter proximal pausing (Figure 1.2) [75]. Indeed pausing *in vitro* only occurs upon addition of both factors [76].

Phylogenetic studies indicate that promoter proximal pausing correlates with the evolution of NELF, as the phenomenon is not observed in species lacking the NELF complex [32, 77]. However, knock-down of NELF does not uniformly affect promoter proximal pausing across the genome [78, 79], suggesting that additional factors contribute to establishing promoter proximal pausing. One factor, Gdown1, is an auxiliary protein [80] that works with Gdown1 negative accessory factor (GNAF) and competes with TFIIF to induce promoter proximal pausing (Figure 1.4) [81]. Interestingly, polymerases containing Gdown1 are uniquely sensitive to activation through interaction with Mediator [80]. Gdown1-mediated elongation repression is also aided through TFIIS, a classical positive elongation factor, although the exact mechanism by

which TFIIS restricts the polymerase instead of helping release it from arrest is unknown (Figure 1.4) [81]. Finally, nucleosome positioning and composition also help establish promoter proximal pausing. The location of paused polymerase appears to depend on the exact position of the +1 nucleosome [82, 83], and addition of histone variants can decrease the prevalence of pausing events [83].

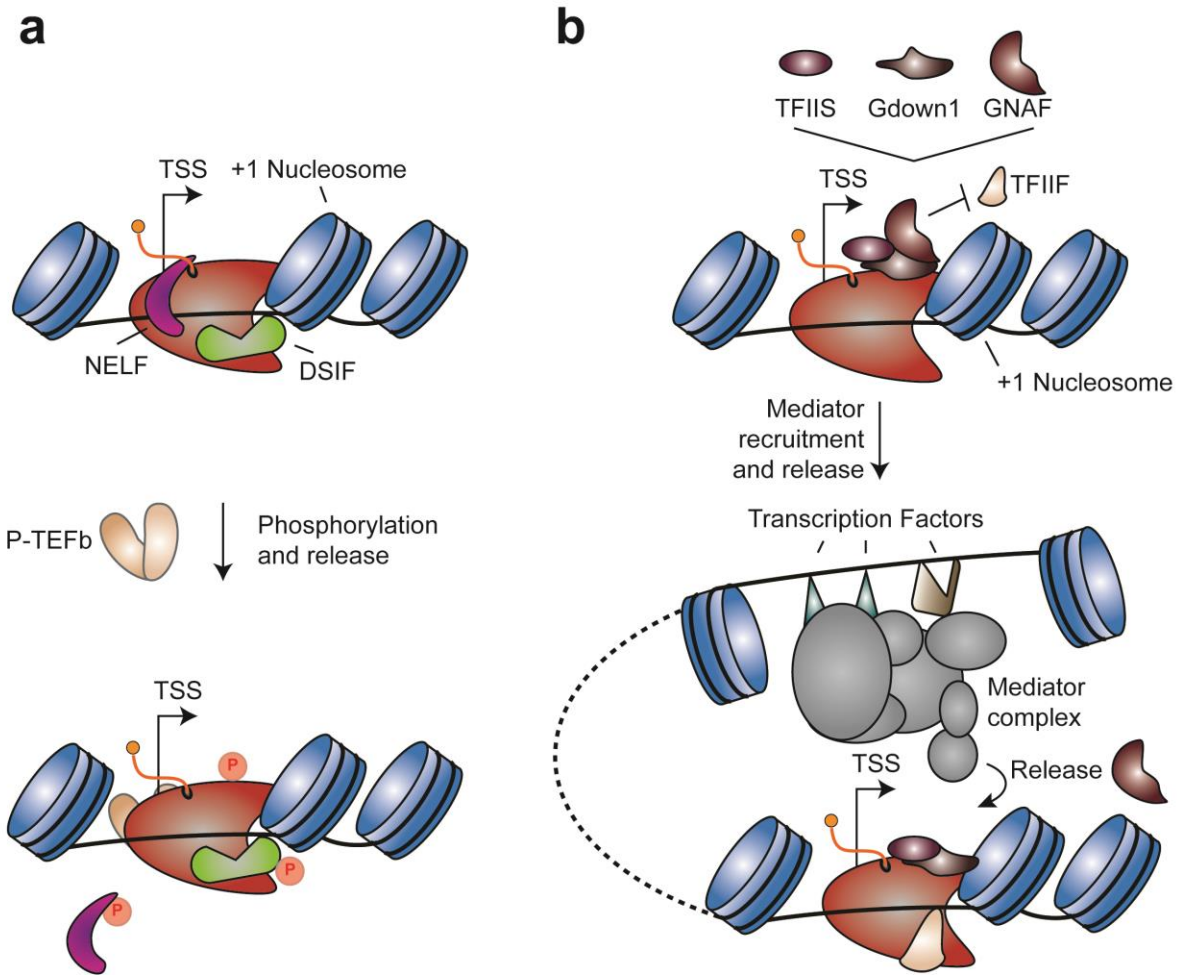

Figure 1.4. Promoter proximal pausing can be established through two non-mutually exclusive mechanisms. Adapted from [32]. (a) ~50 nucleotides into elongation, RNA polymerase II associates with the negative elongation factors NELF and DSIF and pauses. Upon recruitment of P-TEFb, the kinase phosphorylates NELF, DSIF, and the CTD of RNA polymerase II, promoting productive elongation. (b) An auxiliary negative elongation factor, Gdown1, works together with TFIIS and Gdown negative accessory factor (GNAF) to promote pausing near the promoter. Gdown1-mediated promoter proximal pausing can be released through recruitment and activation via the Mediator complex.

Once paused, how does the cell release active polymerases to meet transcriptional demand? *In vitro* transcription assays using nuclear extract identified positive elongation factor b (P-TEFb) as a critical component that provides the stimulation to relieve promoter proximal pausing [59]. Subsequent cloning identified P-TEFb as PITARLE, with further characterization revealing it as a cyclin-dependent kinase (CDK) [84]. P-TEFb is a heterodimer composed of the active kinase subunit CDK9 bound to one of four regulatory cyclins (cyclin T1 [85, 86], T2a, T2b [87], and K [88]), although the majority of CDK9 associates with the T cyclins (Figure 1.5) [89]. Upon recruitment to promoter proximally paused RNA polymerase II (Section 1.3.1), activated P-TEFb phosphorylates the Spt5 subunit of DSIF [90] and the NELF-E subunit of NELF [91]. These events turn DSIF into a positive elongation factor [92] that will stably associate with the elongating polymerase for the entirety of the gene [66], while NELF dissociates from the elongation complex [91]. Although only needed for the initial release of paused polymerase at the 5' end of genes [93, 94], P-TEFb is incorporated into the elongating polymerase as a part of the super elongation complex [95], a collection of eleven-nineteen Lys-rich leukemia and mixed lineage leukemia proteins that associate with RNA polymerase II and aid in overcoming pausing events [96].

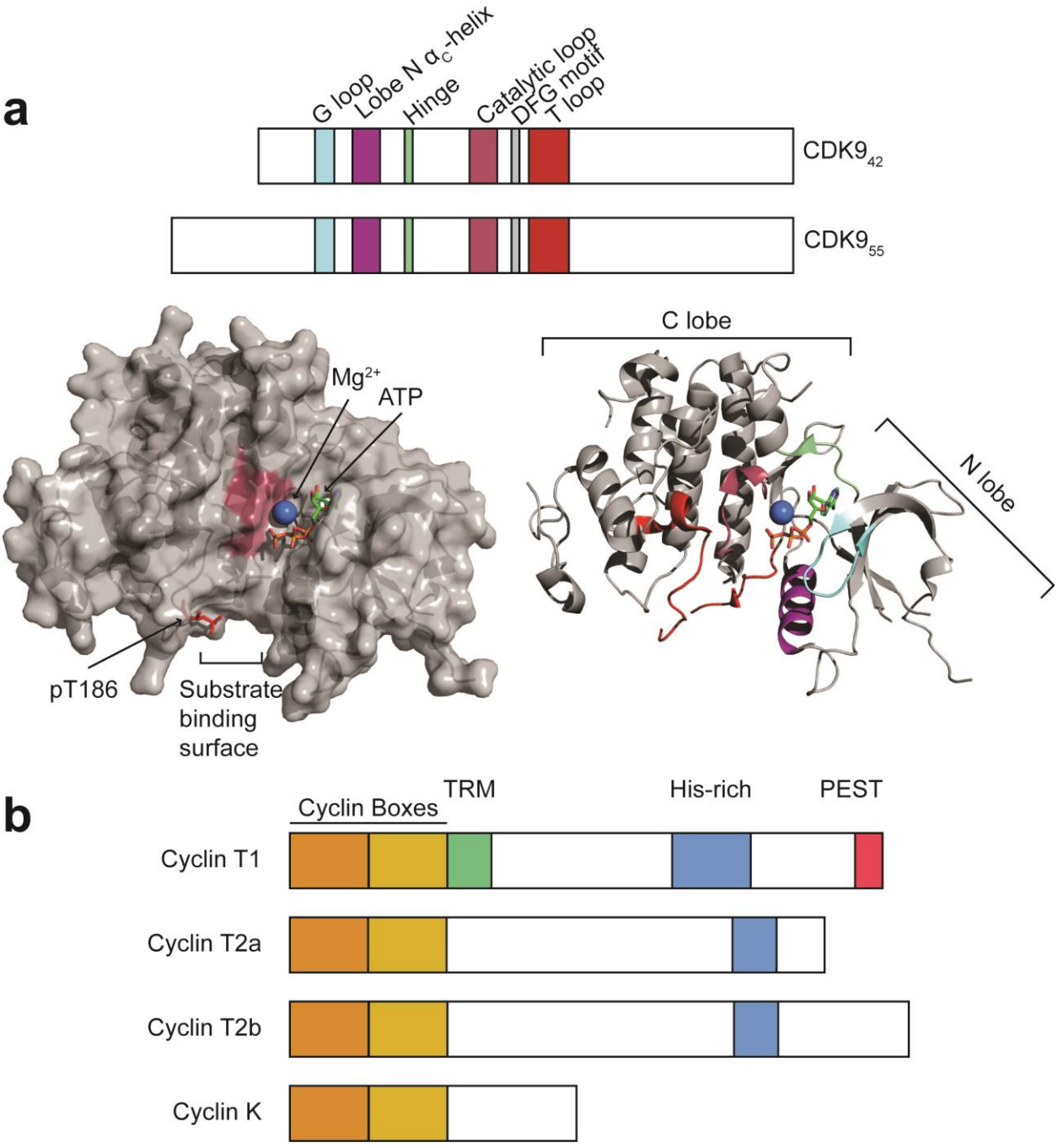

Figure 1.5. P-TEFb sub-domain and domain architecture and subunits. (a) The kinase subunit CDK9 is expressed as two isoforms with molecular weights of 42 kDa and 55 kDa [97]. Box diagram of the relative position of the subdomains (top). G loop: binds ATP, lobe N α_C-helix: binds cyclins, hinge: separates N and C lobes, catalytic loop: residues required for catalysis, DFG motif :coordinates Mg^{2+}, T loop: aligns substrates into binding pocket [84]. Surface representation (left) and ribbon model (right) are shown (PDB ID: 3BLQ) [98]. (b) CDK9 can

bind four regulatory cyclins [84]. Box diagram of domains are shown. Cyclin boxes: bind CDK9, TRM: Tat recognition motif, his-rich: histidine-rich motif, PEST: rich in proline, glutamate, serine, and threonine sequence.

Along with relieving promoter proximal pausing via phosphorylation of DSIF and NELF, P-TEFb is also implicated in phosphorylating the CTD at S2 [99, 100]. However, growing evidence suggest that this phosphorylation event may be an auxiliary function of P-TEFb, as other CDKs (i.e., CDK12 and CDK13) are potentially more important for mediating this post-translational modification [101, 102]. Nonetheless, S2 phosphorylation is an important mark for elongating polymerases as it serves as a docking platform to promote recruitment of other elongation and termination factors [103, 104].

1.2.2 Prevalence and implications of promoter proximal pausing

Initially, promoter proximal pausing was characterized as an elongation control mechanism for expression of inducible or viral genes. With the advent of genome-wide sequencing techniques, the widespread prevalence of this phenomenon has become readily apparent. Indeed, chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) analyses of RNA polymerase II generates four classes of genes: untranscribed genes with no polymerase association, untranscribed genes with promoter proximal pausing, transcribed genes with promoter proximal pausing, and transcribed genes without apparent 5' polymerase accumulation [105]. Of the polymerase-associated genes, about 40-70% of these genes have significant 5' accumulation of RNA polymerase II [39, 66, 106-109], suggesting that there are "P-TEFb-dependent" and "P-TEFb-independent" genes. However, when P-TEFb activity is blocked with drug inhibitors, nearly all transcription is halted [39, 66, 110]. This suggests that either promoter proximal

pausing or P-TEFb-mediated phosphorylation events (i.e., DSIF phosphorylation) occur at every RNA polymerase II-associated gene. The variability in 5' polymerase accumulation, however, most likely arises from differential recruitment of both negative and positive elongation factors.

What is the purpose of promoter proximal pausing? Because the paused polymerases are transcriptionally engaged, promoter proximal pausing serves as a convenient mechanism for rapid gene induction [32]. Indeed, gene ontology analysis of genes with high levels of promoter proximal pausing reveals a significant enrichment for processes related to growth, differentiation, proliferation, and stress responses [105]. Because effective release into productive elongation relies on the activity of one main factor, P-TEFb, this mechanism also serves to induce a concerted response upon appropriate stimulation [111, 112]. Promoter proximal pausing is also important for maintaining nucleosome positioning near the promoter, as knockdown of NELF repositions nucleosomes within promoter elements [113]. This, in turn, decreases future rounds of transcription. Finally, promoter proximal pausing may influence proper termination of divergent transcripts at bi-directional promoters by allowing termination factors to associate and efficiently disengage the polymerase [66, 114].

1.3 REGULATION OF PROMOTER PROXIMAL PAUSING

1.3.1 Recruitment of P-TEFb to paused RNA polymerase II

As the principal factor necessary to relieve promoter proximal pausing [59, 99, 115], P-TEFb is subject to strict control mechanisms to ensure proper transcription. Therefore, recruitment of P-TEFb to paused polymerases is not mediated through simple diffusion of active kinase; instead,

it is regulated via specific interactions with *trans*-acting transcription factors [116]. Known interaction/ recruitment factors include BRD4 [117, 118], JMJD6 [119], NF-κB [120], MyoD [121], c-Myc [122], KAP1 [123], CIITA [124], p53 [125], and VP16 [126]. These transcription factors can directly interact with P-TEFb through P-TEFb-interacting domains (e.g., BRD4 [127] and Tat from HIV1 [128]), while others recruit active P-TEFb via association with regulators of P-TEFb (Section 1.4.2) [129].

As more recruitment partners and mechanisms are discovered, a general theme has emerged that typifies P-TEFb engagement at specific loci (adapted from [129]). Under resting conditions, active P-TEFb is repressed through interaction within a ribonucleoprotein complex (Section 1.3.2). This complex can be tethered to enhancers or promoters, with proteins such as KAP1 responsible for maintaining tethered reservoirs of the kinase (Figure 1.6a) [123]. Free ribonucleoprotein complexes can also be recruited through interactions between complex members and specific transcription factors. Upon transcriptional demand, enzymes associated with these transcription factors catalyze a change in PTM status on either P-TEFb or factors repressing its activity, releasing free P-TEFb (Figure 1.6b). Direct P-TEFb-interacting transcription factors or super elongation factors can then associate with the active kinase for specific recruitment to the paused polymerase. Additionally, the histidine-rich motif found in cyclins T1, T2a, and T2b (Figure 1.5) interact with the CTD of RNA polymerase II [130], providing an additional anchoring mechanism.

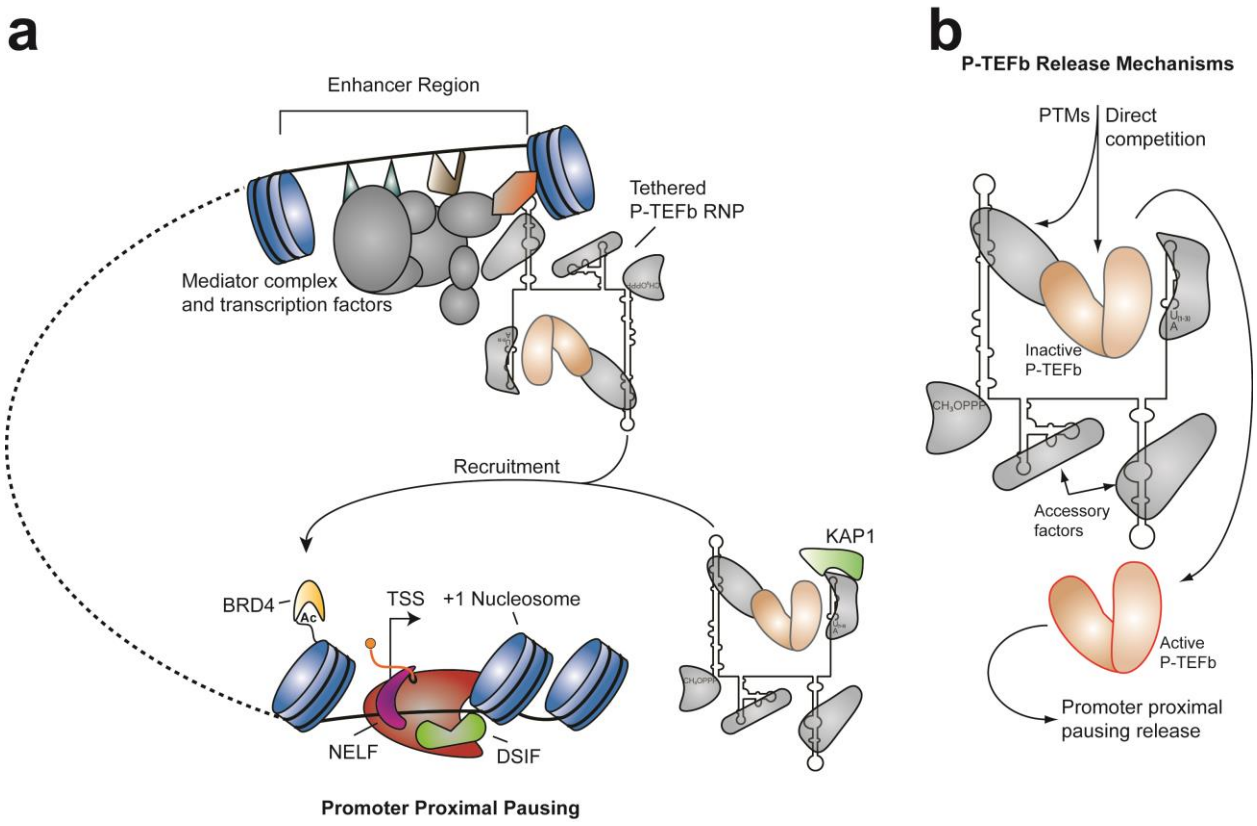

Figure 1.6. General recruitment model for promoter proximal pausing release. Adapted from [129]. (a) P-TEFb, inactivated when bound within a ribonucleoprotein complex (RNP), is tethered or recruited to promoter proximally paused RNA polymerase II. These interactions are mediated through chromatin-associated factors such as BRD4, which binds acetylated histones near promoters [131], and KAP1, which can bind accessory factors within the RNP and recruit the complex [123]. (b) Once recruited, P-TEFb can be released from the RNP through post-translational modifications of P-TEFb or accessory factors as well as by direct recruitment from the complex. Once active, P-TEFb will associate with the polymerase and release promoter proximal pausing.

It is important to keep in mind that promoter proximal pausing is an event that occurs ~ 50 basepairs from the transcription start site [105], and is thus immediately adjacent to both the general transcription factors and transcription activators and repressors acting in *trans* [25]. Therefore, the conditions that either promote or release pausing involve the complex balance between positive and negative elongation factors that are influenced by the three-dimensional architecture of the nucleus. Indeed, promoter-enhancer interactions mediated through chromosomal looping [132] and Mediator [133] have been shown to aid in pausing release, potentially through interactions between Mediator – the enhancer/promoter bridging complex [26] – and components of the super elongation complex [134]. Other enhancers, known as anti-pausing enhancers, specifically recruit and destroy the ribonucleoprotein complex repressing P-TEFb, increasing the levels of active, free P-TEFb near the associated promoters [119].

1.3.2 Multiple levels of regulation affect P-TEFb activity

The critical importance in properly regulating P-TEFb activity cannot be overstated. The first level of regulation occurs with the expression of the kinase itself. The active kinase subunit, CDK9, is expressed as two isoforms resulting from alternative promoter usage, resulting in polypeptides with molecular masses of 42 kDa and 55 kDa (Figure 1.5) [97]. The proportion of expressed isoforms varies between tissue types and across cell differentiation [135]. While the larger isoform appears to undergo constitutive expression, the smaller isoform becomes upregulated upon extracellular signals and activation [135, 136]. In addition, the two CDK9 isoforms can interact with four different cyclins (T1, T2a, T2b, and K – with the two cyclin T2 transcripts arising through alternative splicing of a single gene [87]) to form a total of eight

potential heterodimer complexes (Figure 1.5) [84]. The exact contribution of each complex in regulating P-TEFb function *in vivo* is not fully understood, however.

Once assembled, P-TEFb activity can be modulated through various PTMs. The regulatory cyclin T subunit can undergo acetylation via p300, which releases active P-TEFb from the inhibitory ribonucleoprotein complex (see below) [137]. Like CDK2, CDK9 also possesses a T-loop proximal to the active site. This loop contains a critical threonine (T186) that must be phosphorylated for proper substrate alignment in the active site and full activity (Figure 1.5) [98]. However, unlike CDK2 which requires activation in *trans* [138], CDK9 can autophosphorylate T186 in *cis* [98]. Association of P-TEFb with the repressive 7SK RNA ribonucleoprotein complex requires T186 phosphorylation [139]. Thus, phosphatases such as PPM1G [140] and PP1α [141] not only inhibit P-TEFb activity, but also provide a mechanism by which to release the kinase from the complex. Finally, P-TEFb activity can be indirectly modulated through PTMs of the accessory factor, HEXIM (see below), that represses P-TEFb within the ribonucleoprotein complex. These modifications inhibit the association between either HEXIM and P-TEFb [142-144] or HEXIM and the RNA [145], releasing active P-TEFb.

As I have mentioned throughout the preceding sections, the most notable level of P-TEFb regulation occurs through association with the 7SK small nuclear RNA. Indeed, this novel mechanism by which 7SK RNA serves as a scaffold to form a repressive ribonucleoprotein complex (RNP) with P-TEFb [146, 147] is the predominant mechanism governing active P-TEFb levels in metazoans. Depending on the cell type, nearly 50-90% of the total available P-TEFb is sequestered within 7SK RNPs [143, 146-148], and controlled P-TEFb upregulation (e.g., T-cell activation) concomitantly increases the expression of the inhibitory 7SK RNP components [149, 150]. The function of the inhibitory RNP is to maintain a vast pool of

selectively-recruitable P-TEFb that can be quickly utilized to meet transcriptional demand upon activation. The specifics of this regulatory mechanism will be discussed in the following section.

1.4 7SK RNA: SYNTHESIS, COMPOSITION, AND REGULATION OF TRANSCRIPTION

1.4.1 7SK RNA synthesis and maturation

7SK RNA is an abundant (~2 x 10^5 copies per cell) [151], nuclear [152] non-coding RNA that had an unknown function for nearly three decades until researchers discovered it associated with, and inhibited, P-TEFb [146, 147]. The human genome contains hundreds of pseudogenes encoding 7SK RNA, but the only functional copy is located on chromosome 6 [153, 154]. An RNA polymerase III transcript [155], 7SK RNA is expressed under a strong constitutive promoter [156, 157] that relies on the transcription factors Oct1 [158], SNAPc [159], and Staf1 [160], and terminates upon transcription of a canonical poly-uracil sequence (Figure 1.7) [161]. Its 3' di- or tri-uridylate sequence initially serves as the binding motif for the Lupus antigen (La) protein that protects the transcript from exonuclease activity (Figure 1.7) [162, 163]. Chemical and ribonuclease probing experiments led to a proposed secondary structure containing four stem loops (SI-SIV) [164]. Phylogenetic analysis of 7SK RNAs across metazoans, guided by the strong associated promoter elements, suggests an alternate structure composed of eight highly-conserved motifs (M1-M8) [165]. For the purpose of this thesis, I will defer to the original secondary structural model (Figure 1.7).

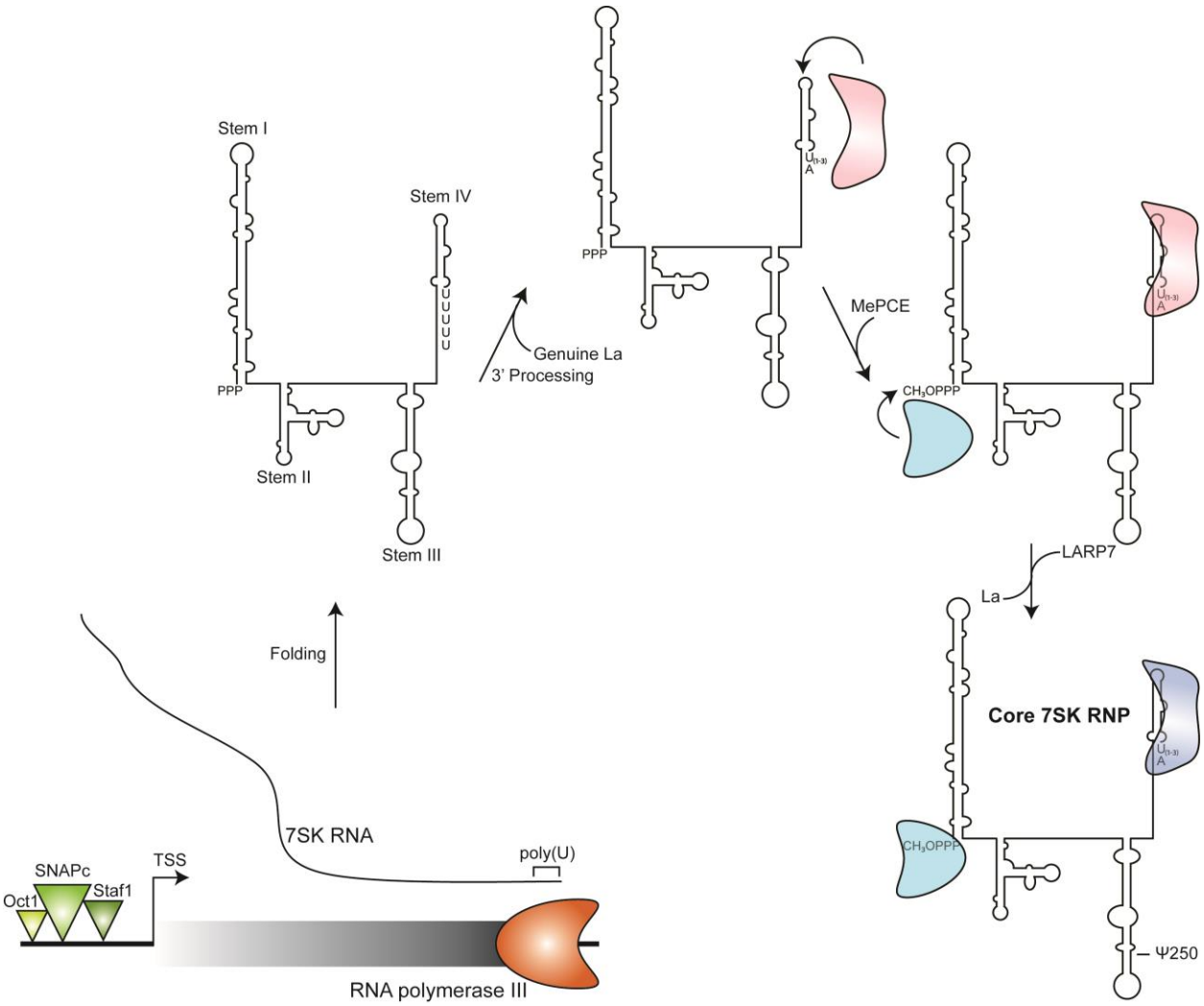

Figure 1.7. Synthesis of the core 7SK RNP. Adapted from [151, 166]. After transcription by RNA polymerase III, 7SK RNA undergoes 3' end processing to remove one, two, or three uracils before addition of an adenine. The 3' end is protected from further degradation by genuine La. The 5' end is capped with a monomethyl group on the gamma phosphate by the methyl-phosphate capping enzyme (MePCE). After methylation, the enzyme remains bound to the 5' end of 7SK RNA. Finally, genuine La is replaced by La-related protein 7 (LARP7), forming the core 7SK RNP.

After synthesis, 7SK RNA undergoes several rounds of post-transcriptional processing. Exonuclease activity trims off one to three uracils before an unknown enzyme adds a single adenosine, giving rise to a heterogenous population of 330-332 nucleotide transcripts [167, 168]. Shortly afterward, La is exchanged for La-related protein 7 (LARP7), which tightly binds and protects the 3' end from further degradation (Figure 1.7) [169-171]. As over 90% of LARP7 is bound to 7SK RNA [166], it appears the sole function of the protein is to aid in 7SK RNA stability and RNP formation [172]. At the 5' end, the methyl phosphate capping enzyme (MePCE), previously known as BCDIN3 [173], adds a single methyl group to the gamma phosphate [174, 175] and binds the transcript (Figure 1.7) [173]. In addition to these terminal processing events, a recent study has found that most 7SK RNA transcripts are pseudouridylated at U250 [176], although the direct role this mark has on 7SK RNA structure and function remains to be discovered.

The 5' and 3' post-transcriptional processing events are critical for 7SK RNA stability. Depletion of either LARP7 [169, 170] or MePCE [173, 177] dramatically reduces 7SK RNA levels. More than merely capping the RNA, LARP7 and MePCE can also directly interact, and 7SK RNA strengthens this interaction [177]. There are two major consequences of the LARP7 – MePCE interaction. 1) LARP7 suppresses MePCE catalysis, preventing removal of the monomethyl phosphate cap [177]. 2) Direct association while binding RNA helps bring together the 5' and 3' ends, forming the core 7SK RNP (Figure 1.7) [178]. From here, 7SK RNA serves as a scaffold to facilitate formation of multiple RNPs that work together to regulate metazoan transcription.

1.4.2 7SK RNA forms multiple RNPs to regulate transcription

To date, 7SK RNA can form four different and mutually exclusive RNPs. The assembly and disassembly of three of the RNPs directly regulate the active concentrations of P-TEFb within the nucleus, while the fourth complex helps suppress RNA polymerase II initiation and elongation at enhancer regions [129]. Comprising a small percentage of the 7SK RNPs, the 7SK-BAF RNP is composed of 7SK RNA and the ATP-dependent nucleosome remodeler BAF complex [179]. Interestingly, this RNP contains neither MePCE nor LARP7, and therefore must be stabilized through an unconventional mechanism. The 7SK-BAF RNP is predominantly recruited to enhancer and super-enhancer regions [179], where the BAF complex positions nucleosomes across enhancer promoters and prevents RNA polymerase II transcription of enhancer RNAs [180].

Formation of the repressive P-TEFb RNP (the 7SK-P-TEFb RNP) relies on the accessory factor hexamethylene bis-acetamide inducible protein (HEXIM) (Figure 1.8) [181, 182]. Humans express two isoforms (HEXIM1 and HEXIM2) arising from adjacent genes on chromosome 17 [166]. The isoforms appear to have overlapping functions, as knock-down of HEXIM1 results in a compensatory increase in HEXIM2 expression [183]. The protein forms a homo- or heterodimer through a C-terminal coiled-coil leucine zipper motif [184, 185]; dimerization is a prerequisite for association with 7SK RNA [184]. Association of the HEXIM dimer with 7SK RNA is required for the binding and inhibition of P-TEFb [184]; a conformational change unmasks the acidic region that binds P-TEFb from an interaction with an arginine-rich motif (ARM) located toward the N-terminus of HEXIM (Figure 1.8) [186, 187]. The ARM is also responsible for mediating dsRNA interactions [188]. Upon binding the distal portion of Stem I of 7SK RNA [189-192], the C-terminal acidic portion interacts with the cyclin T subunit of P-TEFb

(Figure 1.8) [193, 194]. HEXIM contains two aromatic residues, F208 and Y271, that most likely bind within the ATP-binding pocket of CDK9 to restrict the activity of P-TEFb [184]. Once tethered to 7SK RNA through HEXIM, P-TEFb association may be stabilized through interactions between the cyclin T subunit and LARP7 [171, 178] or via direct interaction with Stem IV of 7SK RNA (Figure 1.8) [195].

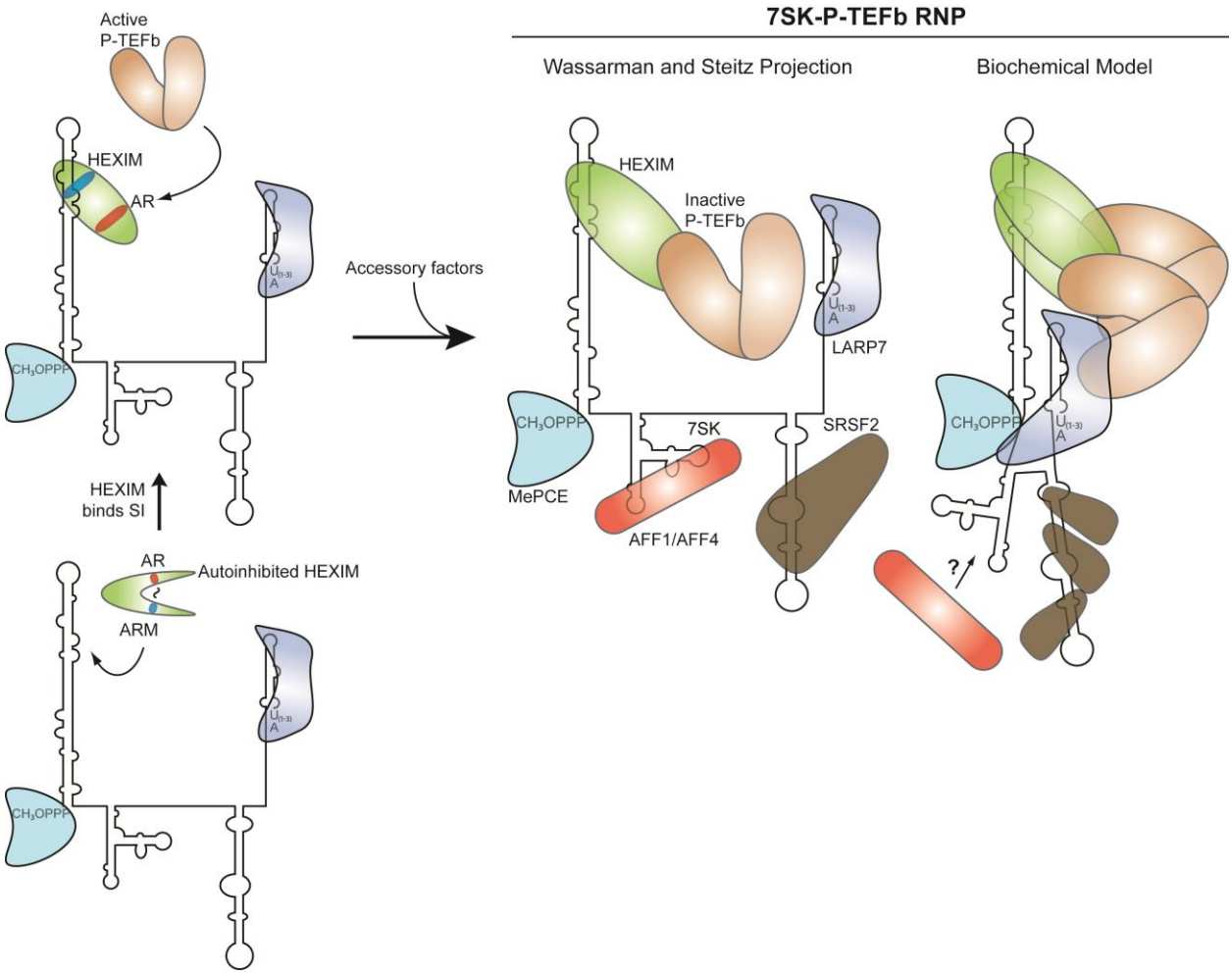

Figure 1.8. Assembly of the 7SK-P-TEFb RNP. Adapted from [129, 151]. HEXIM, initially autoinhibited through interaction between the basic arginine-rich motif (ARM) and acidic region (AR), binds to Stem I of 7SK RNA through the ARM (bottom left). Upon binding, HEXIM can bind and inactivate P-TEFb (top left), forming the core 7SK-P-TEFb RNP. Additional accessory factors may bind to form alternate RNPs (right). The assembled 7SK-P-TEFb RNP is depicted as a basic projection and accurate biochemical model to illustrate the complexity of the particle.

Together, HEXIM and P-TEFb bound to the core 7SK RNP constitutes the 7SK-P-TEFb RNP (Figure 1.8). Additional factors are also implicated in binding this complex, although these interactions and their functions have not been fully characterized. These factors include super elongation complex components AFF1 and AFF4 [196], the splicing protein SRSF2 [197], and the transcriptional repressor CTIP2 [198]. These additional proteins most likely populate specific subsets of 7SK-P-TEFb RNPs and are used to recruit the complex to specific genes.

Once the 7SK-P-TEFb RNP is recruited to a paused RNA polymerase (Section 1.3.1), P-TEFb and HEXIM are released from the complex [146, 147, 199]. Potential mechanisms for inducing this release include post-translational modifications of P-TEFb or HEXIM (Section 1.3.2), demethylation of the 5' monomethyl cap [119], and 7SK RNA restructuring via deposition of the splicing factor SRSF2 onto nascent transcripts [197]. The helicase activity of DDX21 has also been implicated in stimulating this release specifically at ribosomal protein and small nucleolar RNA genes [200]. Excluding activation via 7SK RNA destabilization (removal of the 5' monomethyl cap) [119, 201], P-TEFb and HEXIM dissociation from 7SK RNA does not result in its degradation [170, 202]. Instead, various heterogeneous nuclear ribonucleoproteins (hnRNPs) bind to the remaining core 7SK RNP to create the 7SK-hnRNP RNPs (Figure 1.9). Of the twenty hnRNPs in humans, only hnRNP A1, A2/B1, Q1 and Q3, R, and K have been observed to bind 7SK RNA [202-204]. Surprisingly, these proteins form two mutually exclusive complexes, with hnRNP A1 and A2 binding separately from hnRNP Q and R (Figure 1.9) [204]. It is unknown whether hnRNP K binds either one of the two 7SK-hnRNP RNPs [203], or if the protein forms a unique 7SK-hnRNP RNP. In addition, RNA helicase A has also been found to associate with the 7SK-hnRNP RNP(s) (Figure 1.9) [204].

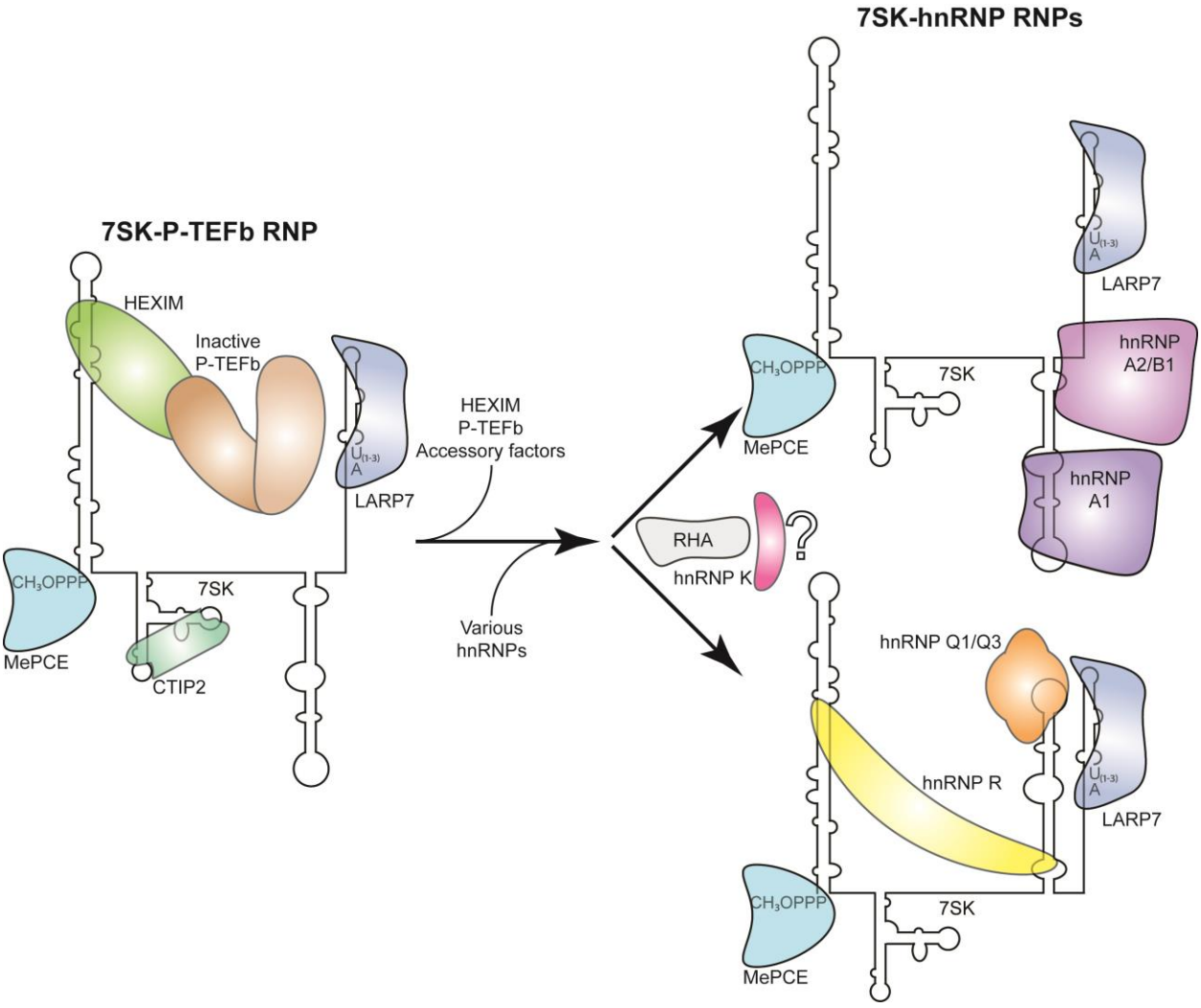

Figure 1.9. 7SK RNA forms two mutually exclusive hnRNP RNPs. Adapted from [129, 151]. Upon dissociation from the 7SK-P-TEFb RNP, the core 7SK RNP binds various hnRNPs. Deletion analysis suggest that all hnRNPs bind Stem III; hnRNPs Q and R also associate with Stem I [204]. hnRNP A1/A2 and hnRNP Q and R have not been found in the same 7SK-hnRNP RNP. The relationship between hnRNP K and RNA helicase A (RHA) and the 7SK-hnRNP RNPs is unknown.

While the precise function of the 7SK-hnRNP RNPs is unknown, several lines of evidence suggest it plays multiple roles in regulating metazoan transcription. First, it may help regulate alternative splicing, as the identified 7SK-associated hnRNPs participate in splice-site selection [205]. Indeed, knock-down of LARP7 or MePCE leads to alternative splicing defects in zebrafish embryos [206]. Second, the balance between the formation of the 7SK-P-TEFb RNP and the 7SK-hnRNP RNPs coordinates the levels of active P-TEFb. Under transcriptional stress (such as UV, actinomycin D, DRB, or flavopiridol exposure), P-TEFb release correlates with increased formation of the 7SK-hnRNP RNPs (reviewed in [129]). Vice versa, knock-down of hnRNP A1 and A2 or hnRNP K augments the formation of the 7SK-P-TEFb RNP [202]. Taken together, these observations imply that 7SK RNA transitions between the two sets of RNPs to control promoter proximal pausing (Figure 1.10).

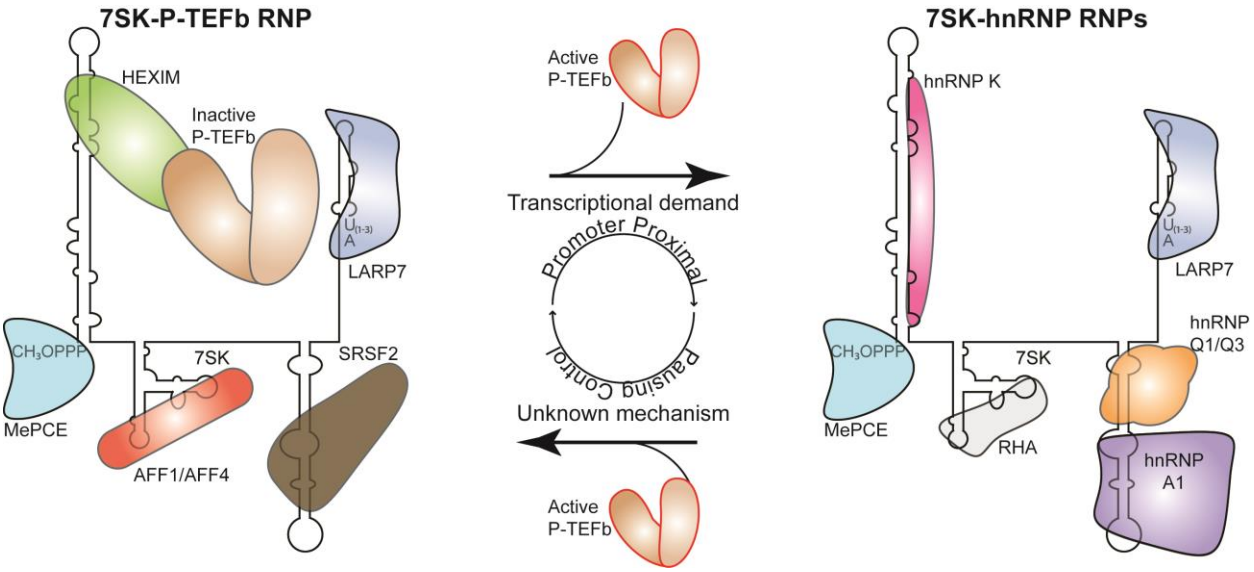

Figure 1.10. The transition between the 7SK RNPs controls metazoan transcription. Adapted from [129, 151]. After recruitment to promoter proximally paused RNA polymerase II, active P-TEFb is released from the inhibitory 7SK-P-TEFb RNP. Various hnRNPs then bind the core 7SK RNP to form the 7SK-hnRNP RNPs. Through an unknown mechanism, hnRNPs are released, upon which HEXIM and other accessory factors may bind and

inactivate P-TEFb, reforming the 7SK-P-TEFb RNP. Together, this cycle fine-tunes the active levels of P-TEFb within metazoan cells.

1.5 OBJECTIVES OF THE PRESENT STUDY

Tight control of P-TEFb ensures proper development, homeostasis, and proliferation of metazoan cells. Indeed, deletion of P-TEFb is embryonic lethal [207], and destabilization of LARP7 leads to increased active P-TEFb levels and metastatic cancer phenotypes in a panel of breast cell lines [208]. As indicated in Figure 1.10, the formation, maintenance, and transition between the 7SK RNPs provide the cell with a multi-variable rheostat to deftly control active P-TEFb levels to meet the transcriptional demands of multi-cellular organisms. There is a large gap in understanding the mechanisms by which these complexes form and transition. In fact, beyond the four initial papers in the early 2000s that described the formation and composition of the 7SK-hnRNP RNPs [170, 202-204], there have been no further studies to explore why these alternate complexes form, nor how the cell controls the transition(s) among them.

In an effort to close this gap, the following chapters will detail two examples of how competition between RNA-binding proteins for 7SK RNA help explain not only the formation of mutually exclusive complexes, but also the mechanisms by which the RNPs transition. Therefore, we focus on pairs of proteins that bind to similar regions, but within different 7SK RNPs. The first pair comprises serine-arginine splicing factor 2 (SRSF2) and hnRNP A1, which are naturally antagonistic splicing factors whose competition for access to pre-mRNA splice-sites upstream of spliceosome recruitment is well-characterized [209]. We find that each protein differentially restructures 7SK RNA upon binding, and that competition between these factors

not only helps maintain RNP formation, but also helps dissociate SRSF2 from the RNP onto nascent RNA transcripts. The second pair, HEXIM and hnRNP K, provide an excellent example of how 7SK RNP control is, in part, modulated through post-translational modifications. Interestingly, we find that P-TEFb itself controls this transition, and suggest that P-TEFb autoregulates active levels and transcriptional processes through phosphorylation of 7SK-associated accessory factors.

While previous studies have examined the interactions between 7SK RNA and hnRNPs via immunoprecipitations, the studies presented herein describe the first detailed biochemical characterizations of the 7SK-hnRNP RNP interactions – and the competition between 7SK-P-TEFb RNP components. They help lay the groundwork for future exploration of all 7SK-associated components and the interaction between them. They also provide glimpses into the complex mechanisms that must synergize to control the transcriptional fate of metazoans. Thus, I hypothesize that 7SK RNA complex formation and transition are the consequences of protein-protein and protein-RNA interaction networks established through direct competition between components. Specifically, because these are (mostly) single-stranded nucleic acid binding proteins, I hypothesize that binding by each factor restructures the RNA in a distinct fashion. This reorganization will mask or expose binding sites to help stabilize recruitment of accessory factors specific for formation of the respective RNP. As most associated factors have separate functions outside of binding the 7SK complex, transition from these RNPs is the result of recruitment away from the complex to attend to these functions. Post-translational modifications may help provide the signal or energy necessary for the "switch." The culmination of these individual signals for each node of the interaction network, therefore, is the driving force for fine-tuning P-TEFb release and association, and thus transcriptional control in metazoans.

2.0 HNRNP A1 AND SRSF2 COMPETITION FOR STEM III OF 7SK RNA AIDS IN RNP CONVERSION AND MAINTENANCE

2.1 INTRODUCTION

Serine-arginine splicing factor 2 (SRSF2) and hnRNP A1 are naturally antagonistic splicing factors. Binding of SRSF2 to exonic splicing enhancers helps recruit the major spliceosome complex, while hnRNP A1 binding to exonic splicing silencers inhibits its formation [34]. More importantly, the two proteins directly compete for access to their respective binding sites on pre-mRNA, whereby binding of one prevents binding of the other [210]. Indeed, competition between these proteins is driven by relative protein abundance and the strength of the binding interactions [211, 212]. Because SRSF2 and hnRNP A1 are predicted to bind to the same stem loop of 7SK RNA, but within mutually exclusive complexes, we wondered if they behave the same way here as on pre-mRNA (Figure 2.1). We hypothesize that hnRNP A1 and SRSF2 compete for access to Stem III, and that this competition restructures the RNA to help maintain formation of one 7SK RNP over the other. In this chapter, I will detail the relationship between the RNA binding domains of SRSF2 and hnRNP A1 with Stem III of 7SK RNA, and how Stem III helps regulate P-TEFb release and RNA polymerase II pausing in cells.

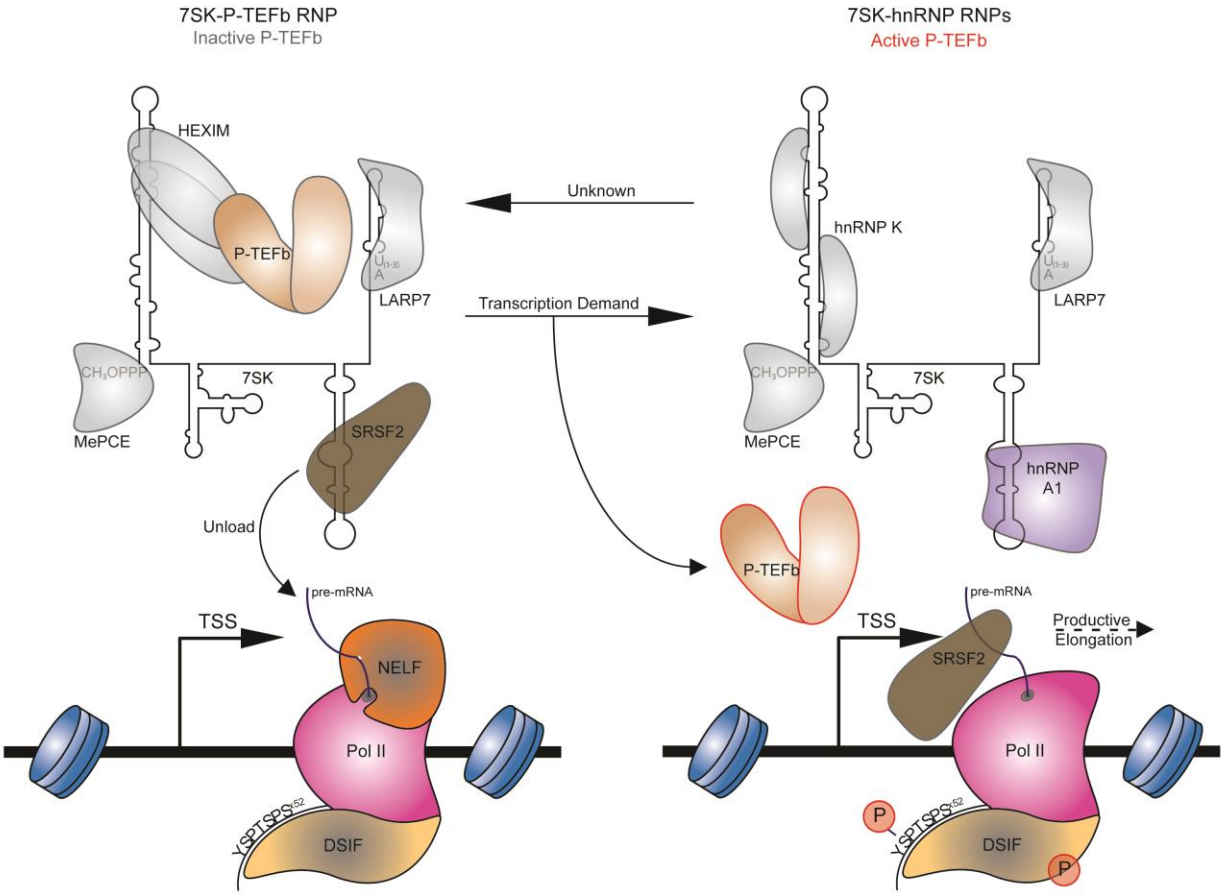

Figure 2.1. 7SK RNA transitions between RNPs to regulate active P-TEFb levels and transcription elongation. P-TEFb is held inactive within the 7SK-P-TEFb RNP and is aided in recruitment to paused RNA polymerase II via SRSF2 (left). Upon transcriptional demand, P-TEFb dissociates from the complex, along with HEXIM and SRSF2, the latter of which binds the nascent pre-mRNA (right). Various hnRNPs, including hnRNP A1, bind 7SK RNA to form the 7SK-hnRNP RNPs.

2.1.1 SRSF2 and hnRNP A1 structure and function

SRSF2 (also known as SC35) belongs to the serine/arginine (SR) family of proteins [213]. Unique among SR proteins, SRSF2 is strictly localized to the nucleus [214] where it is involved in splicing regulation [215] and RNA polymerase II elongation [197, 216]. It is composed of a single N-terminal RNA recognition motif (RRM) followed by a C-terminal RS domain (Figure 2.2a). The RS domain has multiple functions, including non-specific binding of RNA [217] and mediating protein-protein interactions [218]. Additionally, the domain is heavily post-translationally modified, which can determine SRSF2 localization and activity [219, 220]. Specificity for RNA interactions is achieved through the RRM [221], which binds a disparate set of recognition sequences [215]. Nuclear magnetic resonance imaging of the RRM in complex with RNA revealed that the N- and C-termini come together to wrap around and recognize single-stranded RNA, analogous to a closed claw (Figure 2.2a - bottom) [222, 223]. One structural study has also elucidated a minimal recognition sequence of 5'-SSNG-3', where S=C/G and N=A/U/C/G, for the RRM [223].

Like most core hnRNP particle proteins, hnRNP A1 is highly, and almost ubiquitously, expressed in vertebrate tissues [224, 225]. It is involved in an array of RNA metabolic processes, including miRNA biogenesis, alternative splice site selection, nuclear mRNA export, and internal ribosome entry site maintenance [226, 227]. hnRNP A1 contains 2 tandem, antiparallel RRMs [228], an RGG motif that may interact with nucleic acids [229], a glycine-rich C-terminus used for protein-protein interactions [230], and a non-canonical nuclear export/import signal (M9) (Figure 2.2b) [231]. A proteolytic cleavage event results in isolation of just the tandem

RRMs, which were thought to be a different protein named “unwinding protein 1” (UP1) [232, 233].

hnRNP A1 is a single-stranded nucleic acid binding protein, and the crystal structure of UP1 has been solved with both DNA and RNA [228, 234, 235]. While the exact molecular mode of binding is slightly different for DNA versus RNA, both nucleic acids bind along a cleft formed between beta strands 1 and 3 [228, 235]. Specificity for recognizing the SELEX “winner” sequence of 5’-UAGGG(U/A)-3’ is achieved through ionic interactions and base-stacking with aromatic residues such as F17 in RRM1 (Figure 2.2b) [236]. However, domain swapping, deletion, and duplication experiments have shown that the two motifs are non-redundant in binding properties [237]. In recognizing the HIV exon splicing silencer 3 stem loop, only RRM1, aided with aromatic stacking through the linker, makes contact and binds the RNA [235]. Indeed, global analysis of hnRNP A1 interactions highlights unique binding potentials not only for sequence, but also secondary structure, of the RNA [238].

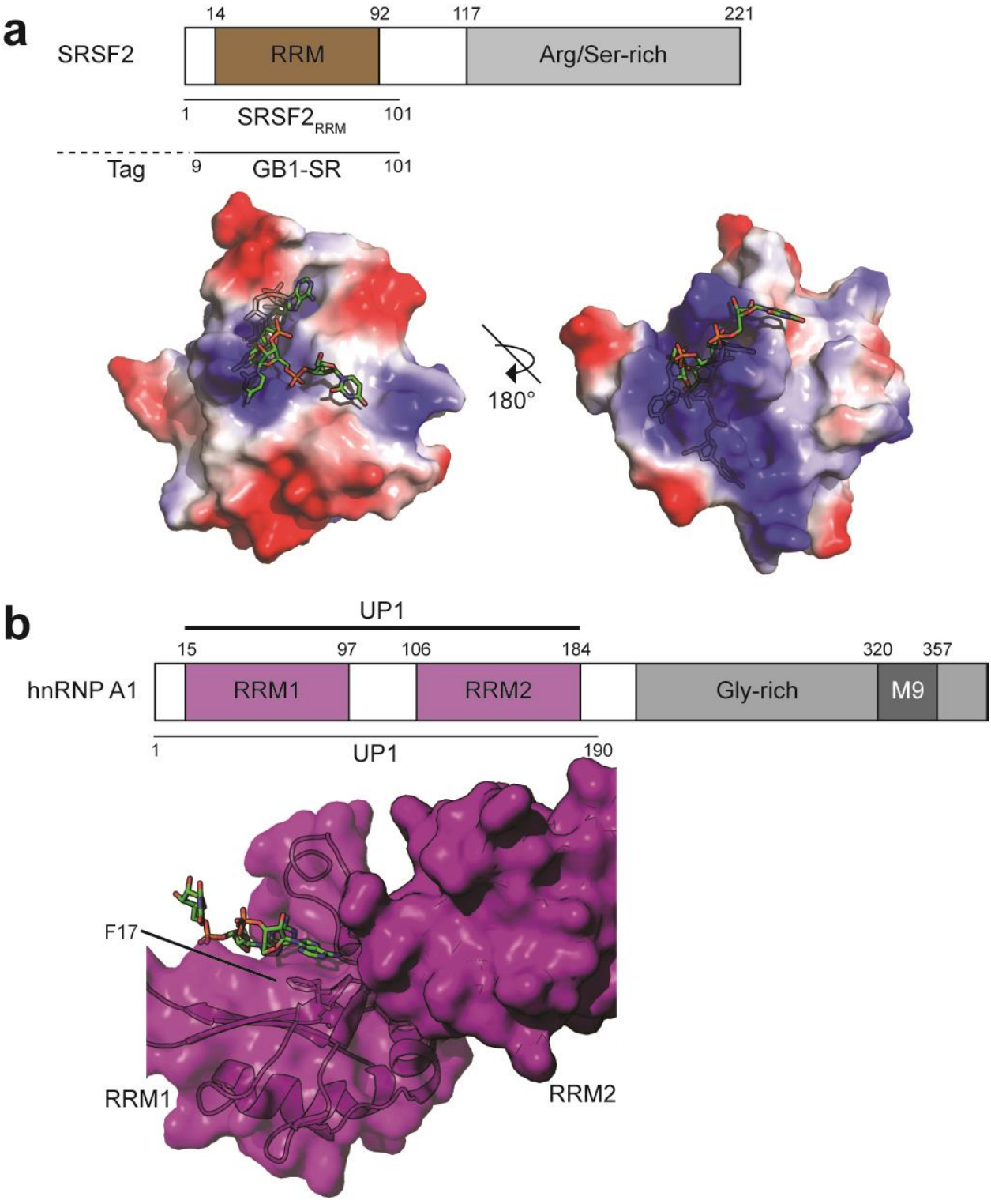

Figure 2.2. Domain architecture and structures of RNA binding domains of SRSF2 and hnRNP A1. (a) SRSF2 domains (top). RRM, RNA recognition motif. Constructs and their corresponding amino acid boundaries used in this study are indicated below the domain diagram. NMR structure of SRSF2 RRM bound to RNA (bottom – PDB ID:

2LEB) [223]. Electrostatic potential is mapped on the surface representation. Red, negatively charged; blue, positively charged. (b) hnRNP A1 domains (top). M9, non-canonical nucleocytoplasmic shuttling sequence. The two tandem RRMs are collectively referred to as UP1 [232, 233]. Construct with corresponding amino acid boundaries used in this study is indicated below. Zoom of the RNA-binding pocket of UP1 from the crystal structure of UP1 bound to RNA (bottom – PDB ID: 4YOE) [235].

2.1.2 SRSF2 and hnRNP A1 associate with 7SK RNA

In 2013, crosslinking immunoprecipitation followed by deep sequencing (CLIP-seq) of serine/arginine splicing factor 2 (SRSF2) in mouse embryonic fibroblasts unexpectedly revealed that the protein binds 7SK RNA [197]. When pulled from cells, SRSF2 co-immunoprecipitates with HEXIM and P-TEFb subunits, placing SRSF2 in the 7SK-P-TEFb RNP. Most of the CLIP-seq signal specifically mapped to the third stem loop (Stem III) of 7SK RNA (Figure 2.1). In contrast to the CLIP-seq results, which showed wide distribution of SRFS2 across gene bodies, chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) of SRSF2 revealed an enrichment of the protein at the very 5' end of gene bodies. Metagene analysis in conjunction with RNA polymerase II ChIP-seq and global run-on sequencing (GRO-seq) revealed that genes with accumulation of SRSF2 5' signal correspond with promoter proximally paused genes and contain SRSF2 recognition sequences within the first 30 nucleotides. Indeed, knock-down of SRSF2 led to an increased accumulation of RNA polymerase II at the promoter. The authors proposed a model in which SRSF2 binds to the 7SK-P-TEFb RNP, helps recruit the complex to paused RNA polymerase II, recognizes the nascent pre-mRNA, and dissociates from the RNP to bind the pre-mRNA (Figure 2.1).

Interestingly, Stem III has also been shown to be the predominant binding site for the majority of the hnRNPs in the 7SK-hnRNP RNPs. Specifically, studies have found that hnRNP

A1 binds 7SK RNA, and this association is dependent on transcriptional demand [170, 202, 204]. Deletion of Stem III reduces hnRNP A1 association as measured via immunoprecipitation, suggesting that hnRNP A1 specifically binds to Stem III or to other protein(s) within this region (Figure 2.1) [204]. Like all hnRNPs within the 7SK-hnRNP RNPs, hnRNP A1 association is mutually exclusive to most members of the 7SK-P-TEFb RNP [170, 202-204], with the exception of MeCPE and LARP7. While the function of hnRNP A1 binding has yet to be elucidated, expression of 7SK RNA without Stem III results in impaired P-TEFb dissociation from the 7SK-P-TEFb RNP, suggesting that hnRNP binding somehow assists in RNP conversion [204].

2.2 MATERIALS AND METHODS

2.2.1 RNA construction design

The transcription template for Stem III was generated using primers Stem III For and Stem III Rev (Appendix A.1) to PCR amplify nucleotides 200-274 from the human 7SK DNA sequence synthesized into pIDTSMART (IDT). The T7 promoter and hepatitis delta virus 3' cleavage overlap were added with PCR to the DNA sequence encoding the hepatitis delta virus ribozyme [239]. Stem III-SHAPE construct was generated from amplification of pIDTSMART-7SK (IDT) using primers Stem III-SHAPE For and Stem III-SHAPE Rev. The T7 promoter, 3' internal hairpin, and reverse transcription primer binding site [240] were inserted with nested PCR. Mutant Stem III constructs were generated using overlap and nested PCR (Appendix A.1). RNA was transcribed *in vitro* at 37°C for 16 hours with final concentrations of 40 mM Tris-HCl, pH 8,

10 mM DTT, 5 mM spermidine, pH 8, 15 mM $MgCl_2$, 4 mM each NTP, and recombinant T7 RNA polymerase that was purified in-house [241]. RNA was gel purified on 8% 29:1 polyacrylamide 1X TBE/7M urea gels, excised, eluted, concentrated, and stored in 10 mM sodium cacodylate, pH 6.5.

2.2.2 Protein purification

UP1 (amino acids 1-190) was cloned from MHS1011-202833012 (Thermo Scientific) using primers UP1 For and UP1 Rev (Appendix A.1) and placed into pET28a using NheI and BamHI (Thermo Scientific) to generate pEU. $SRSF2_{RRM}$ (amino acids 1-101) was cloned from MHS6278-20280894 (Thermo Scientific) using primers $SRSF2_{RRM}$ For and $SRSF2_{RRM}$ Rev, digested with NheI and NotI (Thermo Scientific), and ligated into the NheI and NotI sites of pET28a to generate pES_{RRM}. GB1-SR construct was a generous gift from Lu-Yun Lian (University of Liverpool). RNA-binding null mutants were generated through sequential rounds of site directed mutagenesis using primers in Appendix A.1 [242]. All cloning was confirmed by sequencing (Genewiz). Expression constructs were transformed into BL21 (DE3) *Escherichia coli* and expressed in LB medium with 1 mM IPTG induction at 18°C for 20 hours.

Wild-type and mutant UP1-expressing cells were lysed in lysis buffer (20 mM Tris-HCl, pH 8, 150 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol, 5% glycerol) and purified on a HisTrap HP column (GE Healthcare). Protein was eluted with lysis buffer + 500 mM imidazole, and immediately passed over HiTrap HP Q anion and HiTrap HP SP cation exchange columns (GE Healthcare). Fractions containing UP1 were concentrated to 1.5 mL, injected over a HiLoad Superdex 75 column (GE Healthcare) in lysis buffer without glycerol, concentrated,

buffer exchanged, and stored at -80°C in storage buffer (20 mM Tris-HCl, pH 8, 150 mM NaCl, 1 mM β-mercaptoethanol, 10% glycerol).

$SRSF2_{RRM}$, wild-type GB1-SR, and Y44A GB1-SR constructs were purified with similar protocols. Induced cells were lysed in 20 mM Tris-HCl, pH 7.4, 500 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol, 5% glycerol, and purified in batch using nickel agarose (Thermo Scientific). Protein was eluted by increasing the imidazole concentration to 300 mM. Eluted protein was dialyzed overnight into phosphate buffer (50 mM sodium phosphate, pH 7, 250 mM KCl, 10 mM L-arginine, 1 mM β-mercaptoethanol, 5% glycerol), and applied to tandem HiTrap HP Q anion and HiTrap HP SP cation exchange columns (GE Healthcare). Fractions containing SR protein were concentrated to 1.5 mL, injected over a HiLoad Superdex 75 column (GE Healthcare) in phosphate buffer without glycerol, concentrated, and stored at -80°C in phosphate buffer with 10% glycerol.

2.2.3 Electrophoretic mobility shift assays

RNA was 5' end-labeled with γ-^{32}P-ATP and T4 PNK (Thermo Scientific), gel purified in 8% 29:1 polyacrylamide 1X TBE/7M urea gels, excised, eluted and stored in 10 mM Na cacodylate, pH 6.5. RNA was snap-cooled (95°C for 2 minutes followed by 15 minutes at 4°C) in 100 mM NaCl and 10 mM Na cacodylate, pH 6.5 to generate a single, stably folded species as measured by native electrophoresis and secondary structure probing experiments (see below). Protein was titrated into binding reactions containing 1,000 counts per minute (cpm; <1 nM) annealed RNA, buffer D (20 mM Tris-HCl, pH 7, 0.1 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM EDTA, pH 8, 0.25 mM PMSF, 0.5 mM DTT, 10% glycerol) [212], 2.5 µg yeast total tRNA (Thermo Scientific), and 0.1 µg BSA (NEB), and incubated on ice for 30 minutes. For competition reactions, the first protein

was titrated and incubated with RNA for 45 minutes followed by the addition of the competitor protein and incubation for 35 minutes at 4°C. Reactions were analyzed on either 5% Tris/glycine native 29:1 polyacrylamide gels or 4-20% Tris/boric acid/EDTA native polyacrylamide gels (Bio-Rad), dried, and exposed to an image plate overnight. Plates were scanned on a Fuji FLA-5000 and quantitated with MultiGauge software (Fujifilm).

2.2.4 In-line probing

UP1 was titrated into reactions containing 30,000 cpm annealed RNA (≤ 10 nM), 2.5 µg yeast total tRNA (Thermo Scientific), 0.1 µg BSA (NEB), 10 U RiboLock (Thermo Scientific), and in-line buffer (50 mM Tris-HCl, pH 8.3, 30 mM $MgCl_2$, 100 mM KCl) and incubated at 20°C overnight. Prior to loading samples on a gel, an equal volume of gel loading buffer (10 M urea, 1.5 mM EDTA, pH 8) was added to all reactions. The alkaline digestion ladder was generated by adding 1 µL (~20,000 cpm) 5' end-labeled RNA to alkaline digestion buffer (5 mM Na_2CO_3, pH 11.7, 0.1 mM EDTA, pH 8) and heated at 95°C for 4 minutes. After briefly cooling on ice, an equal volume of 2X formamide buffer (95% deionized formamide, 0.01% bromophenol blue, 0.01% xylene cyanol, 5 mM EDTA, pH 8) was added. The RNase T1 ladder was generated by incubating 1 µL 5' end-labeled RNA in 7 µL gel loading buffer, 1 U of RNase T1 (Ambion), and 1 µL sodium citrate buffer (0.25 M sodium citrate, pH 5) at 50°C for 10 minutes, cooling on ice, and supplementing with 3 µL gel loading buffer and 7 µL ddH_2O. All reactions were loaded on an 8% 29:1 polyacrylamide 7 M urea/1X TBE sequencing gel, run at 90W for 80 minutes, dried, and exposed overnight to an image plate. Plates were scanned with a Typhoon FLA-7000 and analyzed with semi-automated footprinting analysis (SAFA) software [243].

2.2.5 DMS probing

SIII-SHAPE RNA was annealed as described above. Protein was titrated into reactions containing 1 µM RNA, 0.1 µg BSA (NEB), 2.5 µg yeast total tRNA (Thermo Scientific), and buffer D and incubated at 4°C for 30 minutes. Dimethyl sulfate (Sigma) diluted in ethanol was added to a final concentration of 1% to binding reactions and incubated at room temperature for 15 minutes. Reactions containing $SRSF2_{RRM}$ were quenched with 30% β-mercaptoethanol and 300 mM sodium acetate, pH 5.4, prior to phenol/chloroform extraction; UP1 reactions were not quenched but were phenol/chloroform extracted. All reactions were ethanol precipitated and resuspended in 10 µL 0.5X TE. Reverse transcription was carried out using 5 µL of ethanol precipitated sample with end-labeled cDNA primer (5' – GAACCGGACCGAAGCCCG – 3'), enzyme mix (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 10 mM DTT, 4 mM $MgCl_2$, 20 U RiboLock (Thermo Scientific), 0.5 mM each dNTP (Thermo Scientific), and 100 U of SuperScript III (Invitrogen) at 55°C for 15 minutes. Reactions were stopped with 150 mM NaOH and heated at 95°C for 5 minutes, followed by neutralization with acid stop mix (4:25 v/v mixture of 1 M unbuffered Tris-HCl and stop dye (85% formamide, 0.5X TBE, 50 mM EDTA, pH 8, 0.01% bromophenol blue, 0.01% xylene cyanol)). Reactions were separated on 8% 29:1 polyacrylamide 7 M urea/1X TBE sequencing gels, dried, and exposed overnight to an image plate. Plates were scanned with a Typhoon FLA-7000 and analyzed using SAFA [243].

2.2.6 RNase T1 probing

Protein was titrated into reactions containing 10,000 cpm folded RNA (< 5 nM), 2.5 µg yeast total tRNA (Thermo Scientific), 0.1 µg BSA (NEB), 10 U RiboLock (Thermo Scientific), and

buffer D (Section 2.2.3) and incubated at 4°C for 30 minutes. 0.1 U of RNase T1 (Ambion) was added, and reactions were incubated at 20°C for 10 minutes. Reactions were phenol/chloroform extracted, ethanol precipitated, and resuspended in 10 μL 2X formamide buffer. Alkaline digestion and RNase T1 ladders were generated as described above. All reactions were separated on 8% 29:1 polyacrylamide 7 M urea / 1X TBE sequencing gels, dried, and exposed to an image plate. Plates were scanned with a Typhoon FLA-7000 and analyzed with SAFA [243].

2.2.7 Cell culture and transfection

HEK 293 H cells were cultured under standard conditions with DMEM supplemented with 10% FBS. For transfections, low density-plated cells were transfected with 10 μg of plasmid or lipofectamine 2000 (Invitrogen) alone (mock transfection). After 24 hours, media was changed and re-transfected with 2.5 nM siRNA or lipofectamine alone (mock transfection).

2.2.8 Differential salt extraction

24 hours after transfection, cells were equilibrated to room temperature for 5 minutes before adding 20 μL DMSO or 0.1 mM flavopiridol (Sigma-Aldrich) diluted in DMSO and incubated at room temperature for the indicated time. Cells were then harvested on ice, washed once with ice-cold 1X PBS, and differential salt extraction was performed as described [244]. 10 μL 4X SDS loading dye was added to each sample, heated at 85°C for 10 minutes, and 15 μL were loaded on 10% SDS polyacrylamide gels. Gels were transferred to Protran BA 85 nitrocellulose membranes (GE Healthcare), blocked with 3% milk in TBST for 1 hour at room temperature, and incubated with 1:100 α-CDK9 antibody (Santa Cruz, sc-13130) diluted in 3% milk overnight

at 4°C. Blots were washed 3x for 10 minutes with TBST and incubated with 1:1000 α-mouse-HRP antibody (Santa Cruz, sc-516102) diluted in 2% milk for 1 hour at 4°C. After washing 3x for 10 minutes with TBST, blots were imaged on an Amersham Imager 600 with Pierce ECL Western Blotting Substrate (Thermo Scientific) and quantified with Image J [245].

2.2.9 UV stress assay

24 hours after transfection, cells were moved from 37 °C to room temperature and equilibrated for 5 minutes in a hood. Lids were removed, and the cells were irradiated with source UV in the hood for the indicated time. At each time point, total RNA was extracted as detailed below. The following equation was used to calculate the relative elongation ratio: $RER = 2_{HSPA1B}^{(Promoter - Gene\ Body)} * (2_{ACTB}^{(Promoter_{UV+} - Gene\ Body_{UV+})} / (2_{ACTB}^{(Promoter_{UV-} - Gene\ Body_{UV-})})$, where "promoter" and "gene body" refer to C_t values obtained from RT-qPCR with (UV+) or without (UV-) irradiation. "HSPA1B" and "ACTB" reference the genes analyzed. All C_t values were normalized to TBP transcript abundance.

2.2.10 Cellular RNA purification and reverse transcription

After media removal, 0.4 mL of Trizol (Invitrogen) was used for RNA extraction as per manufacturer's protocol. RNA pellets were resuspended in 30 μL water. 50 ng of total RNA was subjected to reverse transcription using the QuantiTect kit (Qiagen) as per manufacturer's protocol.

2.2.11 RT-PCR and RT-qPCR

RT-PCR reactions containing 1 μL cDNA, 0.25 mM dNTPs, 400 nM forward and 400 nM reverse primers, 1x Green DreamTaq (Thermo Scientific) buffer, and 0.1 U DreamTaq polymerase were subjected to 18 cycles (fl7SK) or 24 cycles (β-Actin and GREB1) of PCR. 10 μL of sample were run on 1% TBE agarose gels, imaged on an Amersham Imager 600, and quantified with Image J [245]. RT-qPCR reactions containing 1 μL of 0.5X cDNA, 5 μL iTaq SYBR master mix (Bio-Rad), and 400 nM forward and 400 nM reverse primers were subjected to 40 cycles of PCR on an Applied Biosystems StepOne Plus Real-time Thermocycler. C_t values were calculated with StepOne Software v2.2.2.

2.3 RESULTS

2.3.1 UP1 and SR directly bind Stem III of 7SK RNA

While CLIP-seq and co-immunoprecipitation experiments demonstrated that hnRNP A1 and SRSF2 associate with 7SK RNA Stem III in cells [170, 197], it is not clear whether this association is direct or indirect. We predicted these interactions would be direct because several putative consensus hnRNP A1 (5'-UAGGG(U/A)-3') [236] and SRSF2 (5'-SSNG-3') [223] binding sites are located within the sequence of Stem III (Figure 2.3a). However, most of these are buried within base pairs of the stem, which would hinder recognition by these single-stranded RNA binding proteins [222, 228]. To test the interactions of these proteins and 7SK RNA, we performed electrophoretic mobility shift assays (EMSAs) with Stem III RNA (nucleotides 200-

274) and the RNA-binding domains from human hnRNP A1 and SRSF2. We used only the RNA-binding domains because the C-termini of these proteins aid in both cooperative protein binding [218, 230] and nonspecific RNA interactions [229, 246]. Therefore, the constructs used in these experiments are denoted UP1 and SR (Figure 2.2). In the presence of 0.1 mg/mL BSA and 11 μM tRNA nonspecific competitors, UP1 binds Stem III with a modest K_d of 0.307 ± 0.056 μM (Figure 2.3b), while SR poorly binds Stem III with a K_d of 10.7 ± 1.7 μM (Figure 2.3c). Analysis of the binding curves revealed these proteins bind with a high degree of cooperativity, as measured by the Hill coefficient. UP1 and SR have Hill coefficients of 4.2 and 3.9, respectively (Table 2.1). These results demonstrate a direct interaction of 7SK RNA with the RNA binding domains of hnRNP A1 and SRSF2. The binding affinity of SR, however, suggests that the full-length SRSF2 protein is required to accomplish 7SK RNA binding *in vivo* or that SRSF2 relies on other factors to facilitate access to its binding sites within Stem III.

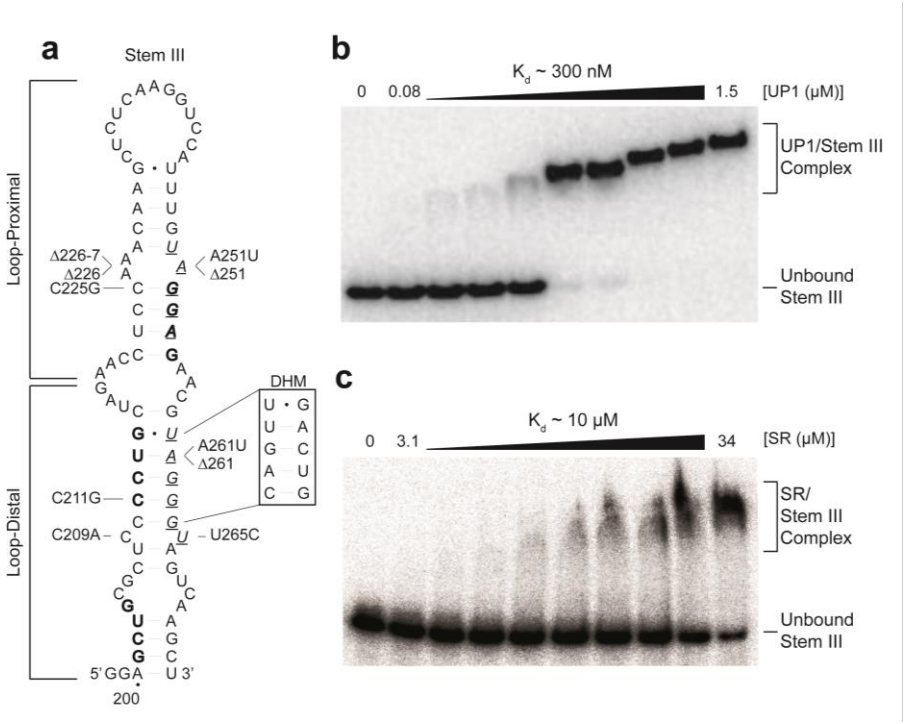

Figure 2.3. UP1 and SR bind Stem III of 7SK RNA. (a) Predicted secondary structure of 7SK Stem III [164]. Boldface, putative SRSF2 binding sites; italic and underlined, predicted hnRNP A1 binding sites. Mutants relevant to this study are indicated; DHM, distal helix mutant. Numbers refer to the sequence in the context of full-length 7SK (GenBank X04236.1). Electrophoretic mobility shift assays of (b) UP1 and (c) $SRSF2_{RRM}$ with trace 5'-radiolabeled Stem III.

2.3.2 Loop-distal site is critical for UP1 interaction with Stem III

To further probe the interactions between UP1 and Stem III, we assessed the affinity of UP1 for several point mutants of Stem III. We mutated two groups of residues: (1) around the putative hnRNP A1 recognition site located nearer to the loop than to the base of Stem III, which we denote as the loop-proximal binding sequence (U250-A254) (Figure 2.3a) [236] and (2) around the putative loop-distal hnRNP A1 binding sequence (U260-U265).

With the exception of Δ227, none of the mutations targeting the putative loop-proximal binding site significantly affect the affinity of UP1 for Stem III (Table 2.1) as measured by EMSAs. Removal of one of the adenosine bulges opposite this binding site (Δ227) decreases the affinity of UP1 for Stem III 1.5-fold. While these mutations do not affect the affinity of UP1 for Stem III, all mutations decrease the cooperativity observed. Mutations C225G, Δ227-8/Δ251, and A251U decrease the cooperative binding of UP1 ~1.5-fold.

Table 2.1. Quantitation of the affinity of UP1 and SR for Stem III and mutants. Pixel density of shifted Stem III was plotted against protein concentration to determine the K_d for individual EMSAs. SR shifts used $SRSF2_{RRM}$ construct. Average K_d values are in μM ± standard deviation calculated from three replicates. "*" indicates 1.5-fold change relative to wild-type. "ǂ" indicates two-fold change relative to wild-type. "N.D.," could not be determined.

		UP1				SR			
		K_d	Relative to WT	Hill Coefficient	Relative to WT	K_d	Relative to WT	Hill Coefficient	Relative to WT
	WT	0.307 ± 0.056	1.00	4.20 ± 0.70	1.00	10.67 ± 1.72	1.00	3.93 ± 0.78	1.00
Loop-Proximal Mutations	C225G	0.325 ± 0.014	1.06	2.49 ± 0.82	0.59*	7.40 ± 3.32	0.69	1.97 ± 0.39	0.50ǂ
	Δ227	0.471 ± 0.106	1.53*	3.12 ± 0.25	0.74	17.27 ± 3.72	1.62*	3.93 ± 2.14	1.00
	Δ227-8	0.336 ± 0.054	1.09	3.05 ± 0.32	0.73	15.19 ± 4.60	1.42	2.62 ± 0.74	0.67
	Δ227-8/ Δ251	0.357 ± 0.020	1.16	2.79 ± 0.38	0.66*	N.D.	N.D.	N.D.	N.D.
	A251U	0.395 ± 0.056	1.29	2.23 ± 0.16	0.53*	20.64 ± 4.63	1.93*	3.32 ± 1.35	0.85
	Δ251	0.327 ± 0.030	1.06	2.96 ± 0.39	0.71	16.84 ± 3.55	1.58*	3.45 ± 2.39	0.88
Loop-Distal Mutations	C209A	1.079 ± 0.224	3.51ǂ	5.24 ± 0.83	1.25	13.82 ± 6.39	1.30	2.64 ± 2.11	0.67
	C211G	0.120 ± 0.011	0.39ǂ	2.77 ± 0.27	0.66*	5.21 ± 0.82	0.49ǂ	2.50 ± 0.26	0.64*
	DHM	0.362 ± 0.122	1.18	3.26 ± 0.48	0.78	3.80 ± 1.85	0.36ǂ	1.86 ± 0.10	0.47ǂ
	A261U	0.379 ± 0.017	1.23	3.28 ± 0.14	0.78	7.33 ± 2.11	0.69	2.77 ± 0.58	0.70
	Δ261	0.271 ± 0.022	0.88	2.70 ± 0.32	0.64*	4.34 ± 2.19	0.41ǂ	1.99 ± 0.32	0.51*
	U265C	0.278 ± 0.010	0.90	2.76 ± 0.60	0.66*	8.98 ± 2.74	0.84	2.97 ± 0.52	0.75

Mutations near to and within the putative loop-distal binding site have varying effects. Mutations C209A and C211G significantly alter the ability of UP1 to bind to Stem III. C209A, which closes the C/U bulge at the 3' end of the predicted binding site, increases the K_d of UP1 for Stem III 3.5-fold to 1.08 μM (student's T-test p=0.02). Conversely, inserting a bulge across from the binding site, with C211G, enhances the affinity 2.6-fold, resulting in a dissociation

constant of 120 nM (student's T-test p=0.03). While mutations within the loop-distal site– A261U, Δ261, and U265C– do not affect the affinity, cooperativity is decreased (Table 2.1).

Together, these data suggest that the loop-distal site is critical for the interaction of UP1 with Stem III. Of note, the higher affinity for this site may be due to its location at the base of the hairpin, which may exhibit more base pair breathing without the constraints potentially placed on the structure by the flanking regions in the full-length RNA. By increasing or decreasing the accessibility of this site, the affinity is significantly enhanced or diminished, respectively. However, most mutations, i.e. – loop-proximal mutations, do not significantly affect the K_d. Rather, they decrease the cooperativity. We interpret these data to mean that the affinity of UP1 for Stem III is primarily driven by initial binding to the loop-distal site.

2.3.3 SR requires access to buried binding sites

SR weakly binds Stem III with 10.67 μM affinity (Figure 2.3c). Two of the three putative SRSF2 recognition sites are hidden among base paired secondary structures; we speculated that the observed poor affinity resulted from the inability of SR to melt Stem III and find these sites (Figure 2.3a). To test this, we utilized the mutants described above in EMSAs with SR. Introducing bulges across from or within predicted SRSF2 binding sites, while preserving the consensus 5'-SSNG-3' sequence [223], enhances the affinity of SR to 7.4 ± 3.32 μM (C225G) and 5.21 ± 0.82 μM (C211G) (student's T-test $p<0.001$) (Table 2.1). Accordingly, mutating the loop-distal helix to shift the consensus SSNG binding site from C211-G214 to C209-C211G, effectively unpairing the first nucleotide of this new putative SRSF2 binding site (mutant DHM), results in a significant two-fold tighter affinity of SR for Stem III (student's T-test $p<0.001$) (Table 2.1). Interestingly, adding a bulge in the middle of one of the binding sites with mutation

A261U does not significantly affect the affinity. These data suggest that SR binding within a structured RNA is strongly influenced by accessibility of the first nucleotide in the binding sequence.

As expected, melting the helices that contain putative SRSF2 binding sites facilitates the interaction of SR and Stem III, while removing bulges around these sites hinders SR binding. Mutations within the loop-proximal adenosine bulges increase the dissociation constant, with Δ227-8 and Δ251 decreasing the affinity ~1.5-fold (Table 2.1). Accordingly, deletion of all three bulges, Δ227-8/Δ251, obliterates SR binding. Eliminating the loop-distal C/U bulge (C209A) also marginally increases the K_d. Together, these data suggest that the poor affinity of SR for Stem III results from the inability of SR to unwind Stem III to access its binding sites and implies a secondary intra- or inter-molecular factor is necessary for efficient association *in vivo*.

2.3.4 UP1 binding opens Stem III

To understand the structural basis for the mutually exclusive occupancy of 7SK RNA by hnRNP A1 or SRSF2, we subjected Stem III to enzymatic and chemical structure probing in the presence of UP1 or SR. Treatment of Stem III with RNase T1, which recognizes and cleaves after single stranded guanosines [247], yields cleavage sites after G218, G240, and G241 in agreement with the predicted secondary structure of Stem III presented by Wassarman and Steitz [164] (Figure 2.4). We also observed cleavage after G214 and G232, which are predicted to adopt wobble pairs with U260 and U246, respectively.

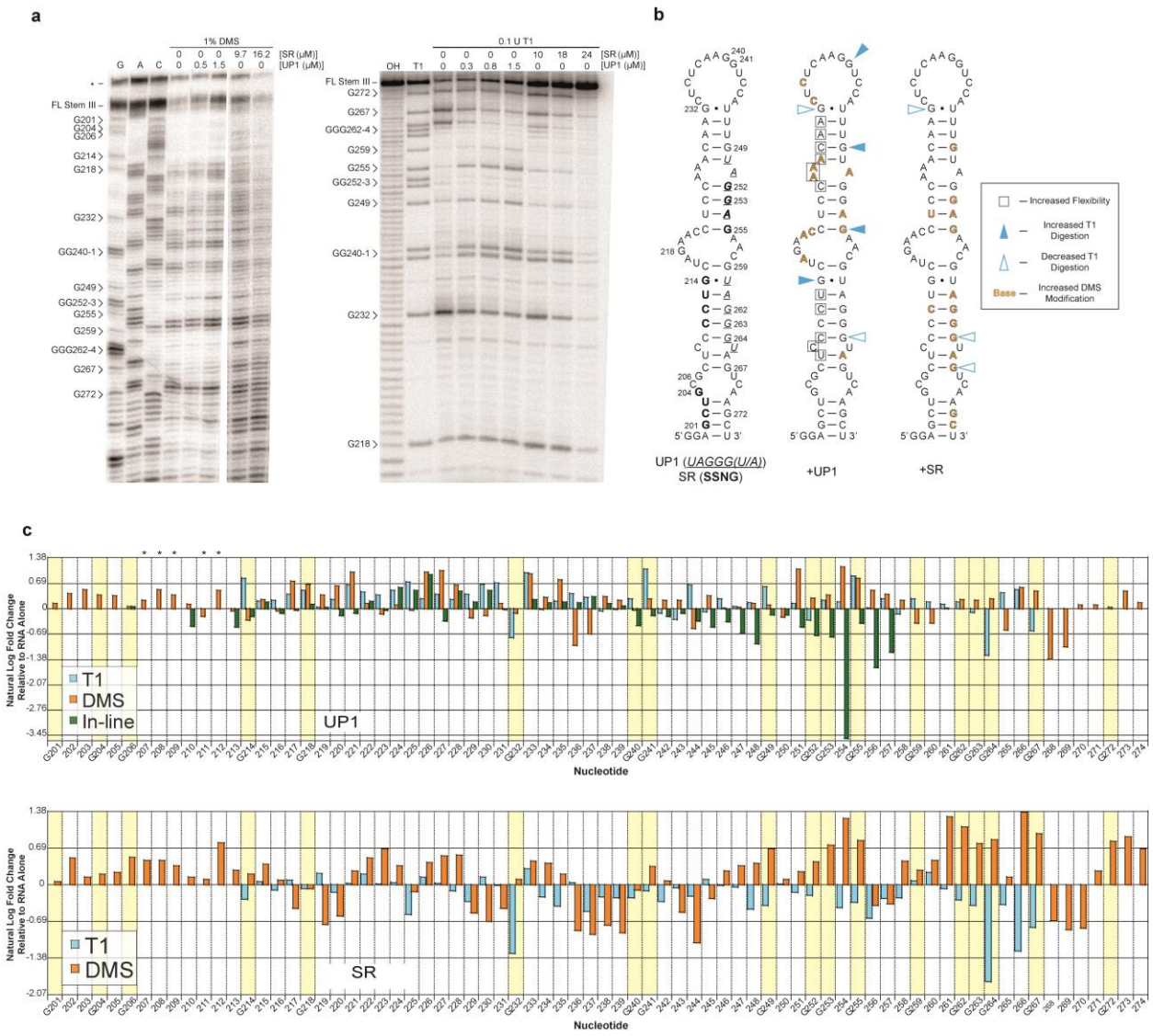

Figure 2.4. UP1 and SR restructure Stem III upon binding. (a) Representative denaturing sequencing gels of chemical and enzymatic probing experiments. Left, DMS treatment of Stem III bound to UP1 or $SRSF2_{RRM}$. G, A, and C, dideoxy-terminated sequencing lanes. "*," full-length SHAPE-Stem III construct. Right, RNase T1 digestion of Stem III bound to UP1 or $SRSF2_{RRM}$. OH, Stem III digested under alkaline conditions to reveal individual nucleotides. T1, denatured Stem III digested with RNase T1 to generate ladder of guanosine-terminated fragments for orientation. (b) Left, putative binding sites of SRSF2 (bold) and hnRNP A1 (italic and underlined). Middle, nucleotides with an UP1-dependent two-fold change in modification or cleavage. Filled arrowheads, increased digestion by RNase T1; outlined arrowheads, decreased digestion by RNase T1; orange outline, increased modification by DMS; boxed, increased cleavage by in-line probing. Right, nucleotides with two-fold change in

modification or cleavage when bound by $SRSF2_{RRM}$. (c) Semi-quantitation of fold-change in cleavage or modification of each nucleotide in the presence of UP1 or $SRSF2_{RRM}$ relative to unbound Stem III. "*" indicates nucleotides that are cleaved in the presence of UP1 but are not flexible in naked Stem III as measured by in-line probing. The position of each G in Stem III is highlighted in yellow for reference.

UP1 binding to Stem III imparts a conformational change to the RNA. Increased RNase T1 cleavage is observed on the 5' side of Stem III, opposite of the two predicted UP1 binding sites (Figure 2.4b). Of note, bands representing G214, G241, and G255 are twice as intense in the presence of UP1 as compared to Stem III alone, suggesting that UP1 binding opens up the stem (Figure 2.4c), consistent with its name of 'unwinding protein.' Accordingly, in-line probing, which measures the general flexibility of RNA via controlled self-cleavage events [248], reveals increased flexibility in nucleotides 207-212 and 224-230, which base pair with the putative binding sites. Decreased flexibility is consistently observed for nucleotides 247-257, which are located within the predicted loop-proximal binding site (Figure 2.4c).

We also conducted DMS probing [249] to fill in the structural details of other nucleotides in Stem III. Increased modification of the adenosines and cytosines in the 5' half of Stem III is observed in the presence of UP1, consistent with the RNase T1 and in-line probing data. Additionally, we observed a general lack of modification throughout the 3' half of Stem III (Figure 2.4c). Together these data indicate that UP1 binds to the 3' half of Stem III, causing Stem III to adopt a flexible, unpaired 5' half.

2.3.5 SR tightly packs Stem III

We next investigated the structural rearrangement of Stem III induced by SR binding. We expected to observe increased RNase T1 sensitivity of nucleotides that base pair with SRSF2 consensus binding sites (5'-SSNG-3') (Figure 2.4b) [223]. To our surprise, we observed a marked reduction in cleavage across the entire stem (Figure 2.4c) for both RNase T1 and minor RNase A contamination. This result suggests that SR may non-specifically coat Stem III, especially at the high concentrations used in our experiments.

To further investigate this, we turned to chemical structure probing methods. Treating SR-bound Stem III with DMS reveals a dramatic SR-dependent rearrangement of Stem III. Most of the adenosines and cytosines in the 3' half of the stem are extensively modified by DMS. We also observed an increase in reverse transcription stops corresponding to guanosines in the 3' half of Stem III, which, while rare, can be modified by DMS [250] and may represent either keto guanine N7 or enol guanine N1 methylation [250, 251]. In fact, a genome-wide sequencing study of DMS modification in *Saccharomyces cerevisiae* found nearly a quarter of the significant modifications mapped to guanosines [252]. We also observed a decrease in modification of nucleotides in the bulges and loop. Modification of nucleotides on both sides of the central bulge (A217, A219, A220/ A256, A257) suggests that the central bulge closes or becomes protected upon SR binding; loop nucleotides (236-245) are less modified upon SR binding, suggesting the collapse of the loop or protection by the protein. The modification of nucleotides in the 5' half of Stem III by DMS increases by less than 2-fold (Figure 2.4c).

While nuclease probing implies non-specific coating of Stem III by SR, the DMS data instead reveals a dramatic rearrangement of Stem III in which the stem becomes tightly packed upon SR binding. The loop and central bulges collapse, while nucleotides across from the

putative SR binding sites become highly accessible to modifications. The simplest interpretation of these data is that SR binding to its consensus sequence causes the phosphate backbone of the previously paired nucleotides to become constrained by the protein. This may flip the bases of the constrained strand outwards towards the solution, exposing them to small-molecule modification, yet hindering the accessibility of the sugar-phosphate backbone to RNase active sites.

2.3.6 UP1 and SR differentially compete for Stem III binding

To date, hnRNPs have not been found to associate with 7SK RNA in the presence of P-TEFb and HEXIM, leading to the prediction that 7SK RNA forms two mutually exclusive sets of RNPs [170, 202-204]. Because SRSF2 co-immunoprecipitates with HEXIM and P-TEFb [197], we hypothesized that hnRNP A1 and SRSF2 compete for binding to Stem III. To test this hypothesis, we conducted competition experiments and analyzed them by EMSA. Because the solubility of the SR construct ($SRSF2_{RRM}$) was not optimal, we expressed and purified a recombinant $SRSF2_{RRM}$ construct fused to a small solubility tag, which we denote as GB1-SR (Figure 2.2a) [222]. The K_d measured for the GB1-SR/Stem III interaction is 10 μM, similar to that of the $SRSF2_{RRM}$ construct (data not shown). RNase T1 digestion of the GB1-SR/Stem III complex displays the same protection pattern as that seen for the $SRSF2_{RRM}$/Stem III complex treated with RNase T1, albeit requiring higher concentrations of the protein (Figure 2.5).

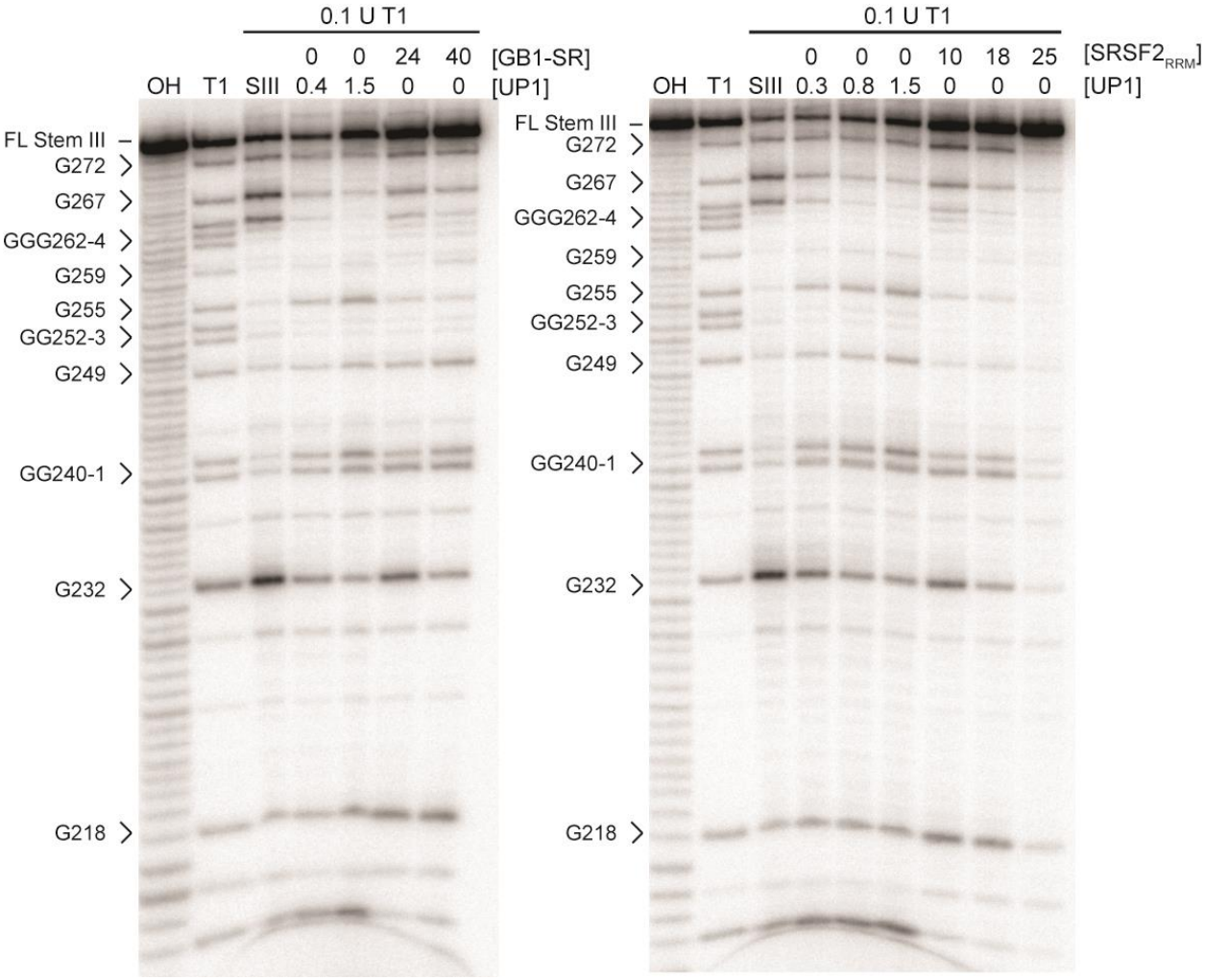

Figure 2.5. GB1 solubility tag does not change the mode of binding of SRSF2$_{RRM}$ to Stem III. RNase T1 digestion of Stem III bound to GB1-SR (left) or SRSF2$_{RRM}$ (right); concentrations are in micromolar. UP1 is included on both gels as a control. OH, Stem III digested under alkaline conditions to reveal individual nucleotides. T1, denatured Stem III digested with RNase T1 to generate ladder of guanosine-terminated fragments for orientation. SIII, folded Stem III digested without the addition of UP1, GB1-SR, or SRSF2$_{RRM}$. Note that high concentrations of GB1-SR (40 μM) reveal a similar cleavage pattern as 18 μM SRSF2$_{RRM}$.

As observed by native gel electrophoresis, when Stem III is pre-bound with a saturating amount of UP1, addition of SR produces a shifted band with a slight increase in mobility. At the highest concentration of SR, 40 μM, the shifted band bifurcates (Figure 2.6a). Conversely, Stem III pre-bound with SR produces two distinct shifts: complex(es) of heterogeneous conformation (a smear above the free probe) and a distinct shift with lower mobility (Figure 2.3c and Figure 2.6). Titration of UP1 condenses the smeared shifts into a single band, with higher concentrations of UP1 reducing the mobility of this band (Figure 2.6b); excess UP1 also binds the free Stem III in the reactions. Of note, the initial SR/Stem III low-mobility band does not change position or intensity upon titration of UP1, suggesting that this band is composed of an SR/Stem III complex. This complex appears to be "locked" since UP1 is unable to disrupt the complex represented by this band, even when it is added at 1.5 μM.

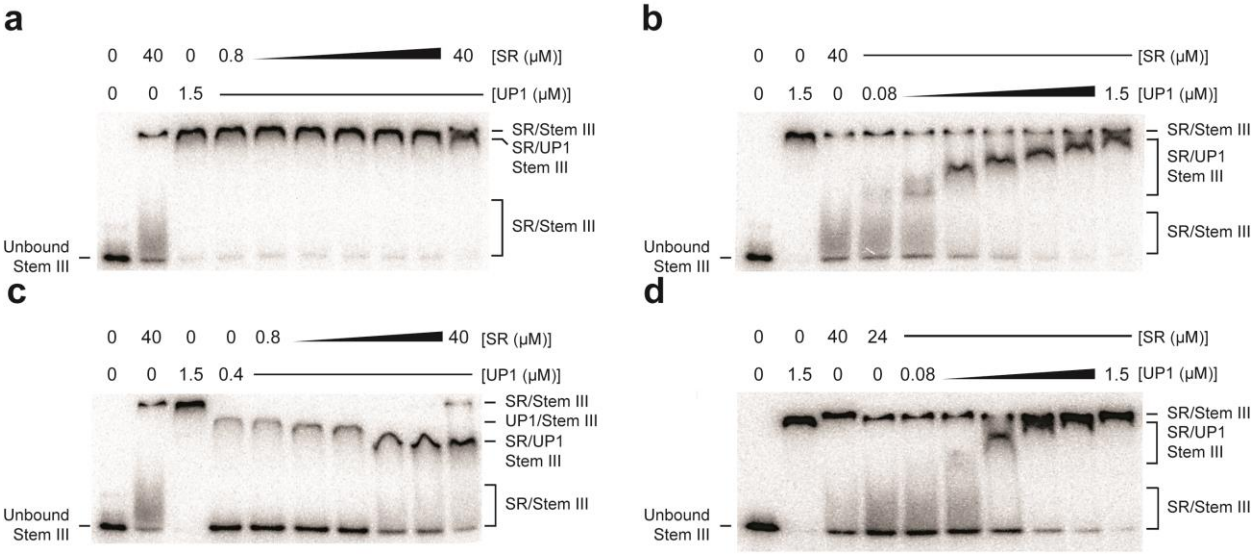

Figure 2.6. UP1 and SR bind Stem III simultaneously, yet UP1 cannot displace SR. EMSAs of competition assays with Stem III pre-bound to UP1 (a and c) or GB1-SR (b and d) and titrated with the other protein.

We next examined if UP1 and SR compete for binding to Stem III under a regime using lower concentrations to pre-bind Stem III. Using concentrations approximating the measured binding constants (Table 2.1), we pre-bound the RNA with either UP1 or SR and then titrated competitor protein up to saturating concentrations. Increasing concentrations of SR causes the UP1/Stem III complex to migrate faster through the gel. The increased mobility of the shifted band appears to occur in two distinct concentration-dependent "phases" (Figure 2.6c). At the highest concentration of SR, the distinct "locked" SR/Stem III complex forms. We cannot determine whether the "locked" complex originates from SR binding free Stem III or Stem III that was initially bound to UP1. However, Figure 2.6a suggests that a high concentration of SR displaces pre-bound UP1 to form this "locked" complex, implying both mechanisms contribute to its appearance. Similar to the saturating condition of pre-bound SR, pre-binding Stem III with lower concentrations of SR results in the formation of two sets of shifts: a smear and the "locked" complex (Figure 2.6d). Titration of UP1 condenses the smeared shift into a lower mobility band whose migration is inversely related to the UP1 concentration (Figure 2.6d).

In order to determine whether the formation of these unique complexes is a consequence of high protein concentrations within polyacrylamide gels, we generated mutants of UP1 (F17A/F108A) [253] and SR (Y44A) [223] that are defective in RNA-binding. Utilizing equimolar concentrations of mutant protein as in Figure 2.6, reciprocal competition reactions with these mutant proteins against Stem III pre-bound to the wild-type UP1 or SR demonstrate that the formation of these new complexes is dependent on both proteins having competency to bind RNA (Figure 2.7). Together, these results and the secondary structure probing suggest that SR and UP1 bind the same stem when present at lower concentrations. SR can displace saturated

UP1/Stem III complexes at sufficiently high concentrations, although UP1 cannot displace saturated SR/Stem III complexes.

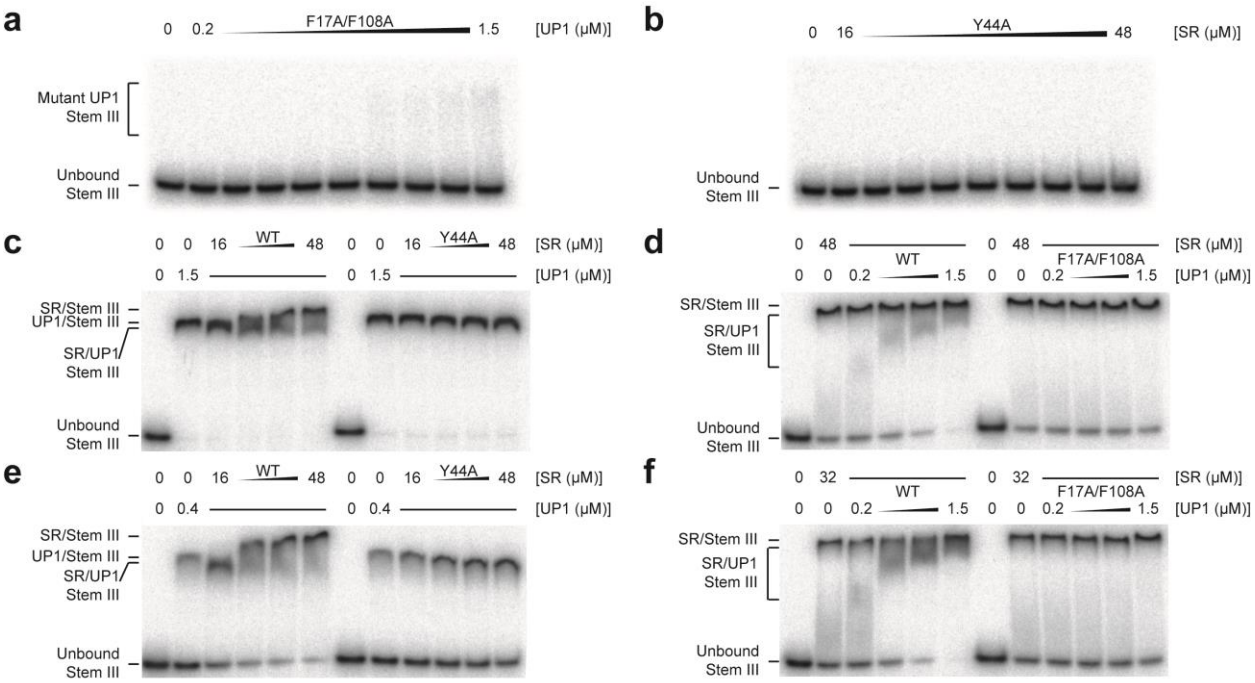

Figure 2.7. RNA-binding mutants of UP1 and SR do not form unique ternary complexes with Stem III. EMSAs of Stem III titrated with mutant UP1 (a) or GB1-SR (b). EMSAs of competition assays with Stem III pre-bound to wild-type (WT) UP1 (c and e) or GB1-SR (d and f) and titrated with either wild-type or mutant competitor protein in equimolar amounts.

2.3.7 UP1 aids in SR recruitment off Stem III

One hypothesis for the function of SRSF2 is to recruit the 7SK-P-TEFb RNP to paused RNA polymerase II via its interaction with SR-sequence rich nascent transcripts [197]. To test this hypothesis, we set up a minimal *in vitro* system that included SR, Stem III, and an RNA oligonucleotide comprising the first 20 nucleotides of the 5' UTR encoded by the human gene *PABPC1* (PABPC1 UTR). This sequence contains four putative overlapping SRSF2 binding

sites, which we used as a proxy for SR-sequence rich 5' UTRs (Figure 2.8a, top). We first assessed the interactions between this oligonucleotide and UP1 or SR. SR poorly binds the radiolabeled PABPC1 oligonucleotide (Figure 2.8a, bottom), exhibiting a smeary and low-affinity shift in an EMSA, while UP1 does not bind (Figure 2.8a, top). We then pre-bound Stem III with SR and titrated cold PABPC1 UTR into the reaction. Interestingly, neither of the two SR/Stem III complexes–smeared or "locked"–are notably displaced by the addition of PABPC1 UTR (Figure 2.8b, bottom). These results suggest that the presence of competing mRNA alone is not sufficient to recruit SR from Stem III.

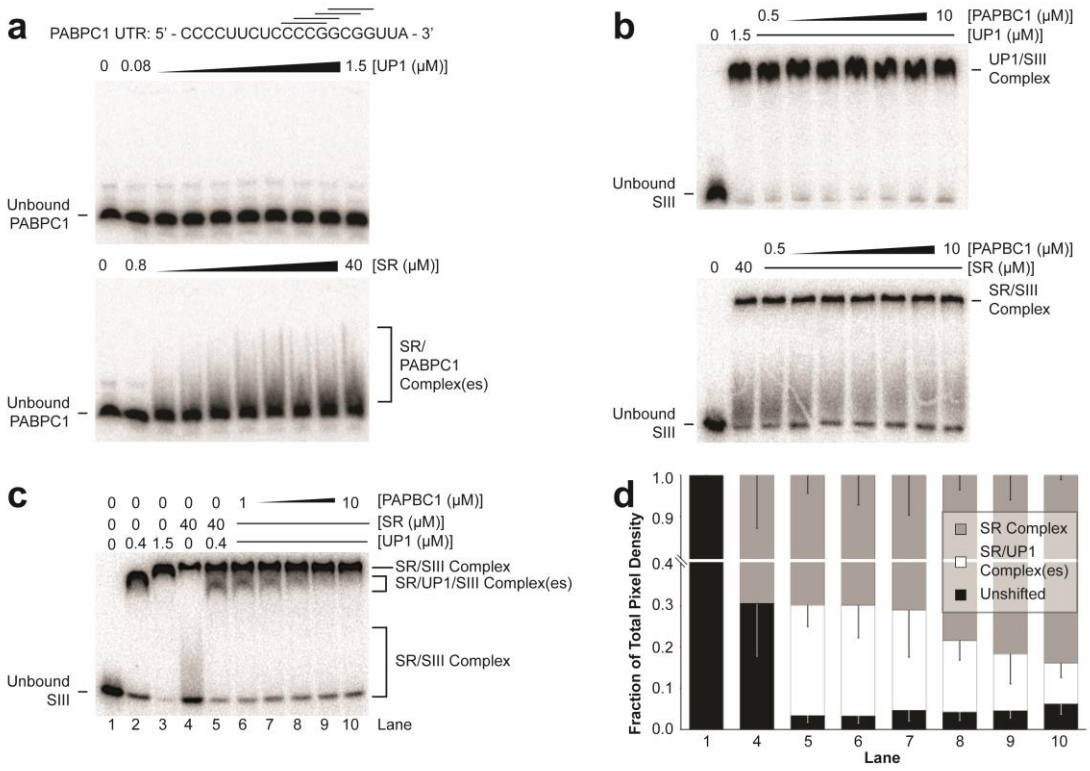

Figure 2.8. UP1 aids in efficient recruitment of SR from Stem III. (a) Top, PABPC1 UTR sequence with four putative SRSF2 binding sites indicated by horizontal lines. Middle and bottom, EMSAs of titrations of UP1 or GB1-SR with PABPC1 UTR, respectively. (b) Native gels of Stem III bound with UP1 (top) or GB1-SR (bottom) and

titrated with cold PABPC1. (c) Native gel with Stem III bound to GB1-SR, competed with UP1, and titrated with cold PABPC1. (d) Quantification of c. Values are n=3 ± S.D.

Our competition data suggest that both UP1 and SR can bind to the same stem, yet immunoprecipitation data show that these two proteins are found in mutually exclusive complexes [170, 202-204]. To explain this discrepancy, we hypothesized that the UP1/SR/Stem III complex is an intermediate complex that forms during the transition between the two 7SK RNPs, and that this transition complex may facilitate dissociation of component proteins. To test this, we performed a competition assay containing both UP1 and SR, Stem III and PABPC1 UTR. When UP1 is added to a preformed SR/Stem III complex, two species result: the "locked" SR/Stem III complex and a higher mobility SR/UP1/Stem III complex (Figure 2.8c, lane 5). Addition of PABPC1 UTR to this reaction decreases the mobility of the intermediate complex (Figure 2.8c, lanes 6-10). Quantification of unbound Stem III, the SR/UP1/Stem III complex, and the "locked" SR/Stem III complex shows that the loss of the SR/UP1/Stem III complex cannot be attributed to non-specific loss of UP1 or SR, as the unbound Stem III fraction does not significantly increase in intensity (Figure 2.8d). Rather, the intensity of the top band increases, suggesting efficient displacement of SR from the SR/UP1/Stem III complex to PABPC1 UTR. It is important to clarify that we cannot distinguish between the SR/Stem III complex or the UP1/Stem III complex, as they co-migrate within the gel (Figure 2.8c, lanes 3 and 4). However, UP1 does not bind the PABPC1 UTR oligonucleotide (Figure 2.8a, middle), and PABPC1 UTR cannot displace UP1 from Stem III (Figure 2.8b, top), implying that the difference in mobility cannot be attributed to the loss of UP1 from the SR/UP1/Stem III complex.

To ensure that the displacement of SR from the Stem III complex was not specific to the nascent mRNA sequence used, we also chose the first 20 nucleotides of the 5' UTR of *TMSB4X*,

a gene with strong SRSF2 enrichment at its promoter and known promoter proximal pausing (Figure 2.9) [197]. As with PABPC1, SR binding results in a smeared, low-affinity shift, while UP1 does not bind the oligo (Figure 2.9a). TMSB4X UTR weakly dissociates SR from Stem III (Figure 2.9b – middle). When compared to dissociation by cold Stem III (Figure 2.9b – top), however, SR displacement from the SR/Stem III complex is significantly reduced (Figure 2.9c), similar to what we observed with PABPC1 UTR (compare Figure 2.8c and Figure 2.9b – middle). Again, the potential intermediate complex containing UP1 and SR selectively disappears upon titration with TMSB4X UTR (Figure 2.9b – bottom). However, addition of UP1 into the reaction has no consequence on SR dissociation from the SR/Stem III complex (Figure 2.9c – compare ± UP1), again consistent with a "locked" complex formation. Together, these results suggest that UP1 helps to efficiently recruit SR to competing 5'-UTR RNA sequences, and this recruitment is dependent on the formation of an intermediary complex that contains both proteins.

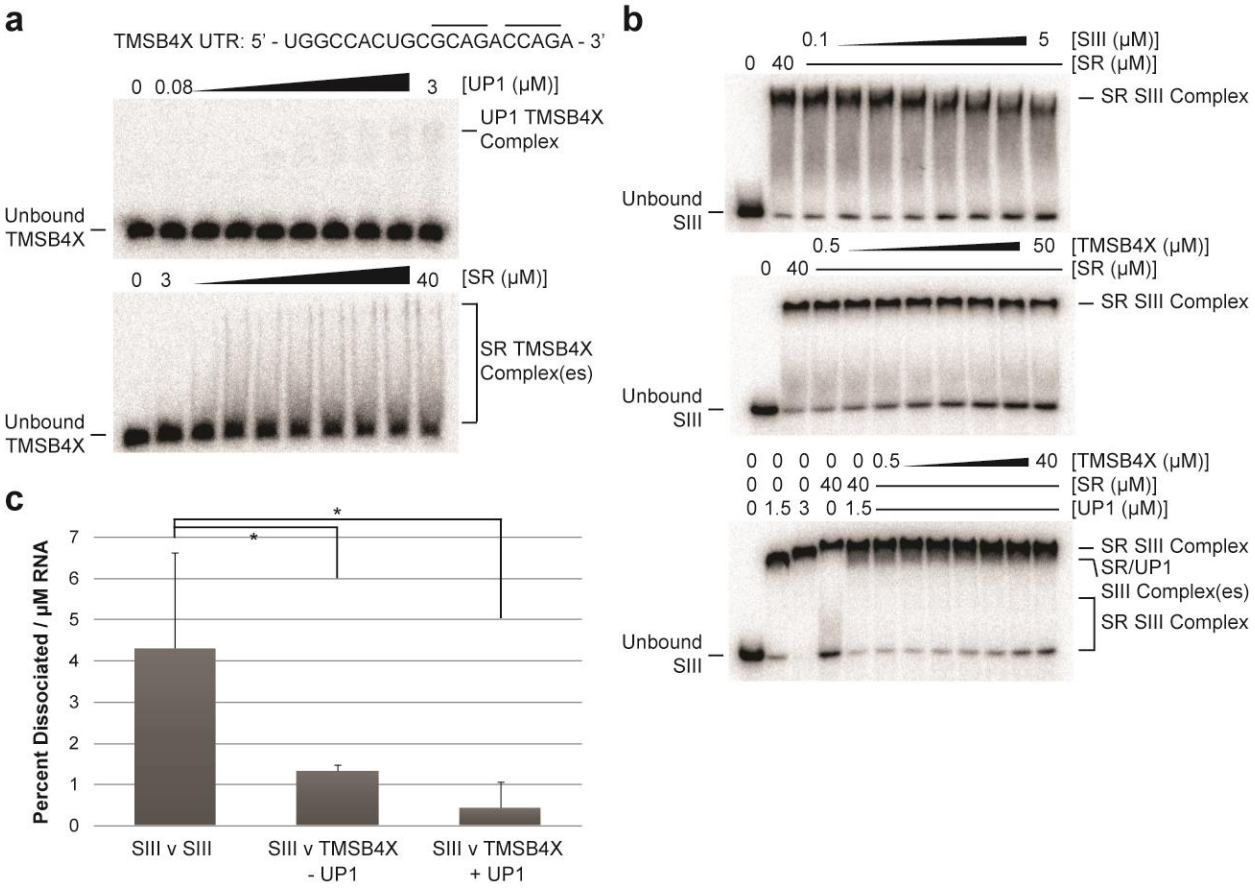

Figure 2.9. SR recruitment to alternative 5' UTR is enhanced with UP1. (a) Top, TMSB4X UTR sequence with two putative SRSF2 binding sites indicated. Middle and bottom, EMSAs of titrations of UP1 or GB1-SR with TMSB4X UTR, respectively. (b) Representative competition EMSAs with Stem III pre-bound to GB1-SR and competed with cold Stem III (top), TMSB4X (middle), or TMSB4X in the presence of UP1 (bottom). (c) Quantification of b. Percent loss of GB1-SR/Stem III complex is normalized to amount of competing RNA. All values are n=3 ± S.D. Student's t test: * $p<0.05$

2.3.8 Stem III mutations disrupt P-TEFb release

Our *in vitro* data support a model in which the competition between hnRNP A1 and SRSF2 for Stem III of 7SK RNA restructures the RNA to contribute to the formation of mutually exclusive complexes and aid in the transition between them. To understand the contribution of Stem III dynamics in the context of the entire 7SK RNA, we examined the effects of Stem III mutations in HEK 293 H cells. To try to preserve endogenous expression levels, we cloned the human 7SK gene (Genebank X05490), including endogenous promoter and terminator sequences, into pMA-T [195]. 24 hours after transfection with 7SK RNA expression plasmids, we transfected the cells again with siRNA targeting nucleotides 221-245 to minimize contributions from endogenous 7SK RNA (Figure 2.10a) [254]. RT-qPCR confirmed that 7SK RNA expression and knockdown does not significantly affect the expression of other coding (TBP mRNA) or noncoding (U2 RNA) messages transcribed by RNA polymerase II (Figure 2.10b).

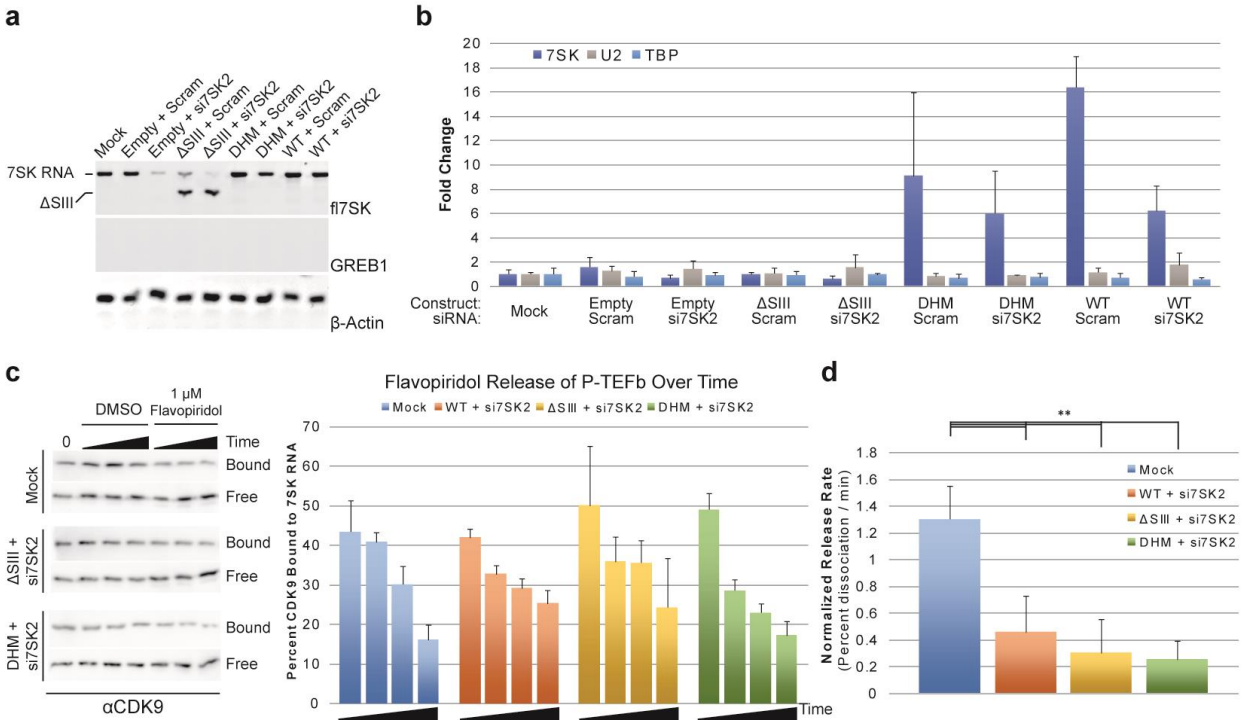

Figure 2.10. Stem III helps release P-TEFb from 7SK RNA under stress. (a) RT-PCR of full length 7SK RNA (fl7SK), growth regulation of estrogen in breast cancer 1 (GREB1) mRNA, and β-Actin mRNA across indicated transfection conditions. Mock, cells transfected with Lipofectamine 2000. Empty, empty vector control. Scram, scrambled siRNA control. ΔSIII and DHM, mutant 7SK RNA constructs. WT, wild-type 7SK RNA. (b) Fold-change of C_t values of indicated transcripts across transfection conditions as measured by RT-qPCR. All values are n=3 ± S.D. (c) Representative Western blots (left) of differential salt extractions on transfected cells subjected to DMSO or 0.1 mM flavopridiol. The levels of CDK9 were quantified in low salt (bound) and high salt (free) conditions and made relative to their total contribution (right). All values are n=3 ± S.D. (d) Rate at which CDK9 is released across time. Percent release is normalized to the release measured in DMSO control. All values are n=3 ± S.D. Student's t test: ** $p<0.01$

Because the most effective siRNA for knocking down 7SK RNA is complementary to Stem III [254-256], we chose to examine the effects of deleting Stem III (ΔSIII – nucleotides 200 to 274) and the distal helix mutant (DHM, Figure 2.3a) – a 6-nucleotide compensatory mutation that affects the binding of SR, but not UP1 (Table 2.1), and is not targeted by this siRNA – on P-TEFb release. Transfected HEK 293 H cells were incubated with DMSO or 100 μM flavopiridol for either 15, 30, or 60 minutes prior to differential salt extraction [244], which is a useful tool for separating 7SK-associated and inhibited P-TEFb (7SK-P-TEFb RNP) from free and active P-TEFb. Flavopiridol is a general cyclin-dependent kinase inhibitor with a higher preference for CDK9 [257]. Treatment with flavopiridol not only halts transcription [258], but also dissociates P-TEFb from the 7SK-P-TEFb RNP [257] and increases formation of the 7SK-hnRNP RNPs [170]. P-TEFb levels were analyzed via Western blot analysis with antibodies against CDK9. As expected, roughly half of total P-TEFb associates with 7SK RNA [204], and treatment with flavopiridol decreases the amount of CDK9 bound to 7SK RNA over time [257] (Figure 2.10c). While expression of ΔSIII and DHM does not completely inhibit the release of P-TEFb upon flavopiridol treatment, these mutations temper the magnitude of release over time compared to mock transfected cells (Figure 2.10c). These changes cannot be attributed solely to increased 7SK RNA levels, as overexpression of WT 7SK RNA results in a stunted decrease in P-TEFb release not observed in our mutants (Figure 2.10c – compare magnitude of bound CDK9 in wildtype over time to mock and mutants). Normalizing the data to the percent-release upon DMSO treatment, we plotted the amount of CDK9 released from bound 7SK RNA at each time point and calculated the rate of release (Figure 2.10d). To our surprise, the conservative 6-nucleotide DHM mutation significantly reduces the rate of P-TEFb release and displays the same decreased rate as losing Stem III altogether (Figure 2.10d). While expression of WT 7SK RNA

also results in decreased release of P-TEFb over time (Figure 2.10d), it must be noted that the triplicate results are highly dissimilar, and must be repeated. Together, these results confirm the importance of Stem III in regulating 7SK RNP homeostasis [204] and suggests protein binding and/or RNA structure of this element directly impacts P-TEFb partitioning.

2.3.9 P-TEFb dysregulation decreases promoter proximal pausing at the *HSPA1B* locus

We next wanted to see if the P-TEFb dysregulation observed in the presence of Stem III mutations affected promoter proximal pausing and transcriptional regulation of stress-induced genes at an endogenous locus. To this end, we examined the induction of HSP70 mRNA transcription from the *HSPA1B* locus upon ultra violet (UV) irradiation [78]. We used RT-qPCR with primer sets designed near the promoter and ~1500 nucleotides into the gene body (Figure 2.11a) to calculate a relative elongation ratio normalized to the elongation ratio observed across the actin (*ACTB*) gene. As a proof of principal, we first irradiated untransfected cells with UV radiation and calculated the *HSPA1B* elongation ratio over time (Figure 2.11b) [78]. We observed a consistent, but not significant, increase in transcript abundance within the gene body relative to the promoter, suggesting our assay does elicit increased transcription at the *HSPA1B* locus. We then irradiated transfected cells with UV for 30 minutes and assayed the relative elongation ratio before (UV-) and after (UV+) UV treatment (Figure 2.11c). Loss of Stem III (ΔSIII) does not significantly affect the transcription of *HSPA1B*. However, the compensatory distal helix mutation not only has a significantly higher transcription ratio prior to UV induction, but it also has a significantly higher UV-induced transcription elongation ratio (Figure 2.11c). Together with the results from the differential salt extraction experiments, these results confirm

that Stem III is important in regulating 7SK RNP dynamics, active P-TEFb levels, and transcription of a stress-induced gene at an endogenous locus.

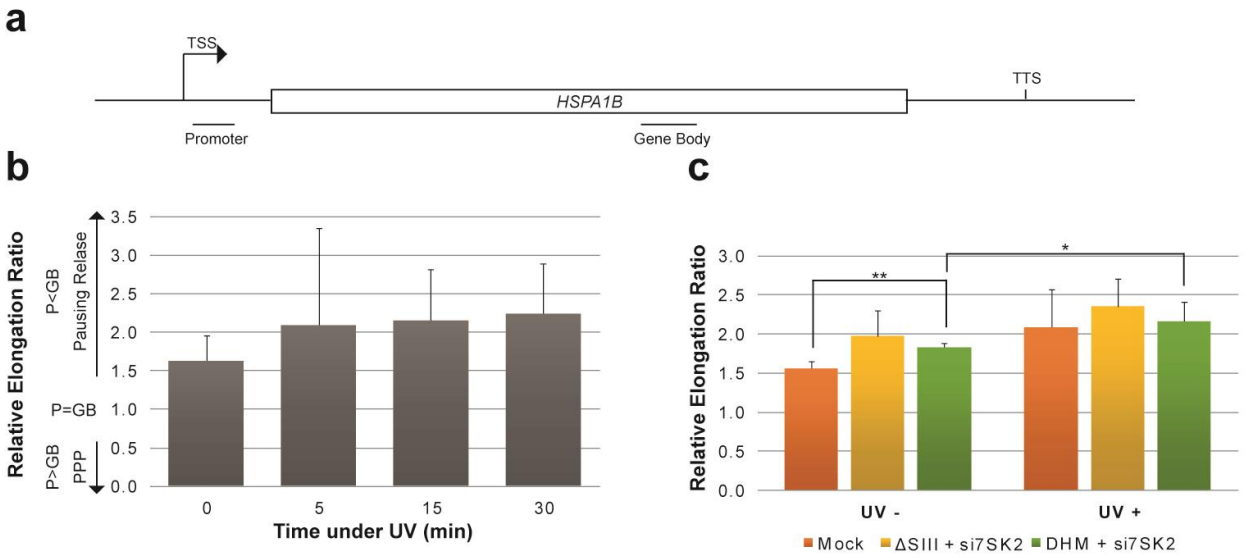

Figure 2.11. Stem III regulates promoter proximal pausing of Hsp70. (a) Schematic of human HSP70 gene *HSPA1B*. Bars below the gene represent approximate locations of PCR amplification sites for promoter (P) and gene body (GB). (b) Relative elongation ratio of RNA polymerase II following exposure to ultraviolet radiation (UV). Elongation ratios are normalized to β-actin elongation rate and TBP transcript abundance. All values are n=3 ± S.D. (c) Relative elongation rate at *HSPA1B* before (UV -) and after (UV +) 30-minute exposure to UV under various transfection conditions. All values are n=3 ± S.D. Student's t test: * $p<0.05$, ** $p<0.01$

2.4 DISCUSSION

Here, we present the first direct evidence that the RNA-binding domains of hnRNP A1 and SRSF2 bind Stem III of 7SK RNA. While UP1 binds Stem III with modest affinity, SR binds the stem with poor affinity *in vitro*; this discrepancy results from the inability of SR to melt the stem and access its single-stranded binding sites, which are located within base paired regions. We

found that UP1 and SR dramatically alter the conformation of the stem: UP1, true to its name of "unwinding protein 1" [232], melts the RNA and creates a flexible 5' side of Stem III, while SR appears to coat and collapse Stem III. We observed that SR and UP1 can bind to the same RNA molecule at low protein concentrations; however, under saturating conditions, SR assumes a "locked" conformation on Stem III that UP1 cannot displace. Indeed, SR is efficiently recruited from Stem III to competing RNA only in the presence of UP1 – and only from the complex formed at low concentrations – potentially reconstituting one aspect of the cellular role of the 7SK-hnRNP RNPs. Finally, introduction of Stem III mutants in cells leads to a disruption of 7SK RNP conversion and P-TEFb release, which leads to dysregulation of promoter proximal pausing at the *HSPA1B* locus upon UV stress. Together, our data suggest that these proteins help remodel 7SK RNA into distinct conformations, contributing to the maintenance of mutually exclusive 7SK RNPs at high concentrations. Further, we suggest that the interplay between these proteins at lower concentrations aids in the transition between the 7SK RNPs.

2.4.1 hnRNP A1 and SRSF2 directly bind and differentially restructure Stem III of 7SK RNA

The putative hnRNP A1 and SRSF2 binding sites hidden among the base pairs and bulges of Stem III are, indeed, binding sites for these proteins (Figure 2.3a). The observed binding affinities alone suggest that 7SK RNA Stem III would more readily complex with hnRNP A1 than with SRSF2 (Table 2.1), in agreement with the observation that ~75% of cellular 7SK RNA partitions into the 7SK-hnRNP RNPs in HeLa cells [170]. The mutational analysis of Stem III shows that the poor binding affinity of SR is due, in large part, to the inability of the protein to access two of its binding sites (Table 2.1). This may be an artifact of using only the single RNA

binding domain of SRSF2, since the serine-arginine-rich C-terminus of SRSF2 also binds RNA [217, 259]. Likewise, it may be an artifact of using only Stem III of 7SK RNA; while our data are consistent with published secondary structure models [164, 165], it remains a possibility that the structure of Stem III might differ in the context of the entire molecule. Nonetheless, our data suggest that for efficient binding, SRSF2 requires Stem III to adopt an open conformation, which may be provided when hnRNP A1 unloads from the 7SK-hnRNP RNP.

Interestingly, the unique rearrangements of Stem III induced by hnRNP A1 or SRSF2 binding have implications for how these proteins might promote mutually exclusive complexes *in vivo*. Secondary structure probing revealed that UP1 binds to the 3' half of the stem, causing the 5' half to become open and flexible (Figure 2.4). This flexibility might propagate along the RNA and help rearrange the secondary structure between Stems I and III to form a 7SK-hnRNP RNP. Since hnRNP A1 binds its substrates cooperatively in the 3'-5' direction [260], perhaps its unwinding activity plays a role, although other protein factors such as RNA helicase A have also been immunopreciptated with 7SK RNA [204] and may instigate a more dramatic restructuring of the RNA. Of note, although the mutational analysis of Stem III suggests a modest role for the loop-proximal UP1 binding site in recognizing hnRNP A1, RNase T1 (Figure 2.4) and in-line (data not shown) structure probing demonstrate that UP1, indeed, binds this site since the guanosines flanking the site (G249 and G255) have increased T1 sensitivity and the nucleotides opposite this site become more flexible. Accordingly, recent work with UP1 and the HIV exon splicing silencer 3 apical loop demonstrates that UP1 binding depends on both sequence and structure recognition [238].

We also made biochemical observations consistent with the hypothesis that SR condenses the stem, exposing the Watson-Crick faces of the bases opposite of its binding sites, and results

in the protection of the bulges and loop (Figure 2.4). This unexpected result is consistent with the NMR structure of SR, which shows a large basic groove on the side opposite of the binding site (Figure 2.2a) [222, 223]: the positively-charged cleft could accommodate the phosphate backbone of the opposite strand, protecting it from nuclease digestion, while exposing the bases for modification by small chemicals. This compaction could help restructure the whole 7SK RNA to bring Stems I and IV closer to one another, a feature that has been observed in the 7SK-P-TEFb RNP [178].

2.4.2 hnRNP A1 and SRSF2 competition may aid in 7SK RNP transition

SRSF2 and hnRNP A1 are antagonistic splicing proteins, with each protein possessing the ability to prevent the binding of the other on pre-mRNA transcripts [210]. Because the proteins populate different 7SK RNPs when extracted from cells, we hypothesized that competition between the proteins for binding to Stem III could help maintain separate 7SK RNPs. Under saturating conditions, each protein occludes the binding of the other, although SR can displace UP1 at concentrations in excess of 40 μM (Figure 2.6a, b). While this concentration is very high for *in vitro* studies, hnRNP A1 exists at ~200 μM in human cell nuclei [261], and SRSF2 is approximately half as concentrated as hnRNP A1 in the cell [262]. However, SRSF2 localizes entirely to the nucleus [214], and thus the local concentration of SRSF2 at any given time may indeed be higher than that of hnRNP A1. Local relative concentrations may therefore play an important role in ensuring the formation of mutually exclusive RNPs. When the concentration of SRSF2 is higher than hnRNP A1, a “locked” SRSF2-bound Stem III would prevent hnRNP A1 from opening and rearranging 7SK RNA when the RNA is bound to P-TEFb. Conversely, when the concentration of hnRNP A1 exceeds that of SRSF2, hnRNP A1 binding will prevent SRSF2-

mediated restructuring and recruitment to paused polymerases (when 7SK RNA is not loaded with P-TEFb) [197].

With lower protein concentrations, however, it appears both proteins can bind to the stem (Figure 2.6c, d). While we cannot unequivocally unravel the composition of these bands, their absence when using RNA-binding mutants (Figure 2.7) suggests two possible mechanisms for their formation: 1) UP1 and SR both bind to Stem III or 2) UP1 and SR can efficiently displace their competitor in such a way that the resulting Stem III-protein complex is structurally different, resulting in bands that migrate to different positions within the gel. If they simultaneously bind to Stem III, this may support a hand-off model in which the dissociation of one protein from Stem III facilitates binding by the other protein during the transition from one 7SK RNP to the other. This hypothesis is especially attractive, considering SR needs help to access its buried binding sites.

Observing complexes that contain both proteins on the same stem suggests 7SK RNA may form a transition complex. The idea of an intermediate Stem III complex agrees with our secondary structure probing data: UP1 binds to the 3' half of Stem III, generating an open and flexible 5' half, in which two of the three putative SRSF2 sites are located (Figure 2.4b). While it is possible that this flexible half of Stem III pairs with other nucleotides of 7SK RNA, it is tempting to speculate that the PABPC1 UTR competition experiment reveals the functional significance of this transition complex: while SR cannot be efficiently recruited from Stem III in the absence of UP1, it unloads to PABPC1 RNA from Stem III when both UP1 and SR are bound (Figure 2.8b, c). This result may explain why knock-down of both hnRNP A1 and A2 [263] or knock-down of SRSF2 [197] increases levels of promoter proximally paused RNA

polymerase II: without hnRNP A, SRSF2 cannot unload onto nascent mRNA sequences, which may hinder P-TEFb release from the complex.

2.4.3 Stem III mutations disrupt P-TEFb release and RNA polymerase II elongation in cells

Deletions of 7SK RNA (~100 nucleotides) that remove Stem III result in a loss of P-TEFb release upon transcriptional stress [204]. However, the previous study did not specifically remove Stem III – a likely stably folded element conserved across vertebrates, basal deuterostomes, and insects [165] – or examine the functional consequences of this phenomenon. We observed that deletion of Stem III of 7SK RNA in a human cell line disrupts P-TEFb release upon transcriptional stress, highlighting the importance of this element in modulating the conversion between the 7SK RNPs. Surprisingly, Stem III containing a 6-nucleotide compensatory mutation has the same stunted release of P-TEFb over time as removal of the entire stem (Figure 2.10d). Unlike the ΔSIII mutation, however, treatment with flavopiridol initially induced a substantial release of P-TEFb (Figure 2.10c, compare mock to DHM transfected cells at the first time point [15 minutes]). While the DHM mutation maintains the base-pairing proposed by Wassarman and Steitz [164], the perturbation of P-TEFb levels exhibited by this mutation is likely due to effects on protein binding affinities (Table 2.1) and disruption of alternative folds of the RNA induced by protein binding. Indeed, when examining the relative transcription elongation ratio of *HSPA1B*, only the DHM mutation significantly affected UV-induced transcription of the gene (Figure 2.11c), highlighting a potential disruption in 7SK RNA alternative folds not seen with a loss of Stem III.

Perhaps unsurprisingly, expression of WT 7SK RNA, even with co-transfection of siRNA against 7SK, resulted in 6-fold more 7SK RNA than mock transfected cells (Figure 2.10b), and this overexpression led to less P-TEFb release upon flavopiridol treatment (Figure 2.10c). In support of the importance of Stem III, however, the phenotype of this overexpression is distinctly different than expression of Stem III mutants (whether at normal levels (ΔSIII) or overexpressed (DHM)) (Figure 2.10c). When measured as a rate, then, overexpression of WT 7SK RNA leads to significantly reduced P-TEFb release compared to mock transfected cells. Therefore, the results obtained in Figure 2.10 with our mutant 7SK RNAs must be taken with a grain of salt as we are unable to definitely distinguish between contributions of overexpression versus mutant sequences.

While Stem III is not believed to directly contact P-TEFb, our *in vitro* and in-cell data can be integrated into a model in which these naturally antagonistic splicing factors help restructure 7SK RNA into distinct, alternatively-folded RNPs that indeed directly impact active and inactive P-TEFb levels (Figure 2.12). We suggest that the 7SK-hnRNP RNPs are not simply a molecular sink for hnRNPs after P-TEFb dissociation, but important regulators of transcription. More than solely modulating general RNA polymerase II transcription through active and inactive P-TEFb levels, the flux between the 7SK RNPs would also modulate the active concentrations of splicing, mRNA transport, and translation factors through binding and releasing specific hnRNPs. Future studies to address the regulation of the complicated interplay between the two types of 7SK RNPs are necessary to understand how this non-coding RNA determines the fate of metazoan transcription.

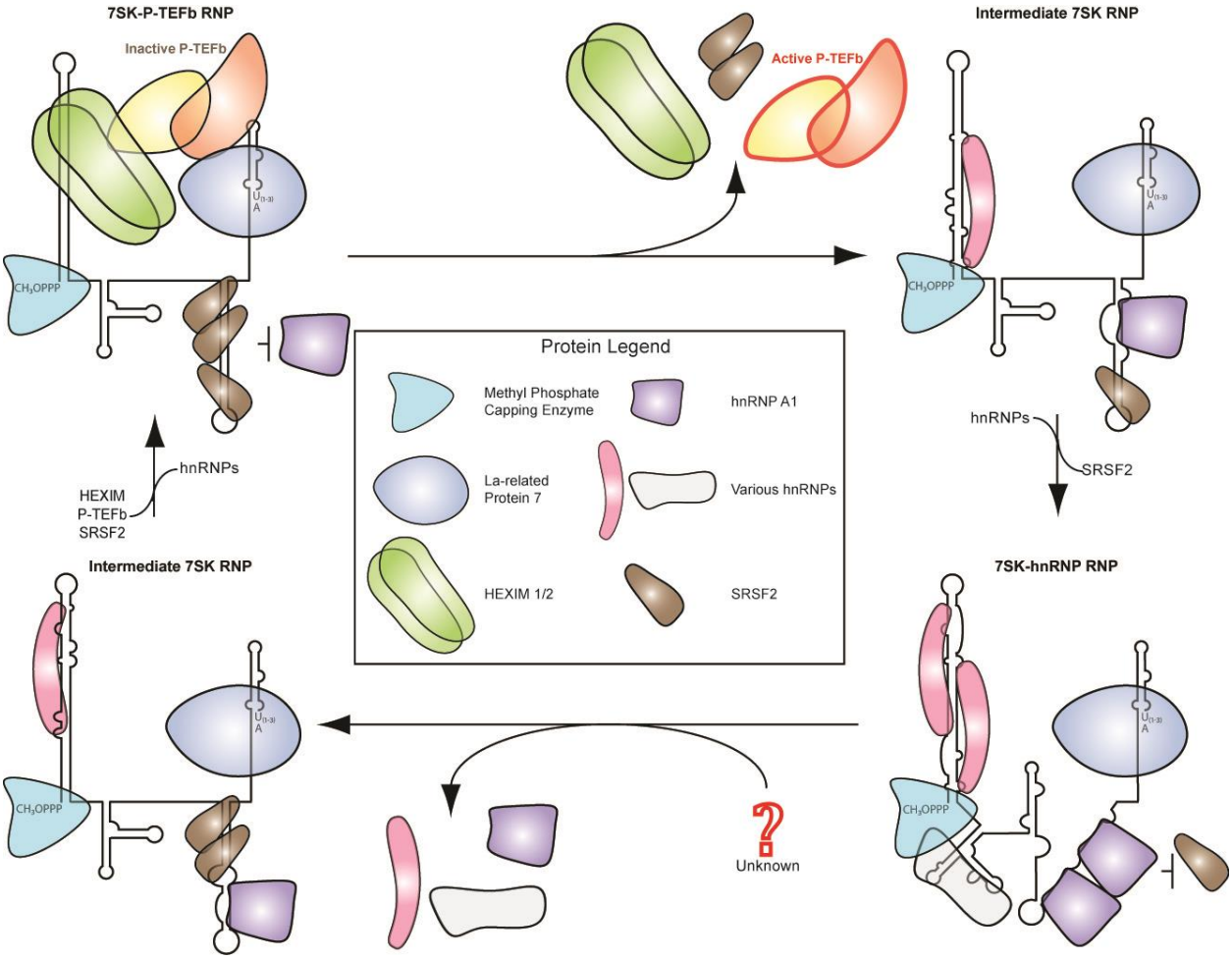

Figure 2.12. Proposed model for the contribution of hnRNP A1 and SRSF2 in the transition between the 7SK RNPs. In the 7SK-P-TEFb RNP (top left), SRSF2 coats Stem III to prevent the binding of hnRNP A1. SRSF2 unloads from the 7SK-P-TEFb RNP onto 5' UTRs in the presence of hnRNP A1, and hnRNP A1 dissociates the remaining SRSF2 proteins (top right). Bound hnRNP A1 sterically blocks SRSF2 from reorganizing the 7SK-hnRNP RNP (bottom right). hnRNP proteins are evicted from the 7SK-hnRNP RNP through an unknown mechanism and SRSF2 binds and remodels Stem III (bottom left) to reestablish the 7SK-P-TEFb RNP.

3.0 P-TEFB-DEPENDENT PHOSPHORYLATION OF HNRNP K ISOFORM 3 MODULATES BINDING TO THE SNRNA 7SK RNP

3.1 INTRODUCTION

The transition between the 7SK RNPs can be considered to behave according to Le Chatelier's principle, whereby knock-down of protein components [202, 203] or drug-mediated inhibition of transcription [170, 202, 204] lead to the increased formation of the alternate RNP. In this regard, the RNA may be thought of as being in equilibrium between the 7SK-P-TEFb RNP and the 7SK-hnRNP RNPs. This, in turn, argues that the formation of these complexes directly regulates active levels of P-TEFb, and therefore promoter proximal pausing. It has been demonstrated that the minimal functional 7SK RNA can be constructed solely with Stems I and IV separated via a short linker [178], suggesting that P-TEFb control is heavily influenced through the ability of HEXIM to bind Stem I and sequester P-TEFb. We have found that hnRNP K binds Stem I; protein-protein interaction networks suggest that hnRNP K and the catalytic subunit of P-TEFb interact [173], leading us to hypothesize and demonstrate that hnRNP K is a novel substrate for P-TEFb. In this chapter I examine how the binding of hnRNP K isoform 3 to Stem I of 7SK RNA is modulated through P-TEFb-dependent phosphorylation at S261, and how this mark may help establish proper termination of RNA polymerase II *in vivo*.

3.1.1 hnRNP K structure and function

hnRNP K, a ubiquitous and highly expressed [262] scaffolding protein, helps bridge nucleic acids and proteins to mediate a myriad of RNA processing events. Beyond the traditional hnRNP roles of mediating alternative splicing [264-266], hnRNP K also regulates transcription initiation [267], anchors chromatin to the nuclear matrix [268], aids in chromatin remodeling events [264, 269], modulates translation [270], and ameliorates DNA-damage-induced apoptosis [271, 272]. In addition, extensive post-translational modification of hnRNP K has been linked to transducing signals to RNA- or DNA-related processes [273]. Indeed, hnRNP K has such diverse and important cellular roles that complete knock-out is lethal [274-276] and substantial knock-down induces cellular apoptosis [277-279]. hnRNP K is upregulated in a number of cancers, and there is a strong correlation between hnRNP K expression and localization and cancer prognosis [280].

hnRNP K has a modular architecture with specific, non-overlapping domains to mediate both nucleic acid interactions and protein-protein interactions (Figure 3.1a). hnRNP K binds both single-stranded [281] and double-stranded DNA [267, 282], as well as single-stranded RNA [283], through three hnRNP K homology (KH) domains (KH1, KH2, and KH3). These domains have a strong preference for poly-pyrimidine tracts [284], placing hnRNP K along with hnRNP E1 and E2, αCP3, and αCP4 into the poly(C)-binding protein family [285]. Multiple biochemical and genetic-interaction studies have shown that the KH domains of hnRNP K work together to increase nucleic acid binding specificity and affinity [286]. In addition to the defined KH domains, hnRNP K also contains several RGG/RG boxes positioned between KH2 and KH3 (Figure 3.1a). While the RGG/RG boxes have not been shown in isolation to bind nucleic acids [287], RGG/RG boxes are the second-most abundant RNA-binding domain [288] and have been shown to aid globular RNA-binding domain interactions [289].

hnRNP K also contains an intrinsically disordered region known as the hnRNP K interacting (KI) domain located between KH2 and KH3 (Figure 3.1a) [273]. This domain likely aids in mediating the over 200 putative protein-binding interactions that have been identified [290]. Interspersed among the RGG/RG boxes within the KI domain are several poly-proline tracts that are classically defined as Src-homology 3 binding domains [291]. The KI domain is also host to a large majority of the post-translational modifications, highlighting how signal transduction pathways not only use hnRNP K as a docking platform, but also modify the protein to help mediate its myriad functions [273]. Immediately adjacent to the KI domain is the non-canonical hnRNP K nuclear shuttling domain (KNS) [292], whose phosphorylation status dictates cytoplasmic accumulation [293]. Together, the individual modules of hnRNP K work within a complex interaction network to influence nearly every major cellular process.

3.1.2 hnRNP K association with 7SK RNA and P-TEFb

hnRNP K not only associates with and mediates messenger RNA-related processes, but it also has roles involving non-coding RNA interactions and functions [294, 295]. RNA-based affinity purification of 7SK RNA followed by mass spectrometry revealed that hnRNP K interacts with 7SK RNA [203]. Co-immunoprecipitations of hnRNP K and 7SK RNA revealed that, like all hnRNP associations with 7SK RNA, hnRNP K does not interact with HEXIM or P-TEFb components, and therefore resides in one of the 7SK-hnRNP RNPs. Interestingly, siRNA-mediated knock-down of hnRNP K led to an increase in the formation of the 7SK-P-TEFb RNP, suggesting that hnRNP K plays a role in maintaining the balance between the two RNPs [203]. While Hogg and Collins found that P-TEFb and hnRNP K do not reside in the same RNP, a high-throughput protein-protein interaction study revealed that hnRNP K and CDK9, but not

cyclin T1, interact [173]. Because CDK2 has been shown to phosphorylate hnRNP K [296], and phosphorylation of hnRNP K modulates the protein's binding and localization [273], we hypothesized that hnRNP K is phosphorylated by P-TEFb. P-TEFb-dependent phosphorylation would induce a change in association between hnRNP K and 7SK RNA, thereby transitioning the 7SK-hnRNP RNP to the 7SK-P-TEFb RNP.

3.2 MATERIALS AND METHODS

3.2.1 RNA construction design

The transcription template and transcription reaction for Stem III was generated as in Section 2.2.1. Stem IV transcription template was generated using 5' Gen For and AAAbottom Rev (Appendix A.2) to PCR amplify nucleotides 296-331 from the human 7SK DNA sequence synthesized into pIDTSMART (IDT). The hepatitis delta virus 3' cleavage overlap was added with PCR to the DNA sequence encoding the hepatitis delta virus ribozyme [239]. Stem I construct was generated using nested PCR (Appendix A.2) to amplify nucleotides 1-108 from the human 7SK DNA sequence with a T7 promoter upstream. Stem IV RNA was transcribed *in vitro* at 37°C for 4 hours with final concentrations of 40 mM Tris-HCl, pH 8, 10 mM DTT, 5 mM spermidine, pH 8, 17 mM $MgCl_2$, 4 mM each NTP, and recombinant T7 RNA polymerase that was purified in-house [241]. Stem I RNA was transcribed *in vitro* at 37°C for 4 hours with final concentrations of 40 mM Tris-HCl pH 8, 10 mM DTT, 5 mM spermidine, pH 8, 15 mM $MgCl_2$, 4 mM each NTP, 0.01% Triton-X, and recombinant T7 RNA polymerase purified in house [241].

RNAs were gel-purified on 8% 29:1 polyacrylamide 1X TBE/7M urea gels, excised, eluted, concentrated, and stored in 10 mM Tris-HCl, pH 8 and 1 mM EDTA, pH 8.

3.2.2 Protein purification

Full-length hnRNP K isoform 2 was cloned from HsCD00520582 (DNASU) using primers hnRNP K For and hnRNP K Rev (Appendix A.2) and placed into pHMG6 using NheI and EcoRI (Thermo Scientific) to generate pHK. Wildtype hnRNP J was generated through site-directed mutagenesis of pHK using primers in Appendix A.2 to generate pHJ. Mutant hnRNP J constructs were generated through site-directed mutagenesis of pHJ (Appendix A.2) [242]. The plasmid for human-expression of hnRNP J was generated by PCR amplifying pHJ template using primers MycJ For and MycJ Rev and placed into pCMV6 using AsiSI and NotI (Thermo Scientific). HEXIM construct was a generous gift from David Price (University of Iowa). GST was expressed from pGEX-6P-1 (GE Healthcare). GST-yCTD construct was a generous gift from Craig Kaplan (Texas A&M University). All cloning was confirmed by sequencing (Genewiz). hnRNP K and J expression constructs were transformed into BL21 (DE3) RIPL *Escherichia coli* and expressed in autoinduction medium at 18°C for 20 hours. HEXIM expression construct was transformed into BL21 (DE3) RIPL *Escherichia coli* and expressed in LB medium with 0.1 mM IPTG at 18°C for 20 hours. GST and GST-yCTD expression constructs were transformed into BL21 *Escherichia coli* and expressed in LB medium with 1 mM IPTG at 18°C for 20 hours.

Wildtype and mutant hnRNP K and J-expressing cells were lysed in lysis buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol, 5% glycerol) and purified in batch method using HisPur Ni-NTA Resin (Thermo Scientific). Protein was eluted with lysis buffer + 500 mM imidazole, and dialyzed overnight into TEV cleavage buffer (20 mM

Tris-HCl, pH 8, 150 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol, 5% glycerol) in the presence of TEV protease purified in house [297]. After a 20-minute centrifugation at >24,000 x g, supernatant was immediately passed over a HisTrap HP column (GE Healthcare) equilibrated in TEV cleavage buffer. Nickel column flow-through was then applied to HiTrap HP Q anion and HiTrap HP SP cation exchange columns (GE Healthcare) equilibrated in Q/S starting buffer (20 mM Tris-HCl, pH 8, 150 mM NaCl, 1 mM β-mercaptoethanol, 5% glycerol) and eluted in Q/S elution buffer (20 mM Tris-HCl, pH 8, 1 M NaCl, 1 mM β-mercaptoethanol, 5% glycerol). Fractions containing hnRNP K or J were confirmed via SDS-PAGE and concentrated and stored at -80°C.

HEXIM-expressing cells were lysed in 20 mM Tris-HCl, pH 8, 20 mM imidazole, 500 mM NaCl, 1 mM β-mercaptoethanol, 5% glycerol and purified in batch using nickel agarose (Thermo Scientific). Protein was eluted by increasing the imidazole concentration to 300 mM. Eluted protein was dialyzed overnight into lysis buffer and applied to a HisTrap HP column (GE Healthcare). Fractions containing HEXIM protein were concentrated to 60 µM to avoid precipitation observed at higher concentrations (personal observation) and stored at -80°C.

GST and GST-yCTD were purified in a similar fashion. Induced cells were lysed in lysis buffer (25 mM Tris-HCl, pH 8, 350 mM NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol, 10% glycerol) and purified in batch using glutathione agarose resin (GoldBio). Protein was eluted upon addition of lysis buffer supplemented with 10 mM reduced glutathione and dialyzed overnight into 25 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol, and 10% glycerol buffer. After centrifugation to remove debris, the supernatant was applied to HiTrap HP Q anion and HiTrap HP SP cation exchange columns (GE Healthcare). Fractions containing either GST or GST-yCTD were concentrated and stored at -80°C.

3.2.3 Electrophoretic mobility shift assays

RNA was 5' end-labeled with γ-^{32}P-ATP and T4 PNK (Thermo Scientific), gel purified in 8% 29:1 polyacrylamide 1X TBE/7M urea gels, excised, eluted and stored in 10 mM Na cacodylate, pH 6.5. RNA was snap-cooled (95°C for 2 minutes followed by 15 minutes at 4°C) in 100 mM NaCl and 10 mM Na cacodylate, pH 6.5. Protein was titrated into binding reactions containing 1,000 counts per minute (cpm; <1 nM) annealed RNA, hnRNP K binding buffer (16 mM Tris-HCl, pH 8, 20 mM NaCl, 11.2 mM β-mercaptoethanol, 4% glycerol), 2.5 µg yeast total tRNA (Thermo Scientific), and 0.1 µg BSA (NEB), and incubated on ice for 35 minutes. For competition reactions, the first protein was titrated and incubated with RNA for 25 minutes followed by the addition of the competitor protein and incubation for 15 minutes at 4°C. Reactions were analyzed on 5% Tris/glycine native 29:1 polyacrylamide gels, dried, and exposed to an image plate overnight. Plates were scanned on a Fuji FLA-5000 and quantitated with MultiGauge software (Fujifilm).

3.2.4 Kinase assays

Except for GST-yCTD (diluted to a final concentration of 1 µM), proteins were diluted to 20 µM in a total volume of 7.5 µL prior to addition of 5 µL of 5x reaction buffer (40 mM MOPS-NaOH, pH 7, 1 mM EDTA, 5 mM β-glycerophosphate) and 2.5 µL of P-TEFb (Millipore) diluted in dilution buffer (20 mM MOPS-NaOH, pH 7, 1 mM EDTA, 0.01% Tween-20, 5% glycerol, 0.1% β-mercaptoethanol, 1 mg/mL BSA) to 20 ng/µL. No-kinase reactions contained 2.5 µL of dilution buffer. Reactions containing kinase inhibitors received 1 µL of 1 mM flavopiridol or DRB diluted in DMSO. All reactions were initiated upon addition of 10 µL γ ATP Mix (25 mM

MgAc, 0.3 mM ATP, 5-25 nCi/µL ^{32}P-γ-ATP (Perkin Elmer)) and set at 30°C for 20 minutes prior to addition of 15 µL of acid stop mix (3% phosphoric acid, 20 mM Tris-HCl, pH 6.8, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol). 9 µl of each reaction was run on 10% SDS 19:1 polyacrylamide gels, dried, and exposed to an image plate overnight. Plates were scanned on a Fuji FLA-5000 and quantitated with ImageJ [245].

3.2.5 Cell culture and transfection

HeLa cells were cultured under standard conditions with DMEM (Thermo Scientific) supplemented with 10% FBS. For transfections, low density-plated cells in 24-well plates were transfected with 1 µg of plasmid or lipofectamine 2000 (Invitrogen) alone (mock transfection).

3.2.6 Nuclear fractionation

HeLa cells were grown to ~90% confluency in 6 cm plates. 50 µL of either DMSO or 0.1 mM flavopiridol diluted in DMSO were added directly to the media and incubated at 37°C for 30 minutes. Release reactions were treated for 30 minutes as described prior to removal of media and incubation at 37°C for 15 minutes with fresh media. Plates were washed 2x with 2 mL of 1X PBS (Fischer Scientific), manually scraped into 1 mL 1X PBS, and transferred to a microcentrifuge tube. 50 µL of resuspended cells plus 10 µL 4X SDS loading dye was used to generate whole cell extracts. Nuclear fractionation was performed using a nuclear extraction kit according to protocol (Abcam) with one exception: nuclear pellet was washed 1x with 300 µL pre-extraction buffer (generated as per kit instruction) and transferred to new tube prior to addition of extraction buffer. Protein concentrations were measured using a Bradford assay

(Thermo Scientific). Equal concentrations of protein were loaded on 10% SDS polyacrylamide gels. Gels were transferred to Protran BA 85 nitrocellulose membranes (GE Healthcare), blocked with 3% milk in TBST for 1 hour at room temperature, and incubated with either 1:5000 α-YY1 antibody (Abcam, EPR4652) or 1:5000 α-α-tubulin (Thermo Scientific, DM1A) diluted in 3% milk overnight at 4°C. Blots were washed 3x for 10 minutes with TBST and incubated with 1:5000 α-mouse-HRP antibody (Santa Cruz, sc-516102) (α-α-tubulin) or 1:5000 α-rabbit-HRP antibody (Santa Cruz, sc-2357) diluted in 1% milk for 1 hour at 4°C. After washing 3x for 10 minutes with TBST, blots were imaged on an Amersham Imager 600 with Pierce ECL Western Blotting Substrate (Thermo Scientific) and quantified with Image J [245]. Blots were stripped with mild stripping buffer according to protocol (Abcam), blocked with 3% milk in TBST for 1 hour at room temperature, and incubated with 1:5000 α-hnRNP K (Santa Cruz, sc-28380) diluted in 3% milk overnight at 4°C. Washes, secondary antibody (1:5000 α-mouse-HRP), imagining, and analysis were performed as above.

3.2.7 Cellular RNA purification and reverse transcription

24-hours after transfection, media was removed, and 0.25 mL of Trizol (Invitrogen) was used for RNA extraction as per manufacturer's protocol. RNA pellets were resuspended in 25 µL water. Protein was extracted from organic phase as per protocol and subjected to Western blot analysis. 250 ng of total RNA was subjected to reverse transcription using the QuantiTect kit (Qiagen) as per manufacturer's protocol.

3.2.8 RT-PCR and RT-qPCR

RT-qPCR reactions containing 1 μL of 0.75X cDNA, 5 μL iTaq SYBR master mix (Bio-Rad), and 500 nM forward and 500 nM reverse primers were subjected to 40 cycles of PCR on an Applied Biosystems StepOne Plus Real-time Thermocycler. C_t values were calculated with StepOne Software v2.2.2.

3.3 RESULTS

3.3.1 hnRNP K binds Stems I and III of 7SK RNA

Formaldehyde crosslinking followed by immunoprecipitation of hnRNP K in human cells shows a strong enrichment for 7SK RNA via northern blot analysis, suggesting a direct interaction between the two *in vivo* [203], although the study did not identify where the protein binds 7SK RNA. Using a thermodynamic scoring matrix developed from a serial analysis of gene expression assay and yeast three-hybrid screens [298], we identified three high-probability hnRNP K binding sites: two located within Stem I and one along the 5' half of Stem III (Figure 3.1b). Quantification of EMSAs performed with purified recombinant hnRNP K and Stem I RNA (nucleotides 1-108) reveals modest binding with a K_d of 5.3 ± 1.7 μM (Figure 3.1c top left). However, RNA tagged-based affinity purification of 7SK RNA displays a 4-fold stronger enrichment for hnRNP K isoform 3 than for hnRNP K isoforms 1 or 2 [203]. Therefore, we cloned, expressed, and purified recombinant human hnRNP K isoform 3 (hereafter referred to as hnRNP J, the historical designation for this alternatively spliced isoform of hnRNP K [299]

(Figure 3.1a)) and performed EMSAs with Stem I. hnRNP J has a significantly higher affinity for Stem I than hnRNP K, with a K_d of 2.8 ± 0.03 μM (student's T-test p=0.04) (Figure 3.1c bottom left). We also tested hnRNP J binding to Stems III (nucleotides 200-274) and IV (nucleotides 295-331) via EMSAs. hnRNP J has a lower affinity for Stem III (K_d = 3.8 ± 0.3 μM) than for Stem I (Figure 3.1c top right), in agreement with the scoring matrix (Figure 3.1b). We observed no appreciable binding to Stem IV, demonstrating a specific interaction to Stems I and III (Figure 3.1c bottom right). Together, our data reveal a direct interaction between 7SK RNA and hnRNP K isoforms 2 and 3, with the strongest preference for the alternatively spliced isoform 3 binding to Stem I.

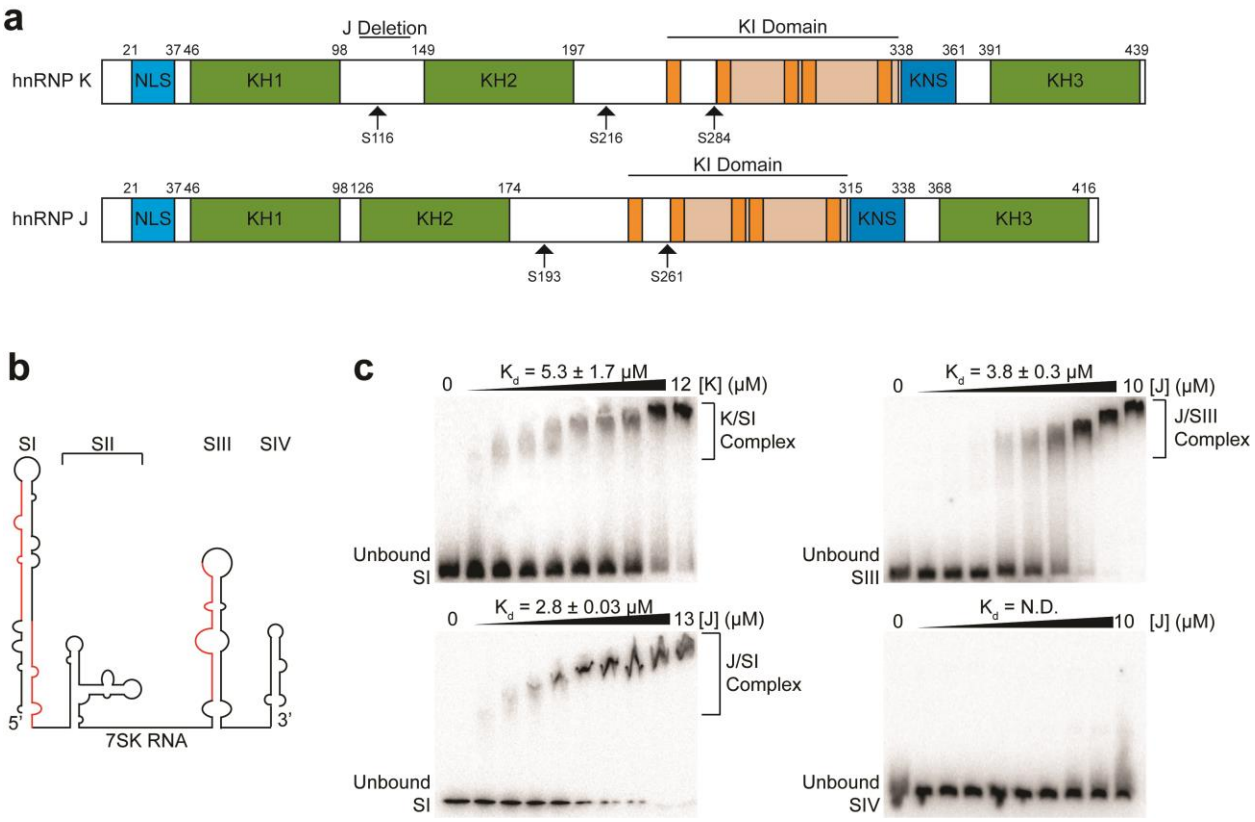

Figure 3.1. hnRNP K isoforms 2 and 3 bind Stems I and III of 7SK RNA. (a) Bar diagrams of hnRNP K isoforms 1 and 2 (top) and isoform 3 / hnRNP J (bottom) [285, 300] with domain boundaries above. NLS, nuclear localization signal; KH, hnRNP K homology domain; KI, hnRNP K interacting domain; KNS, hnRNP K nuclear shuttling domain. Orange bars represent RGG/RG motifs and tan bars indicate Src-homology 3 binding regions [273]. Arrows point to potential cyclin-dependent kinase phosphorylation sites [301]. (b) Cartoon of predicted secondary structure of 7SK RNA [164] with stem loops I-IV labeled above (SI-SIV). Red regions of RNA, potential hnRNP K binding sites [298]. (c) Representative EMSAs of titrations of hnRNP K (K) or hnRNP J (J) with Stems I, III, and IV. Average K_d of three (J) or six (K) replicates ± standard deviation is shown above each gel. N.D., not determined.

3.3.2 P-TEFb phosphorylates hnRNP J at S261

Because hnRNP K is a major docking protein with over 200 potential protein-protein interaction partners [290], binds 7SK RNA (Figure 3.1 and [203]), and interacts with the catalytic subunit of P-TEFb [173], we next investigated whether P-TEFb could directly phosphorylate hnRNP J. Using group-based prediction software [301] of potential phosphorylation sites by cyclin-dependent kinases 2, 5, and 7, we identified two high-probability phosphorylation sites at S193 and S261 on hnRNP J (Figure 3.1a). We then performed standard kinase assays according to manufacturer's protocol (Millipore, #14-685) using commercially-available P-TEFb and either wild-type (WT), S193A, S261A, or S193/261AA hnRNP J (Figure 3.2a and b). We used glutathione S-transferase (GST) as a negative control and GST fused to the *Saccharomyces cerevisiae* C-terminal domain of Rpb1 (GST-yCTD) [302], a known substrate of Bur1/2, the yeast ortholog of P-TEFb [303, 304], as a positive control. To confirm that our results are specific to phosphorylation by P-TEFb, we also performed the reactions in the presence of the general CDK inhibitor flavopiridol [257] or the P-TEFb-specific inhibitor DRB [115].

While incubation of GST with P-TEFb yielded minimal incorporation of radioactive phosphate (Figure 3.2a and b), we observed a dramatic increase in signal for GST-yCTD and WT hnRNP J upon P-TEFb addition, suggesting efficient phosphorylation of these substrates by the kinase. Addition of either 40 μM of flavopiridol or DRB (competed with 100 μM of cold ATP in the reactions) results in drastically diminished signal, confirming that phosphorylation is specific to P-TEFb. Surprisingly, S261A displays a significant decrease in ^{32}P incorporation compared to WT (Figure 3.2c) (student's T-test p=0.03). A double alanine mutation at S193 and S261 displays the same level in reduction of radiation incorporation as S261A (Figure 3.2b). Pre-

binding hnRNP J to Stem I of 7SK RNA does not increase or decrease the signal (data not shown), indicating that P-TEFb-dependent phosphorylation of hnRNP J is independent of nucleic acid interaction. While S261A does not completely abolish ^{32}P incorporation, similar results obtained through *in vitro* kinase assays and P-TEFb [305] and our data together suggest that P-TEFb directly phosphorylates hnRNP J predominantly at S261.

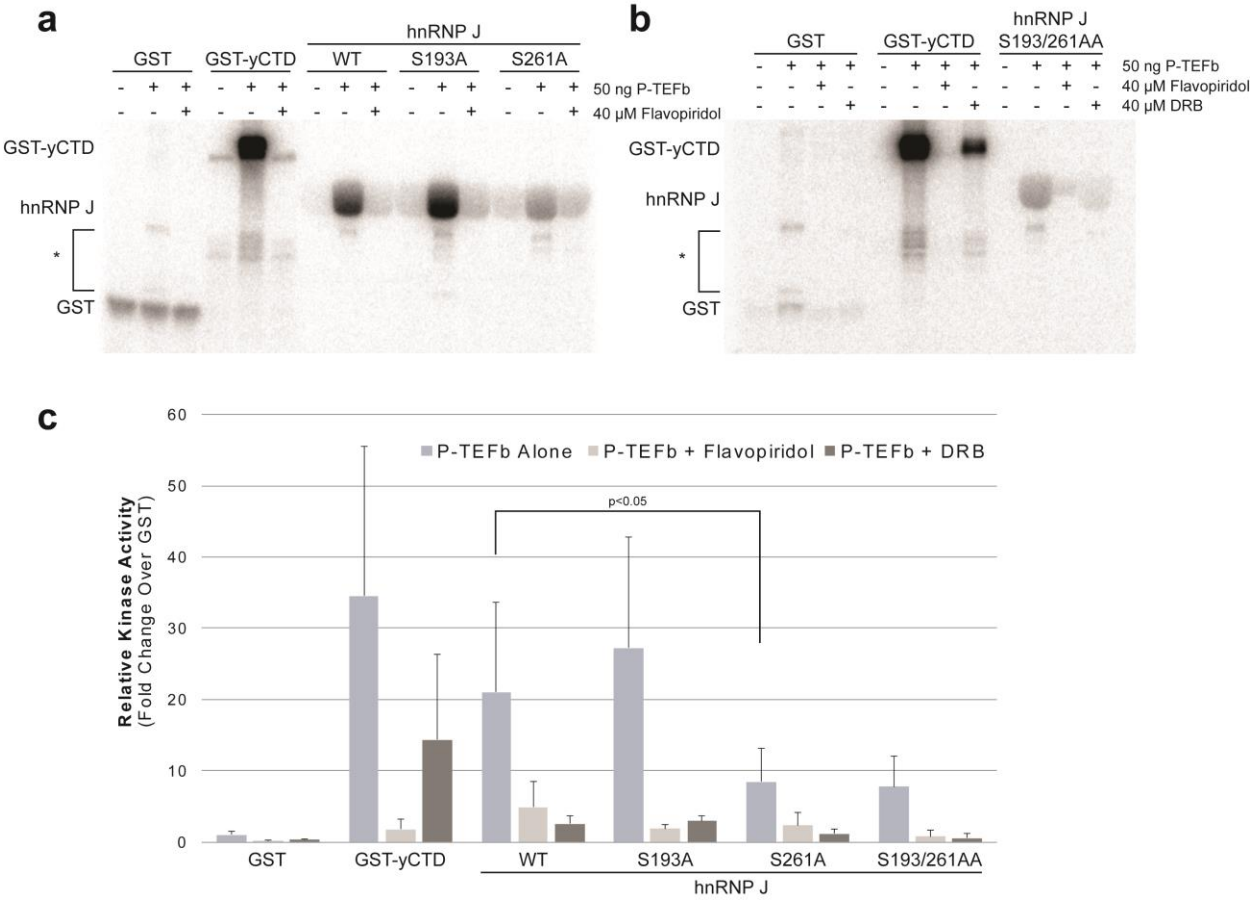

Figure 3.2. P-TEFb directly phosphorylates hnRNP J. (a and b) Representative SDS polyacrylamide gels of substrates phosphorylated with 50 ng total protein of commercial P-TEFb (Millipore) in the presence of trace ^{32}P-γ-ATP. GST, glutathione S-transferase; GST-yCTD, GST fused to the final 26 hepta-peptide repeat from *Saccharomyces cerevisiae* Rpb1 [302]. "*" indicates non-specific and degradation products. (c) Quantification of kinase assays. Background-subtracted (P-TEFb "-" from P-TEFb "+") total pixel density is normalized against pixel density from GST + P-TEFb. Values are averages from three independent replicates ± standard deviation.

3.3.3 Phosphorylation of S261 decreases affinity for 7SK RNA

S261 is located within the KI domain of hnRNP J (Figure 3.1a), far from the canonical RNA-binding domains (KH domains). Therefore, in order to elucidate a potential function for P-TEFb phosphorylation of hnRNP J at S261, we closely examined the local environment of S261. To our surprise, S261 is located directly adjacent to a stretch of RGG/RG boxes interspersed with two poly-proline Src-homology 3 binding motifs (Figure 3.1a) [273]. As RGG/RG motifs are the second most common RNA-binding motif in eukaryotes [288], we decided to examine if phosphorylation at S261 might modulate hnRNP J binding to 7SK RNA. To this end, we generated S261A (phospho-null) and S261D (phospho-mimetic) constructs of hnRNP J and subjected them to EMSA analysis with Stem I RNA (Figure 3.3). Unexpectedly, both S261A (3.6 ± 0.4 μM) and S261D (4.0 ± 0.4 μM) have significantly less affinity for Stem I than WT hnRNP J (2.8 ± 0.03 μM) ($p=0.01$ and $p=0.006$, respectively) (Figure 3.3b). However, S261D has marginally less affinity for Stem I RNA than S261A does (Figure 3.3a), although neither construct binds as poorly as hnRNP K (Figure 3.3b). These results suggest that the RGG/RG motifs around, and including, S261 mediate a portion of the interaction with Stem I of 7SK RNA; however, more experiments are necessary to tease out the exact contribution of the RGG/RG boxes towards association with 7SK RNA elements.

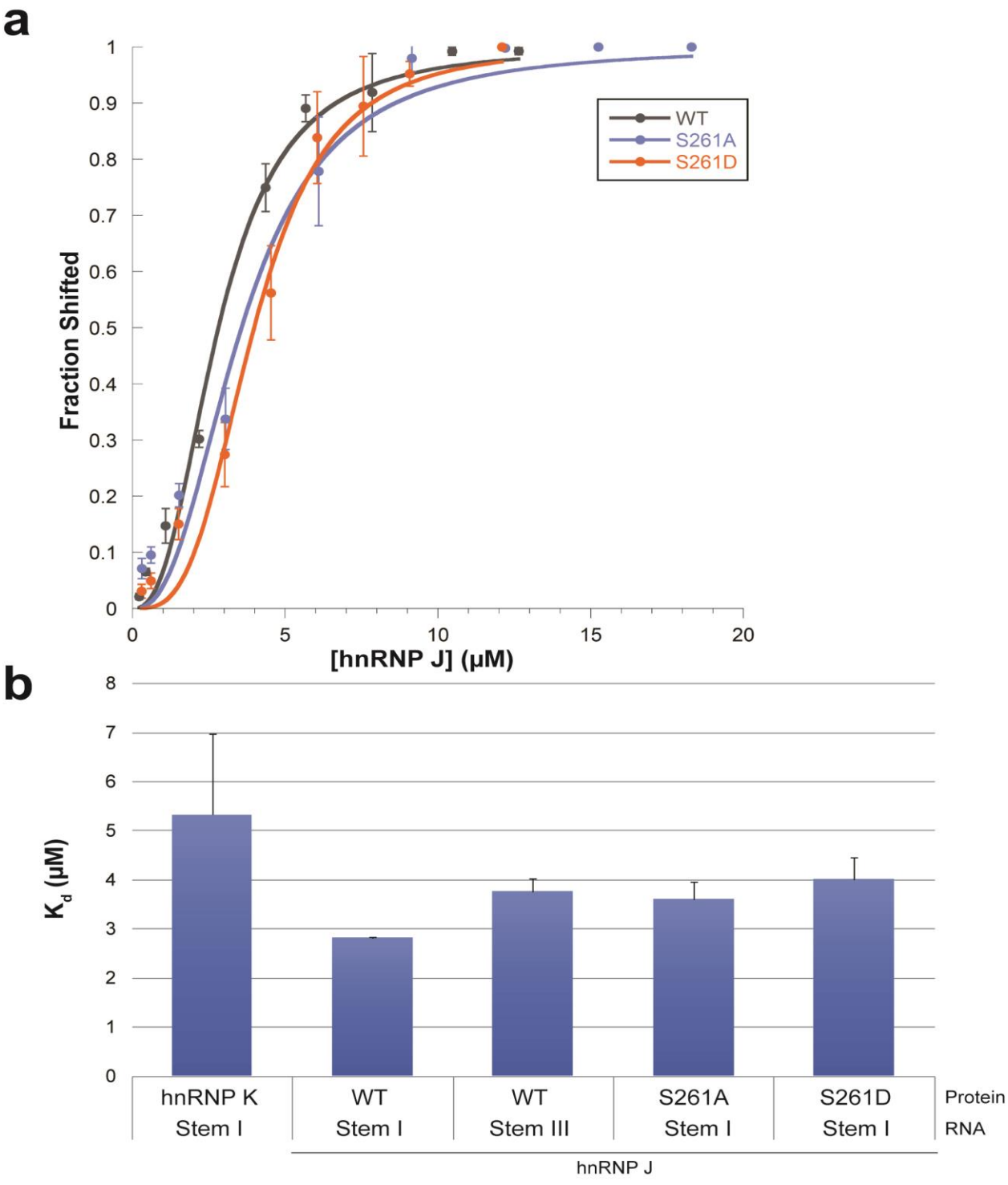

Figure 3.3. Phospho-mimetic hnRNP J binds Stem I with less affinity. (a) Representative binding curves calculated from the average ± standard deviation of four EMSA replicates with the given hnRNP J protein construct and Stem I of 7SK RNA. (b) Average K_d ± standard deviation calculated from three to six EMSA replicates with the indicated protein and 7SK RNA constructs.

We next wanted to assay whether direct phosphorylation via P-TEFb could recapitulate our phospho-mimetic results. Therefore, we first performed a non-radioactive kinase assay with or without P-TEFb and WT hnRNP J. Instead of adding acid stop mix, the kinase assays were titrated into EMSA reactions with Stem I (Figure 3.4a). In order to fully saturate the 100 μM of hnRNP J used in the reactions, we conducted the kinase assays with either 100 μM or 200 μM ATP. The large concentration of ATP required for the kinase assays subsequently hindered the binding of hnRNP J to Stem I, as evidenced by the 3 to 4-fold larger K_d in the negative control reactions at 100 μM and 200 μM ATP, respectively. However, in the presence of P-TEFb, we saw a consistent decrease in affinity compared to negative control reactions. To decrease the large negative charge contribution from the saturating ATP concentrations, we repeated the assay at 200 μM ATP, but purified the kinase assay over a Sephadex G-25 micro-spin column prior to use in EMSAs. Unfortunately, this resulted in weak binding. When the small fraction that shifted is plotted against protein concentration, however, it appears that, again, reactions containing P-TEFb bind marginally worse to Stem I than negative control reactions (Figure 3.4b). As an alternative approach, we generated hnRNP J N-terminally tagged with maltose binding protein (MBP-hnRNP J). Immobilization of MBP-hnRNP J to amylose resin allowed us to thoroughly wash the reactions prior to elution. When used in EMSAs with Stem I, we could partially restore binding affinities near WT hnRNP J (4.8 ± 0.3 μM versus 2.8 ± 0.03 μM, respectively) (Figure 3.4c). As seen before, addition of P-TEFb marginally, yet consistently, decreases the affinity of hnRNP J for Stem I. These results, in conjunction with the data from Figure 3.3, indicate that P-TEFb-dependent phosphorylation of S261 of hnRNP J results in a decreased affinity for 7SK RNA.

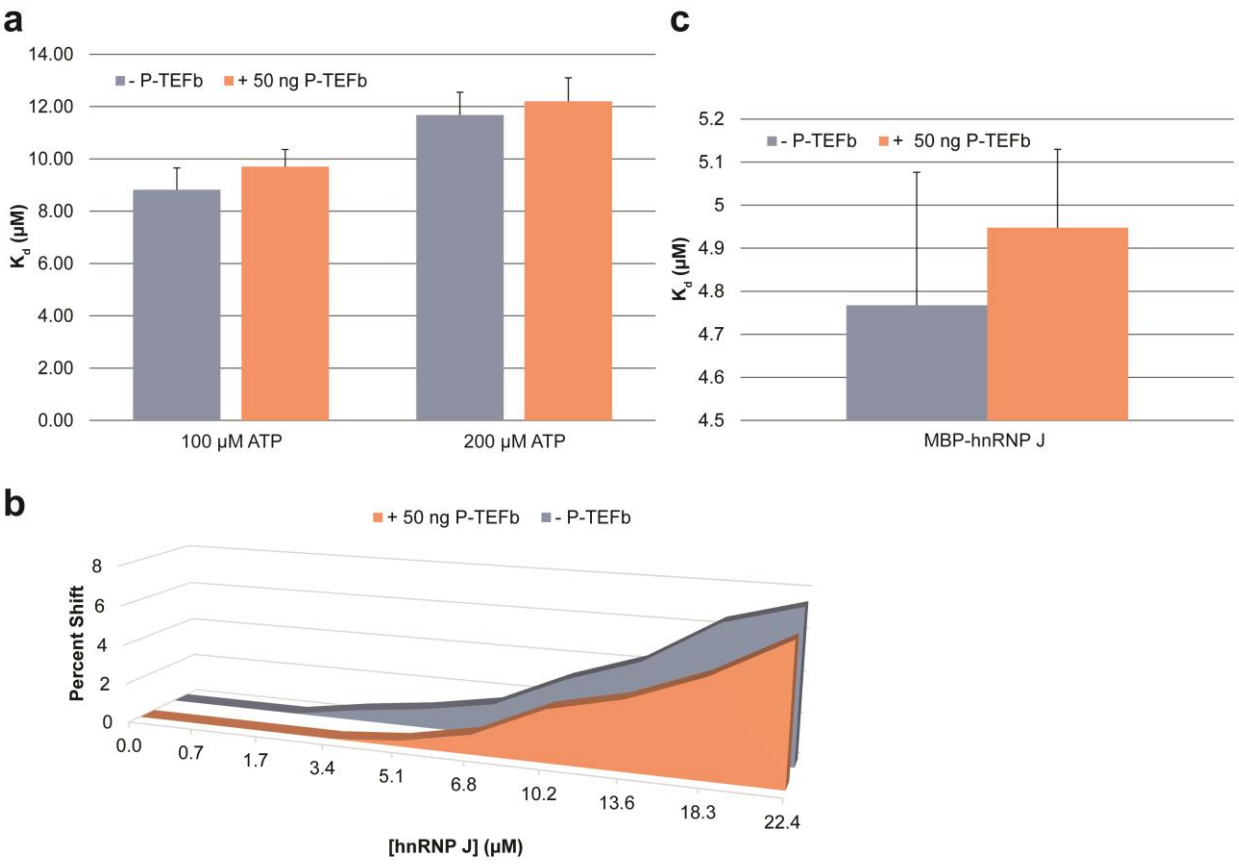

Figure 3.4. Direct phosphorylation of hnRNP J by P-TEFb decreases affinity for Stem I. (a) Average K_d ± standard deviation calculated from three EMSA replicates using hnRNP J directly from kinase assays containing either 100 μM or 200 μM ATP. (b) Reactions as in (a) were subjected to G-25 spin column purification (GE Healthcare, #27-5325-01) prior to use in EMSAs with Stem I. Average percent shift of two replicates are plotted. (c) hnRNP J N-terminally tagged with maltose binding protein (MBP) pre-bound to amylose agarose resin was subjected to kinase assays prior to elution and EMSAs with Stem I. Average K_d ± standard deviation of four replicates is plotted.

3.3.4 Phosphorylation of S261 increases HEXIM competition for Stem I

To date, when immunoprecipitated from cells, most hnRNPs have been observed to bind to Stem III of 7SK RNA [151]. While we observed hnRNP J binding to Stem III (Figure 3.1c), the highest-affinity interaction occurs with Stem I (Figure 3.3b), and the predicted binding sites (Figure 3.1b) of hnRNP J directly overlap known HEXIM-interaction motifs [190, 191, 195] and nucleotides [189]. Whereas hnRNP K has the potential to interact with double stranded nucleic acids [267, 282], the protein predominantly associates with single stranded RNA [298, 306, 307], while HEXIM has been shown to specifically recognize and bind structured RNA [188, 191]. Additionally, the proteins appear to assemble on mutually exclusive complexes, and siRNA-mediated knock-down of hnRNP K drives increased accumulation of the 7SK P-TEFb RNP in cells [203]. We therefore examined whether HEXIM and hnRNP J compete for binding to Stem I of 7SK RNA.

Initial attempts to visualize competition assays through EMSAs resulted in HEXIM/Stem I and hnRNP J/Stem I complexes having shifts with identical mobilities within native gels (data not shown). We therefore utilized the MBP-hnRNP J fusion protein to add additional mass and charge in order to efficiently separate the complexes on a native polyacrylamide gel (Figure 3.5a). Of note, EMSAs with MBP-hnRNP J and Stem I revealed that the N-terminal tag does significantly decrease affinity for Stem I ~2-fold (Figure 3.6), and therefore the following results must be interpreted with caution in comparing “endogenous” hnRNP J and HEXIM competitions. Titration of HEXIM into reactions containing pre-bound MBP-hnRNP J initially results in a non-specific smear of higher mobility (Figure 3.5a left). At higher concentrations, however, formation of HEXIM/Stem I complexes appears. The non-specific smear is not

recapitulated when MBP-hnRNP J is titrated into reactions containing pre-bound HEXIM/Stem I complexes (Figure 3.5a right), suggesting that hnRNP J binding restructures the RNA, and HEXIM displacement of hnRNP J initially forms a heterogeneous population of Stem I complexes.

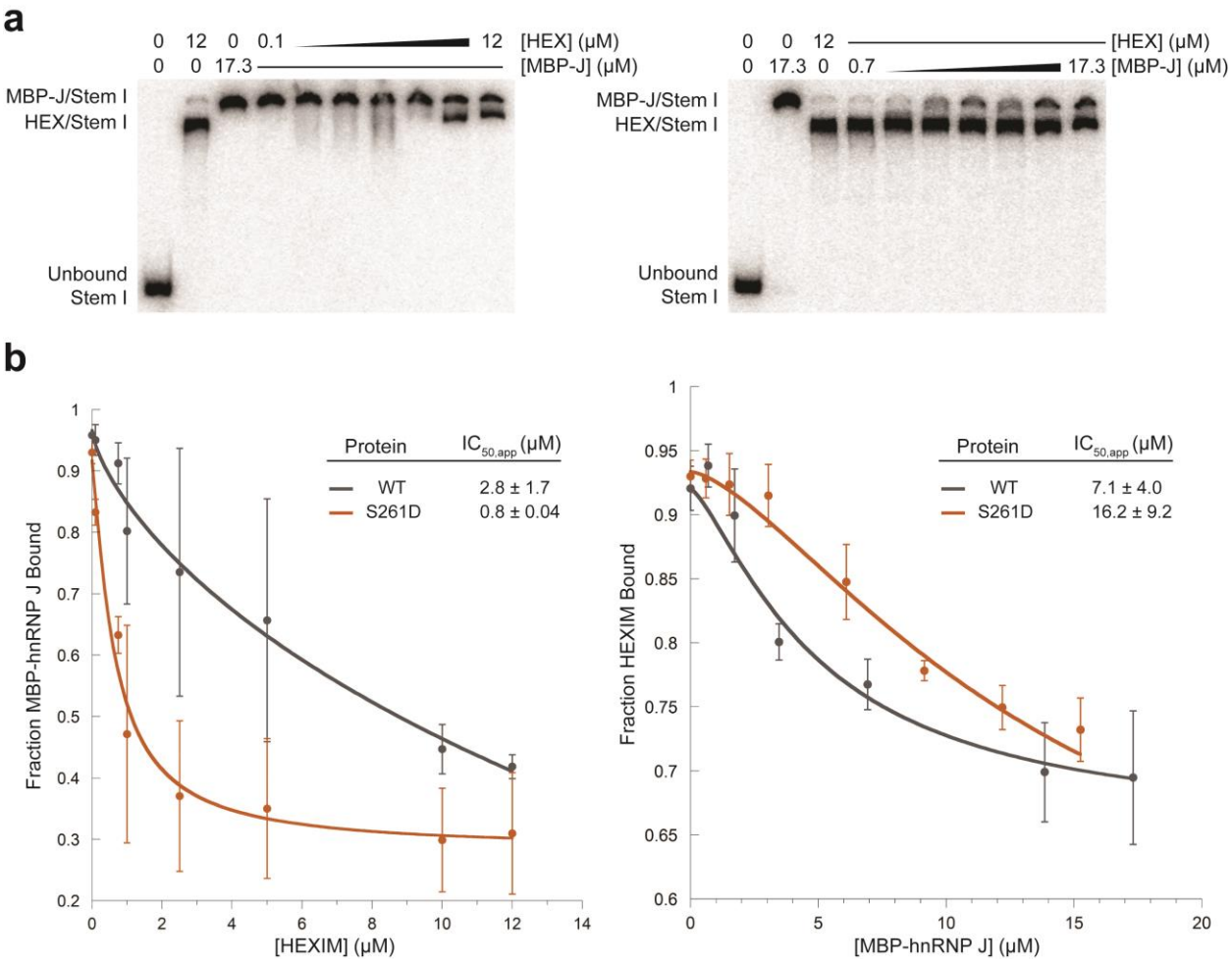

Figure 3.5. HEXIM efficiently competes for binding Stem I. (a) EMSAs of competition assays with Stem I pre-bound to MBP-hnRNP J (MBP-J, left) or HEXIM (HEX, right) and titrated with the other protein. (b) Average competition curves plotting fraction pre-bound protein shift against competitor protein. $IC_{50,app}$ is calculated from three replicates ± standard deviation.

To better understand the competition between HEXIM and MBP-hnRNP J, we calculated $IC_{50,app}$ values for each competition. In agreement with published affinities for HEXIM and Stem I (~500 nM) [191] and our MBP-hnRNP J – Stem I affinity of 6.1 μM (Figure 3.6), HEXIM displays a higher degree of competition for Stem I than MBP-hnRNP J (2.8 ± 1.7 μM versus 7.1 ± 4.0 μM, respectively). We next investigated whether S261 phosphorylation, which lowers hnRNP J affinity for Stem I (Figures 3.3 and 3.4), would impact HEXIM competition for Stem I. To our surprise, HEXIM dissociates half of the S261D MBP-hnRNP J/ Stem I complex at a 3.5-fold lower concentration than WT protein (0.8 ± 0.04 μM), while it takes over 2-fold more S261D hnRNP J to dissociate half of HEXIM from Stem I (Figure 3.5b). These results suggest that S261 not only aids in mediating nucleic acid interactions, but phosphorylation of this residue by P-TEFb may aid in increased HEXIM-mediated dissociation of hnRNP J from 7SK RNA.

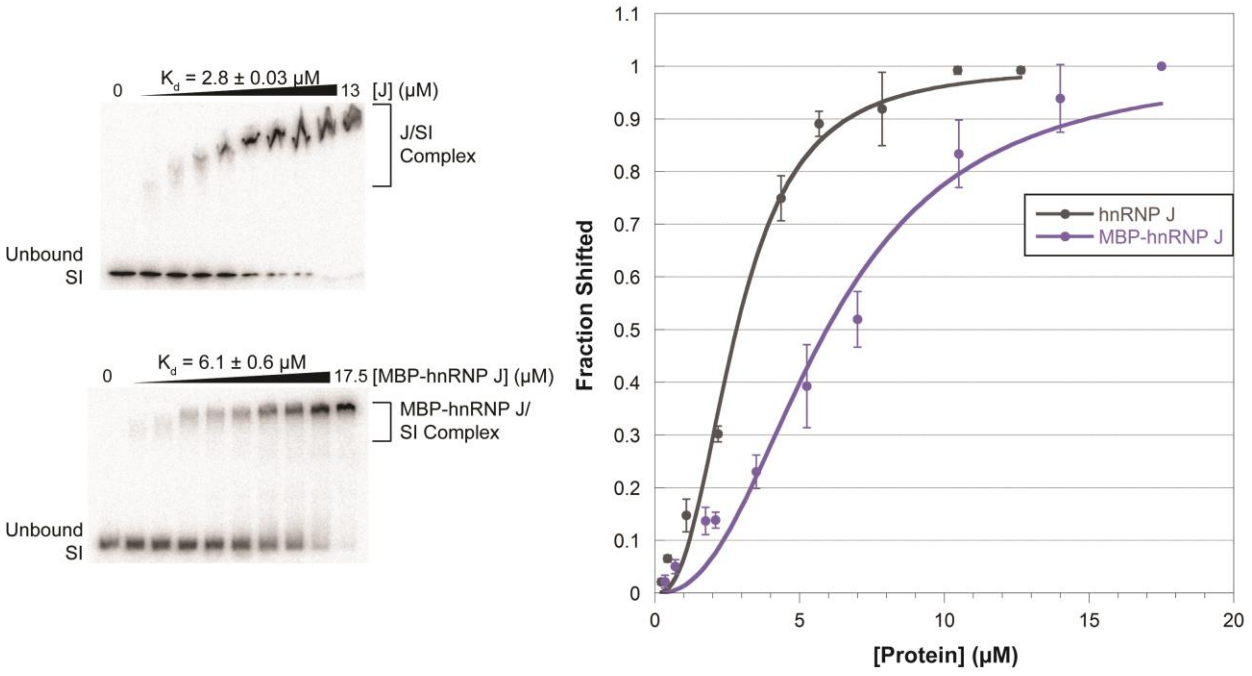

Figure 3.6. Addition of N-terminal maltose-binding protein tag decreases affinity of hnRNP J for Stem I of 7SK RNA. Left – Representative EMSAs with either WT hnRNP J (top) or MBP-tagged WT hnRNP J (MBP-

hnRNP J, bottom). K_d is calculated from three (hnRNP J) or six (MBP-hnRNP J) replicates ± standard deviation. Right – curve fit for average fraction of Stem I shifted ± standard deviation of three (hnRNP J) or six (MBP-hnRNP J) replicates.

3.3.5 P-TEFb-dependent phosphorylation of hnRNP K does not induce nuclear export

Previous studies have shown that S284 (hnRNP J S261) can be phosphorylated by extracellular signal-regulated kinase 1 and 2 (ERK1/2) [293, 308, 309]. Phosphorylation of S284 and S353 by ERK1/2 drives cytoplasmic accumulation of hnRNP K upon mitogen stimulation [293]. We therefore decided to see if P-TEFb-dependent phosphorylation would also relocalize hnRNP K to the cytoplasm. To address this question, we performed nuclear fractionation [310] of HeLa cells treated for 30 minutes with either DMSO or 1 μM flavopiridol prior to release for 15 minutes in drug-free media (Figure 3.7). Inhibition of P-TEFb with flavopiridol dissociates the protein from the 7SK P-TEFb RNP [170], thus increasing active concentrations of P-TEFb upon drug release.

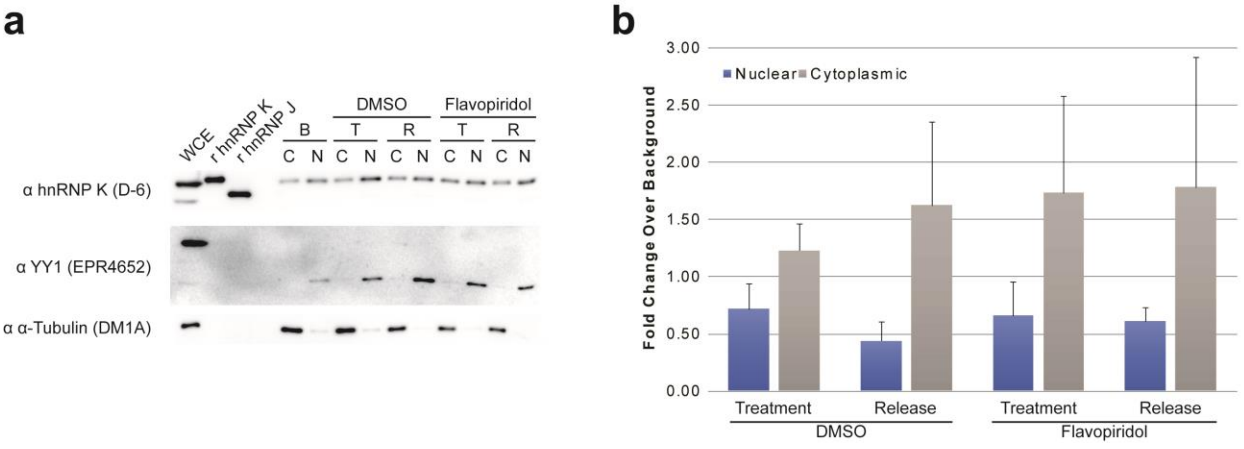

Figure 3.7. Cells treated with flavopiridol do not accumulate hnRNP K in the cytoplasm. (a) Representative Western blots of nuclear fractionation experiments with cells treated (T) for 30 minutes with either DMSO or 1 μM flavopiridol prior to release (R) for 15 minutes. WCE, whole cell extract; B, background (no treatment); C,

cytoplasmic fraction; N, nuclear fraction. Recombinant (r) hnRNP K and J are used as size marker controls. YY1, nuclear control; α-tubulin, cytoplasmic control. (b) Quantification of (a). Pixel density of hnRNP K bands were normalized to loading controls. Average of four replicates ± standard deviation is plotted.

Western blot analysis of the nuclear and cytoplasmic fractions with appropriate controls (YY1 and α-tubulin, respectively) demonstrates effective fractionation with minimal cross-contamination (Figure 3.7a bottom panels). As reported previously [203], hnRNP J is poorly expressed (Figure 3.7a 1st lane) and was thus not included in our analysis. We observed higher hnRNP K signal in nuclear fractions relative to cytoplasmic fractions, in agreement with previous immunofluorescence data [292, 293]. Treatment with either DMSO or flavopiridol results in increased cytoplasmic accumulation of hnRNP K with a concurrent loss of nuclear signal (Figure 3.7b). Release of either treatment marginally increases the observed phenotype; however, we do not observe any difference between treatment with DMSO or flavopiridol. These results suggest that P-TEFb-dependent phosphorylation does not induce cytoplasmic accumulation of hnRNP K under the regime tested.

3.3.6 S261 aids in *MYC* transcription termination maintenance

In an effort to identify a potential role for P-TEFb-dependent phosphorylation of S261 of hnRNP J, we once again carefully examined the local architecture of the KI domain (Figure 3.1a). Poly-proline repeats often mediate protein-protein interactions [311], and hnRNP K can potentially interact with over 200 proteins [290]. Recent work has identified XRN2, an exonuclease required for proper RNA polymerase II termination [312], as a target of P-TEFb, whereby phosphorylation increases XRN2 activity [305]. Interestingly, it has been suggested that hnRNP

K also interacts with and recruits XRN2 to the 3' end of genes [313]. Therefore, we hypothesize that P-TEFb-dependent phosphorylation of hnRNP J at S261 not only aids in dissociation from the 7SK-hnRNP RNPs (see above), but also serves as a marker that aids in the recruitment of XRN2, and thus proper termination maintenance.

To test this hypothesis, we overexpressed WT, S261A, and S261D C-terminally Myc-FLAG tagged hnRNP J (hnRNP J-MF) in HeLa cells (Figure 3.8b). 24-hours after transfection, total RNA was extracted, converted to cDNA, and used in RT-qPCR with primer sets placed before or after the poly-adenylation signal (PAS) at the loci *CCNB1*, *FRAT2*, and *MYC* (Figure 3.8a) [305]. Transcript abundance amplified after the PAS was normalized across treatment conditions to the amplicon generated before the PAS, and then compared to mock-transfected cells to calculate the fold change. An increase in transcripts generated past the PAS is indicative of termination defects. We observed a different phenotype at each locus tested (Figure 3.8c). Overexpression of hnRNP J-MF has no effect on proper termination at *CCNB1*. While overexpression of WT hnRNP J-MF does not induce a termination defect at *FRAT2*, expression of either S261A or S261D increased accumulation of transcripts past the PAS by ~1.5-fold (Figure 3.8c). However, overexpression of hnRNP J, regardless of the construct tested, significantly (student's T-test $p \leq 0.002$) increases transcript abundance after the PAS at *MYC* (Figure 3.8c). Of note, expression of S261A hnRNP J-MF marginally exacerbates the termination defect detected at *MYC* compared to WT. Together, these results suggest hnRNP J may aid in proper termination of RNA polymerase II, potentially through recruitment of XRN2 mediated via the phosphorylation status of S261. Additional studies are required to examine the link between S261 phosphorylation, P-TEFb, XRN2, and termination; to identify a direct

interaction between XRN2 and hnRNP J via the KI domain; and to dissect the prevalence of hnRNP J-mediated termination throughout the genome.

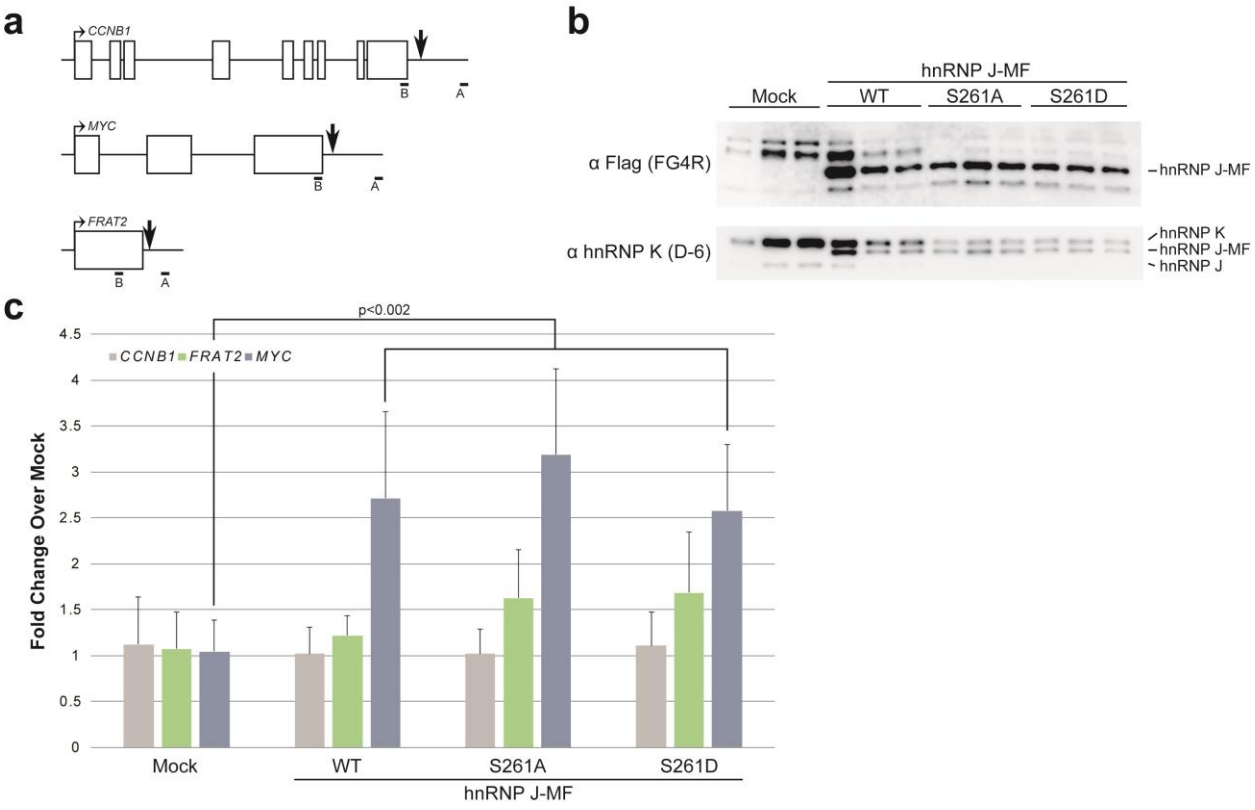

Figure 3.8. Overexpression of hnRNP J S261 mutants increases accumulation of transcripts past poly-adenylation signals. (a) Cartoon representation of genes analyzed. Boxes indicate exons; arrows are annotated poly-adenylation signals (PAS). General location of primers used for RT-qPCR set before (B) and after (A) the PAS. Adapted from [305]. (b) Representative Western blot of three independent transfections of C-terminally Myc-FLAG-tagged hnRNP J (hnRNP J-MF). (c) Signal amplified after the PAS is normalized to transcript abundance before the PAS as measured via RT-qPCR. Average of six replicates ± standard deviation is plotted.

3.4 DISCUSSION

Here we report how competition between HEXIM and an hnRNP for access to Stem I of 7SK RNA is regulated through phosphorylation via P-TEFb itself. We found that hnRNP K isoforms 2 and 3 (hnRNP J) directly bind Stems I and III of 7SK RNA. In agreement with immunoprecipitation data from cells [203], hnRNP J has significantly higher affinity for Stem I than hnRNP K. We also report the *in vitro* identification of two new substrates for P-TEFb: hnRNP K and J at S284 or S261, respectively. Furthermore, we find that P-TEFb-dependent phosphorylation not only decreases the affinity of hnRNP J for Stem I of 7SK RNA, but it also enhances the dissociation of hnRNP J via HEXIM competition. In identifying a potential significance for P-TEFb-dependent phosphorylation in cells, we discovered that P-TEFb phosphorylation does not induce cytoplasmic accumulation of hnRNP K. Instead, we find that S261 of hnRNP J may play a role in regulating the recruitment of XRN2 to the 3' end of select RNA polymerase II genes. Together, these results suggest a mechanism by which P-TEFb autoregulates incorporation into the 7SK-P-TEFb RNP by phosphorylating hnRNP J, thereby decreasing affinity of hnRNP J for Stem I while simultaneously increasing HEXIM-mediated dissociation from Stem I. Once bound to Stem I, HEXIM can associate with P-TEFb and hold it catalytically inactive. Furthermore, once dissociated from 7SK RNA, phosphorylated hnRNP J may interact with termination machinery and maintain proper termination boundaries for RNA polymerase II (Figure 3.9).

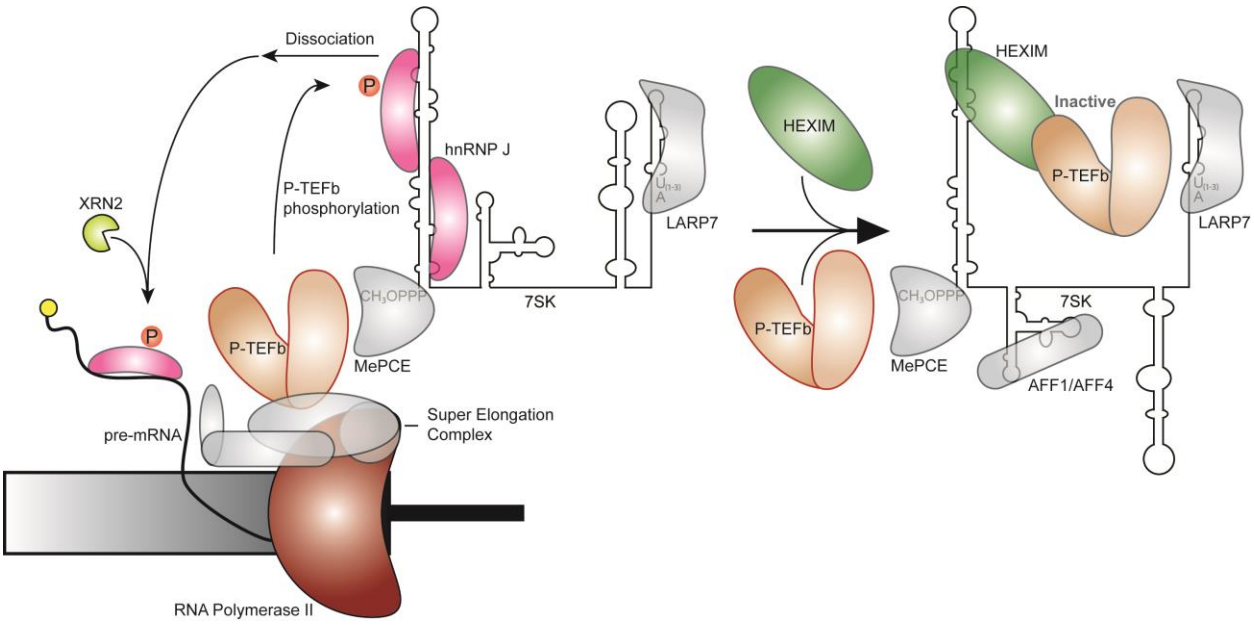

Figure 3.9. P-TEFb autoregulation through phosphorylation of hnRNP J. As P-TEFb, traveling with RNA polymerase II as part of the super elongation complex, nears the 3' end of a gene, it can phosphorylate hnRNP J bound to 7SK RNA. This increases the dissociation rate of hnRNP J and allows HEXIM to bind Stem I, thereby binding and inactivating P-TEFb within the 7SK-P-TEFb RNP (right). Phosphorylated hnRNP J may then interact with termination machinery to aid in recruitment of XRN2 and properly terminate RNA polymerase II.

3.4.1 hnRNP K and J directly interact with 7SK RNA

Initial characterization of the interaction between hnRNP K and 7SK RNA did not elucidate the region(s) to which hnRNP K binds [203]. Utilization of a scoring matrix where every nucleotide is weighted across the consensus sequence of 5'- cCAUc(N_{2-7})wCCCw(N_{7-18})UCAYc-3' (where W=A/T, Y=C/T, N=A/G/U/C, and subscripts denote length of bases) [298] identifies three high-probability binding sites (Figure 3.1b). Our results confirm that the scoring matrix accurately predicts hnRNP K binding, as hnRNP J has a higher affinity for Stem I than Stem III of 7SK RNA (Figure 3.3b). Because hnRNP J did not bind Stem IV and all experiments were performed

in the presence of nonspecific competitors, we are confident that the observed interactions are specific and would likely present the *in vivo* binding regions of hnRNP J.

We also found that hnRNP K isoform 3 binds 7SK RNA with higher affinity than hnRNP K isoform 2 does. This agrees with immunoprecipitation data in which 4-fold more hnRNP J associates with 7SK RNA than hnRNP K [203]. Unlike hnRNP K, hnRNP J lacks 23 amino acids within the linker between the RNA-binding domains KH1 and KH2 (Figure 3.1a). Biochemical and biophysical studies of hnRNP K suggest that KH1 and KH2 act as a concerted unit when binding to nucleic acids [287, 314]. However, the linker between KH1 and KH2 is large enough to force a spacing constraint between the 4-nucleotide recognition sequences (see the above consensus sequence) [298]. With a shorter linker, KH1 and KH2 in hnRNP J may behave more like the tandem KH domains of NusA [284], in which the KH domains bind a continuous stretch of nucleic acids [315]. This architecture, and alternative binding mode, may explain the significantly higher affinity gained by hnRNP J.

3.4.2 P-TEFb phosphorylates hnRNP J at S261

To date, P-TEFb is considered to only phosphorylate three substrates: the C-terminal domain of RNA polymerase II [99], the Spt5 subunit of DSIF [90], and the NELF-E subunit of NELF [91]. However, a chemical genetic screen expressing a modified CDK9 subunit identified a large number of P-TEFb-specific substrates – including XRN2 [305]. Here, we provide evidence that hnRNP J is also phosphorylated by P-TEFb *in vitro* (Figure 3.2). We found that P-TEFb predominantly phosphorylates S261, which resides in a canonical CDK-recognition sequence of (S/T)Px(R/K) (i.e. SPRR) [316-318]. S261 (S284 in hnRNP K) can also be phosphorylated by ERK1/2 [293, 308], which suggests that phosphorylation of S261 is a central post-translational

modification important for transducing regulatory stimuli to affect hnRNP J. Unlike phosphorylation via ERK1/2, which induces shuttling of nuclear hnRNP K out into the cytoplasm [293], inhibition and subsequent release of P-TEFb through flavopiridol does not affect localization patterns of hnRNP K (Figure 3.7). However, our study does not address the phosphorylation of hnRNP K or J by P-TEFb *in vivo*. Additionally, while drug treatment did not affect the localization of hnRNP K in cells, we were not able to directly examine the effect on hnRNP J, which may, through an unknown mechanism, be the only hnRNP K isoform target of P-TEFb *in vivo*.

In the context of 7SK RNA, we found that a phospho-mimetic mutant, S261D, and WT hnRNP J phosphorylated by P-TEFb *in vitro* have decreased affinity for Stem I (Figures 3.3 and 3.4). Furthermore, S261D hnRNP J is better dissociated from Stem I by HEXIM (Figure 3.5). While the published affinity for HEXIM is 9-fold better than hnRNP J [191] (~500 nM versus 2.8 µM) and HEXIM has a lower $IC_{50,app}$ than hnRNP J (2.8 µM versus 7.1 µM), hnRNP K is ~27x more abundant in cells [262, 319]. A back-of-the-envelope calculation for hnRNP J concentration (assuming 4-fold lower levels than hnRNP K [203]) reveals it is ~7x more concentrated than HEXIM in the nucleus. Therefore, our results strongly suggest that the decreased affinity and competition induced by S261 phosphorylation may play a critical role in tipping the balance towards restructuring the 7SK-hnRNP RNPs into the 7SK-P-TEFb RNP via swapping of hnRNP J for HEXIM. Surprisingly, this implies that P-TEFb autoregulates active kinase levels and aids in conversion to the 7SK-P-TEFb RNP (Figure 3.9).

3.4.3 S261 of hnRNP J may help regulate transcription termination

A chemical genetic screen found that P-TEFb phosphorylates XRN2 to stimulate its exonuclease processivity. Inhibition of P-TEFb through several different drugs led to large termination defects and accumulation of transcripts beyond the PAS [305]. A separate study found that hnRNP K localizes to the 3' end of genes. Knock-down of hnRNP K also led to accumulation of transcripts 3' of the PAS, and while RNA polymerase II levels were not reduced near the PAS, XRN2 recruitment was drastically decreased [313]. Because P-TEFb is incorporated into the super elongation complex and travels with RNA polymerase II [96], it stands to reason that active P-TEFb must be present during termination. Disassembly of the elongation complex would therefore release active P-TEFb. The free kinase needs to be inhibited through the 7SK-P-TEFb RNP to prevent spurious unpausing and proper localization to other paused RNA polymerases. We therefore hypothesized that P-TEFb phosphorylation of hnRNP J not only aids in dissociation from the 7SK-hnRNP RNPs, but may act as a mark to enhance XRN2 recruitment and subsequent phosphorylation by P-TEFb. This is an especially attractive model considering S261 is located immediately N-terminal to a stretch of RGG/RG boxes and poly-proline tracts, two important elements that aid in protein-protein interaction [288, 320].

When we examined transcript abundance past the PAS for *CCNB1*, *FRAT2*, and *MYC*, we found that each gene behaves differently in the presence of either WT, S261A, or S261D hnRNP J overexpression (Figure 3.8). Unfortunately, no clear consensus can be reached from our initial data. However, it appears that S261A overexpression produces marginally higher transcript accumulation 3' of the PAS than S261D or WT hnRNP J overexpression, suggesting that hnRNP J-mediated termination maintenance relies on both the addition and the removal of the phosphate moiety. This may be tied to arginine methylation, as the terminal arginine in the P-TEFb

recognition sequence of SPRR has been shown to be methylated [321]. Notably, a number of studies have found that arginine methylation affects the phosphorylation status of serines located 3 amino acids away (i.e. S261) [322-325], and arginine methylation near poly-proline tracts (a $poly(Pro)_6$ tract starts at amino acid 265, one after the P-TEFb-recognition sequence) influences protein recognition [326]. Although several more experiments are required to effectively demonstrate a direct link between P-TEFb, hnRNP J, and XRN2 and termination, these initial results provide a hint at the complex regulatory loop between elongation and termination mediated through P-TEFb.

4.0 CONCLUSIONS AND FUTURE DIRECTIONS

Given the prevalence and importance of promoter proximal pausing [105], it is surprising to find that very few studies [170, 202-204] have focused on understanding the functional consequence, or even transition between, the 7SK RNPs. Indeed, almost every study on 7SK RNA focuses on P-TEFb association and dissociation (i.e., the 7SK-P-TEFb RNP). Therefore, I took a different approach and began with an investigation into the effects of hnRNP binding to 7SK RNA. In the preceding chapters, I have described my efforts to elucidate the contributions of hnRNPs A1 and K/J in maintaining the alternative 7SK RNPs with the goal of eventually understanding the complex choreography of protein binding, RNA restructuring, and RNP transitioning that is necessary to intimately regulate protein-coding transcription through RNA polymerase II. Perhaps unsurprisingly, I have uncovered that the maintenance and transition between the 7SK RNPs is not a simple stochastic model of protein association and dissociation, but a complex interplay between competing proteins fighting for access to the highly abundant nuclear non-coding RNA. Therefore, local concentrations between the competing proteins will be a large driving force in the regulation of P-TEFb. In fact, all hnRNPs that bind 7SK RNA also shuttle between the nucleus and the cytoplasm, providing a convenient mechanism for quickly modulating active transcriptional levels [205].

What factors could influence the local concentrations of competing proteins? Because SRSF2 is strictly a nuclear protein [214], the concentration of hnRNP A1 must drive the

competition. Interestingly, hnRNP A1 contains a non-canonical nuclear shuttling domain known as M9 [231]. Without a classical nuclear localization signal (NLS), the M9 domain forces hnRNP A1 to shuttle between the nucleus and cytoplasm through a mechanism that is dependent on active transcription, such that increased transcription concomitantly increases nuclear accumulation [327, 328]. Additionally, hnRNP A1 has other functions throughout the nucleus, including mediating alternative splicing and telomere maintenance [226]. Therefore, alternative functions of hnRNP A1 could drive changes in local concentrations. Thus, by increasing transcription, P-TEFb activity would initially increase hnRNP A1 concentrations within the nucleus, aiding in dissociation of SRSF2 from 7SK-P-TEFb RNPs. However, as the demand for functions outside of binding 7SK RNA increase with the extra transcriptional load, local hnRNP A1 concentrations will decrease and allow SRSF2 to help reform the 7SK-P-TEFb RNP.

In the case of hnRNP K/J and HEXIM, I found that post-translational modifications (PTMs) drive the competition. hnRNP K isoforms 1 and 2 shuttle between the nucleus and cytoplasm [205]; however, unlike hnRNP A1, the classical N-terminal NLS overrides transcriptional dependence for shuttling [292, 327, 329]. Instead, PTMs, predominantly phosphorylation, have been shown to elicit cytoplasmic accumulation [293, 330]. Interestingly, we and others [330] have found that hnRNP K isoform 3 (hnRNP J) appears to strictly localize within the nucleus. Additionally, hnRNP J, HEXIM1, and HEXIM2 are not as highly expressed as the other proteins investigated in this thesis [319]. Therefore, to tip the scale towards formation of the 7SK-P-TEFb RNP, P-TEFb itself will phosphorylate hnRNP J, leading to decreased affinity of hnRNP J for Stem I and increased HEXIM-mediated dissociation. Like hnRNP A1, it appears that upon phosphorylation, hnRNP J proceeds to mediate alternative functions, such as regulating termination, which would decrease local active concentrations.

A common theme runs throughout both mechanisms: changes in local concentrations appear to be mediated through P-TEFb itself. This could be an indirect consequence, such as generally increasing transcription, which drives demand for alternative hnRNP A1 functions. It could also be a direct mechanism, such as phosphorylating hnRNP J and mediating 7SK RNP dissociation and RNA polymerase II termination. Either way, it appears that P-TEFb autoregulates active levels within the nucleus by destabilizing the 7SK-hnRNP RNPs and promoting the formation of the inactive 7SK-P-TEFb RNP (Figure 4.1). This idea of autoregulation is supported by a study that found that rising levels of active P-TEFb directly leads to increased transcription of HEXIM1 mRNA by unpausing a special reserve of RNA polymerase II located upstream of the annotated promoter [331]. This, in turn, increases HEXIM1 production, and thus drives formation of inhibitory 7SK-P-TEFb RNPs.

The results gleaned from chapters 2 and 3 indicate that the 7SK-hnRNP RNPs are not just a "molecular sink" for binding various hnRNPs. Nor are they merely "place-holder" complexes that tie up 7SK RNA until it is needed for inactivating P-TEFb. Instead, it appears that specific proteins are binding to specific regions to elicit specific changes that work together to regulate metazoan transcription (Figure 4.1). Indeed, there is evolutionary constraint on "core" 7SK RNA elements [165, 332, 333], particularly Stems I, III, and IV. Because most nucleotide expansions and contractions occur between "core" elements [165], this implies that the nucleotide sequence is necessary not only for dictating specific protein interactions, but that folding of the RNA must be important for the correct function (see Chapter 2, Figures 2.10 and 2.11, for an example with Stem III). In order to truly understand how 7SK RNA regulates transcription, future studies should address the functional consequence of forming at least two different 7SK-hnRNP RNPs. Does each 7SK-hnRNP RNP have its own regulation and competition? Do they regulate specific

subsets of genes through modulating active levels of the bound hnRNPs? Do they restructure 7SK RNA differently – and does this impact maintenance and transition? These questions and more can be used as a framework to aid in establishing how this abundant non-coding RNA regulates transcription in metazoans, as well as the grounds for investigation into novel anti-cancer-related therapies (see below).

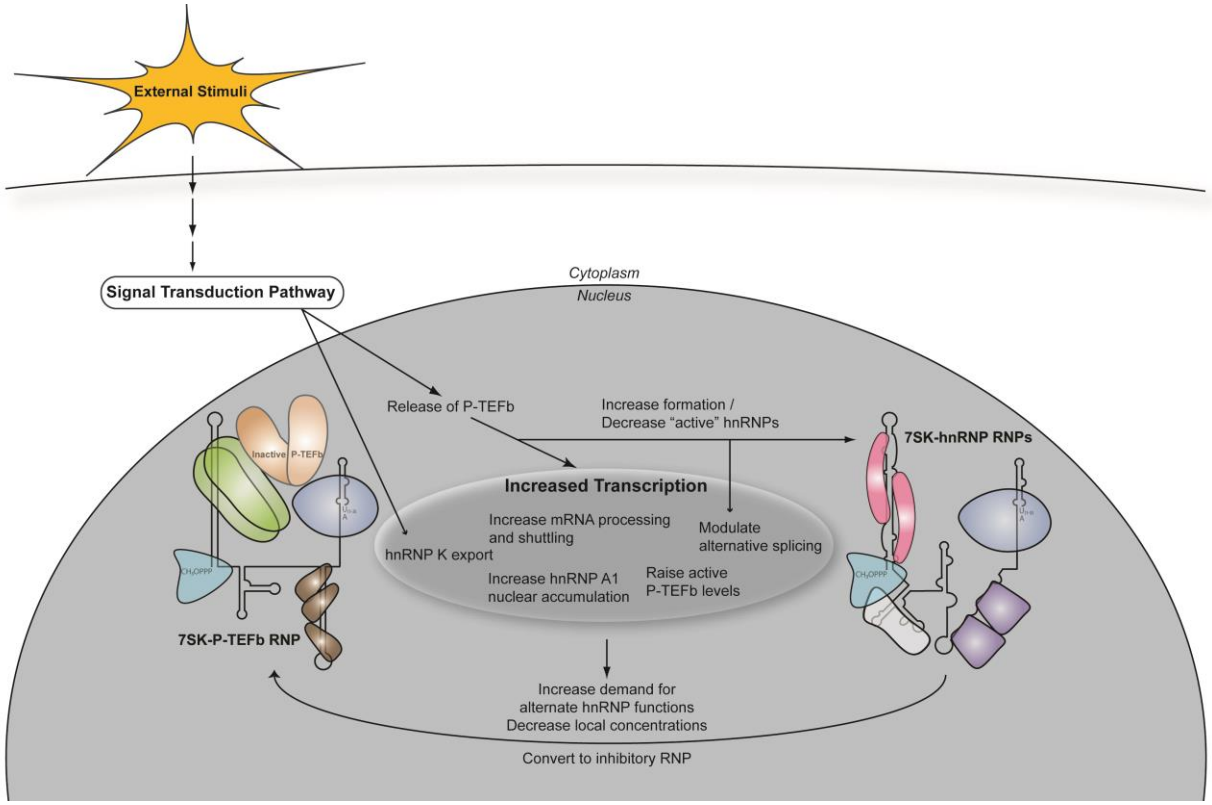

Figure 4.1. Working model of transcription-dependent P-TEFb autoregulation.

4.1.1 Implications for disease

Dysregulation of 7SK RNA is directly linked to several disease states. Because 7SK RNA regulates promoter proximal pausing, which controls genes governing development and proliferation [105], 7SK RNP dysregulation results in uncontrolled cell proliferation (one hallmark of cancer progression [334]). However, 7SK RNP components are also implicated in

the progression of cardiac hypertrophy (through decreased expression of HEXIM [335, 336]) and HIV activation (through hijacking P-TEFb via Tat [192] to specifically increase transcription of the viral genome [337]). As P-TEFb itself is an essential protein not only in lower eukaryotes [338], but also during development in metazoans [207], most disease-causing mutations occur within the protein components regulating 7SK RNA stability or function.

Several disease-causing phenotypes are linked to LARP7 mutations. Loss of function of LARP7 exposes the 3' end of 7SK RNA to exonucleases and therefore leads to its decreased abundance [169]. Indeed, decreased expression of 7SK RNA due to a loss-of-function mutation in LARP7 is suspected to cause a novel familial form of primordial dwarfism [339], highlighting the importance in regulating P-TEFb during development. Spontaneous mutations in LARP7 are correlated with several forms of cervical and gastric cancer [340, 341], suggesting that tight regulation of 7SK RNA expression is necessary to maintain quiescence. Indeed, one study found that LARP7 knock-down results in metastatic phenotypes throughout a panel of human breast cancer cell lines, while reintroduction of LARP7 into a highly invasive metastatic cell line significantly decreases colony formation [208]. Together, these studies demonstrate the critical importance in maintaining 7SK RNA levels towards preserving cellular development and homeostasis.

Because of the pleiotropic effects of hnRNP A1 (Section 2.1.1), the protein is involved in a number of diseases ranging from neurodegeneration to cancers [226]. Many of these diseases are correlated with splicing defects caused by changes in hnRNP A1 abundance [226]. While no study has directly linked diseases to hnRNP A1 binding to 7SK RNA, artificial destabilization of 7SK RNA through LARP7 or MePCE siRNA-mediated knock-down does affect the alternative splicing of candidate genes in zebrafish embryos [206]. As there is more hnRNP A1 (~10^7) [261]

in the cell than 7SK RNA (~10^5) [151], it is highly likely that maintenance of the 7SK RNPs aids in controlling the "active" concentration of hnRNP A1. Indeed, hnRNP A1 is usually overexpressed in cancers [342-344], which, combined with the data from Chapter 2, suggests that a large pool of hnRNP A1 will help drive formation of the 7SK-hnRNP RNPs and increase active P-TEFb levels. It is hard to determine what contribution is specific to hnRNP A1 and 7SK RNA regulation versus alternative splicing, telomere maintenance, mRNA transport, and translation [226], but it is probable that all these processes are intimately tied into 7SK RNP regulation via regulation of transcription.

Like hnRNP A1, hnRNP K is grossly overexpressed in several cancers including colorectal, liver, lung, breast, and oral cancers [280, 345]. Surprisingly, aberrant localization to the cytoplasm is strongly correlated with poor prognosis [346, 347]. Because shuttling of hnRNP K is predominantly regulated via PTMs [273], this finding underscores the critical importance in regulating upstream signaling pathways. P-TEFb-dependent phosphorylation does not appear to induce hnRNP K localization in HeLa cells (Figure 3.7), and therefore does not contribute towards the aberrant cytoplasmic localization found in cancers [346, 347]. As hnRNP K is involved in a myriad of processes (Section 3.1.1), including regulating the transcription of *MYC* [267, 348], untangling the exact contribution of hnRNP K to 7SK RNP regulation and cancer will be a daunting task. However, my work in Chapter 3 helps narrow the focus onto hnRNP J. In support of the notion that hnRNP J aids in regulation of transcription through 7SK RNP maintenance, initial characterization of hnRNP K isoforms noted that a smaller isoform becomes upregulated in proliferating and transformed cells [299]. Our work (Figure 3.7 and immunofluorescence data not shown) and that of others' [330], strongly suggests that hnRNP J is exclusively nuclear, raising the intriguing possibility that it is specifically involved in a subset of

regulation processes distinct from the major hnRNP K isoforms – including 7SK RNP maintenance and transcription termination.

It is important to note that almost all studies into 7SK RNA use a cancer-derived cell line (in particular HeLa) or stem cell line. It will be interesting to compare the composition of 7SK RNPs across a variety of cell lines – ranging from non-transformed primary cells to highly metastatic cancer-derived cell lines. This information will be critical in building a model for 7SK RNP dysregulation across disease progression. Understanding how the cell regulates not only the expression, but the composition of the 7SK RNPs, will provide us with novel avenues of research for drug therapies, understanding the factors that contribute to disease progression, and could potentially lead to the identification of biomarkers for cancer diagnosis and prognosis.

4.1.2 Towards understanding the 7SK RNPs *in vitro*

I began this thesis with the intent of examining the binding of every 7SK RNA-related hnRNP to 7SK RNA. Because promoter proximal pausing is only affected upon knock-down of all hnRNP A members [263], I suspect that hnRNP A2/B1 binding plays a similar, potentially redundant, role as hnRNP A1. Instead, future investigations should focus on examining the roles of hnRNP Q and R. These proteins were previously shown to associate with both Stems I and III through expression of exogenous constructs followed by immunoprecipitation [204]. Like hnRNP K/J, hnRNP Q and R could either bind both regions separately (Figure 3.1c), or, they could bind both stems simultaneously and form a bridge between them [151]. Interestingly, hnRNP Q and R do not bind to the same complex as hnRNP A1 or A2/B1 [204], so the 7SK-hnRNP Q/R RNP may have a novel three-dimensional structure and play a specific regulatory role independent of the 7SK-hnRNP A RNP. It also may have an independent mechanism for hnRNP removal than the

7SK-hnRNP A RNP. Indeed, similar to hnRNP K, phosphorylation of hnRNP Q has been shown to drive cytoplasmic accumulation [349], suggesting that the mechanism for dissociation will probably rely on direct activation via transcription regulators such as P-TEFb.

Once the various 7SK-hnRNP RNPs have been investigated biochemically, future endeavors should focus on establishing the composition of the 7SK-hnRNP RNPs in cells. This can be accomplished using immunoprecipitation of endogenous RNPs followed by quantitative mass spectrometry. It would also be helpful to examine if the different 7SK-hnRNP RNPs regulate distinct subsets of genes by investigating changes in genomic data – including both recruitment of hnRNPs to chromatin through ChIP experiments and RNA-sequencing profiles – across drug treatments, siRNA-mediated knock-downs, or cell lines.

One study found that RNA helicase A interacts with 7SK RNA as a potential component of a 7SK-hnRNP RNP [204]. Two other RNA helicases, DDX21 [200] and DDX6 [350], also interact with 7SK RNA; however, these helicases appear to aid in dissociation of P-TEFb, and thus belong within the 7SK-P-TEFb RNP. Besides phosphorylation of hnRNP J via P-TEFb (Chapter 3), there are no direct mechanisms for efficient dissociation of hnRNPs from the 7SK-hnRNP RNPs. Binding of RNA helicase A to 7SK RNA may serve two functions, neither of which are mutually exclusive. It could hold together secondary structure elements (e.g., packing the HIV1 genome into the nucleocapsid [351]). As a helicase, it could also actively unwind 7SK RNA upon stimulation and aid in hnRNP dissociation. Like the hnRNPs that bind 7SK RNA, RNA helicase A also shuttles between the nucleus and cytoplasm [352]. Therefore, the signal that activates RNA helicase A to unwind 7SK RNA and potentially promote hnRNP dissociation would be a prerequisite for activating other functions of the helicase such as aiding in transcription, mRNA shuttling, and translation [353].

Once the contribution of each 7SK-binding protein has been established, the final frontier will place all the parts together through reconstituting the 7SK-hnRNP RNPs *in vitro*. This will open a new avenue of research to examine the effects of order-of-addition, their competition, establishment of mutually exclusive complexes, and how the RNPs transition. Most importantly, working with the entire 7SK RNA will allow us to investigate how the binding of single-stranded binding proteins to one region of the RNA influences the global fold. Our work and others [195] imply that the folding of 7SK RNA strongly impacts its function. Thus, it is imperative to examine the secondary structure of the *in vitro*-reconstituted 7SK RNPs upon protein addition and dissociation using non-base specific modification methods such as selective 2'-hydroxyl acylation analyzed by primer extension [354]. Finally, since the structures of many of the individual proteins (or domains thereof) that bind 7SK RNA have already been solved [98, 193, 194, 222, 223, 228, 355-357], we can use these data to help fill in models obtained through cryo-electron microscopy of the 7SK RNPs. Together, these data will allow us to examine the intimate relationship between protein binding and RNA structure that work together to regulate metazoan transcription.

4.1.3 Towards exploring *in vivo* consequences of the 7SK RNPs

In Section 4.1.2, I outlined a series of *in vitro* assays that can be used as a basis for understanding the contribution of individual components of the 7SK RNPs, eventually leading to biochemical and structural studies of reconstituted 7SK RNPs. However, 7SK RNA does not exist within an isolated vacuum, solely dictated by the presence of artificial salt compositions and concentrations, laboratory buffers, bovine serum albumin, semi-purified tRNA, and recombinant proteins. The functions mediated by 7SK RNA occur within the crowded and complex

environment of the nucleus [358]. Unfortunately, due to the critical function of 7SK RNA in regulating transcription, trying to manipulate 7SK RNP components in human cells is not only incredibly difficult, but hard to interpret (see Figures 2.10 and 2.11). Therefore, it will be beneficial to create a 7SK-like system *in vivo* using a non-metazoan eukaryote model. I propose using the budding yeast *Saccharomyces cerevisiae* due to the ease of genetic manipulations [359], conserved transcription system (albeit without promoter proximal pausing and having minimal alternative splicing) [360], and endogenous expression of a P-TEFb-like kinase (Bur1 and Bur2) [304].

It is important to note that the introduction of the 7SK RNA system in yeast is not to perfectly recreate promoter proximal pausing. While it was once thought that P-TEFb performed the functions of both Ctk1 and Bur1 (yeast transcriptional kinases), recent evidence claims that Ctk1 in yeast is orthologous to CDK12 (reviewed in [361]), suggesting that Bur1 is the true ortholog of CDK9. While Bur1 and CDK9 only share 43% sequence identity [362], the functional conservation of phosphorylating DSIF (Spt4/Spt5) would suggest that a 7SK RNP-like system could target Bur1/Bur2. Therefore, the goal is to create a Bur1/Bur2 sequestration system using 7SK RNA that would mimic P-TEFb repression and activation (via transitioning from one 7SK RNP to another).

Creating a yeast model to systematically examine the effects of protein binding or 7SK RNA structure on transcription regulation will be a large undertaking. One must introduce within the genome, at a minimum, a copy of 7SK RNA, LARP7, MePCE, HEXIM1, and, potentially, NELF components, as *in vitro* studies suggest efficient pausing is only accomplished when both DSIF and NELF are present [363]. Yeast encode orthologous subunits of P-TEFb (Bur1/Bur2) [304] and DSIF (Spt4 and Spt5) [69]; however, if HEXIM cannot interact with Bur1/Bur2 – as

determined through *in vitro* binding and kinase assays, CDK9 and CycT1 will also be introduced into the genome. All ectopic components should be expressed under an inducible promoter so that one could turn on and off the system in the case of lethality. If expression of the 7SK-P-TEFb RNP is lethal, it may provide an avenue to study in detail how promoter proximal pausing evolved. One may instead find the closest relative to unicellular eukaryotes with the 7SK RNP system and express those components. Then the system can be slowly built towards mammalian origins through addition of increasingly complex components.

After successful creation of the 7SK RNP expression system, initial exploration could leverage the power of genomic studies, including ChIP-seq [364], GRO-seq [106], and RNA-seq [365], to examine how the metazoan 7SK RNA affects global yeast transcription. Without expression of NELF and establishment of promoter proximal pausing, sequestration of Bur1/Bur2 should present phenotypes analogous to *bur1Δ*/*bur1-ts* or chemical inhibition [366, 367]. Once initial studies have mapped the consequences of expressing the 7SK-P-TEFb RNP, then one could introduce mutated 7SK RNA to explore the contribution of each RNA element and how it regulates Bur1/Bur2 activity. Finally, this system can also be used to examine what protein factors are necessary to induce dissociation. Initial studies could examine the efficacy of existing yeast orthologs (e.g., Bdf1 (human Brd4 [368])) in dissociating the 7SK-P-TEFb RNP *in vivo* before undertaking expression of metazoan-specific transcription factors.

While alternative splicing is not as robust in yeast as it is in metazoans [369], and yeast do not express orthologs of SRSF2, hnRNP A1, hnRNP Q, hnRNP R, or RNA helicase A, yeast do encode an ortholog of hnRNP K (PBP2 [370]). Therefore, it will be interesting to see if any other yeast splicing regulators or RNA-processing machinery will bind to 7SK RNA and form an alternative 7SK RNP when Bur1/Bur2 is released. If indeed we find that specific RNA-binding

proteins are necessary for the proper function of 7SK RNA (transitioning between alternative RNPs to regulate active Bur1/Bur2 levels), one may need to add some or all of the hnRNP RNP particles into genomic loci under inducible promoters as well. With expression of a complete *in vivo* artificial 7SK RNP system, we will be able to systematically examine the relative importance of every 7SK RNP, how individual protein components cross-talk to modulate active Bur1/Bur2 levels, how separate functions of hnRNPs aid in 7SK RNP transitioning, and begin to address whether 7SK RNA itself may modulate the diverse functions of distinct hnRNPs.

Chapter 3 explores the possibility that hnRNP J mediates transcription termination via its phosphorylation of S261 by P-TEFb at specific genes. The results are preliminary, but future endeavors can take advantage of human cell lines to thoroughly explore this avenue. Initial studies can investigate if hnRNP J directly interacts with XRN2 and the subset of genes in which this regulation may take place. Additionally, my results suggest that S261 helps mediates this interaction; however, it is unclear if this interaction is dependent on phosphorylation of S261, or if the phosphorylation status of S261 modulates PTM placement on adjacent residues. If hnRNP J does regulate termination, it raises several intriguing questions about the purpose of the hnRNP K isoforms. Why is hnRNP J expressed at such low abundance? What does it do? Could it be specific to regulation with 7SK RNA and termination of distinct genes? Is hnRNP J strictly nuclear, and if so, what prevents utilization of the KNS? Together with the results from Chapter 3, the data gleaned from these experiments will add much-needed depth towards understanding the intricate roles of the hnRNP K isoforms in regulating the basic cellular process of transcription and gene expression.

4.1.4 Concluding remarks

Ultimately, to understand 7SK RNA and its functions is to understand P-TEFb regulation and promoter proximal pausing. I have taken a unique, and quite often overlooked, perspective towards this goal by devoting my investigations into the 7SK-hnRNP RNPs. I have provided the first direct biochemical examination of hnRNPs A1 and K/J binding to 7SK RNA. I found that their binding is not merely stochastic, but rather a complex interplay between the competing factors. I have also explored the functional consequences of these *in vitro* observations in human cells. I have validated [204] that Stem III helps regulate P-TEFb release upon transcriptional stress, and that either protein binding, RNA restructuring, or both are necessary for this regulation (Chapter 2). I have also found that hnRNP J, a splice variant of hnRNP K, may also [313] contribute to proper termination, and this function may be regulated, in part, through P-TEFb-dependent phosphorylation (Chapter 3).

The current working model for 7SK RNA-mediated transcription regulation attaches little importance to the 7SK-hnRNP RNPs. While several lines of evidence (including data presented within this thesis) have shown that disruption of hnRNP binding, either through decreased expression of the proteins [203, 263] or large-scale deletions of portions of the RNA itself [204], affects promoter proximal pausing, few models have integrated these impacts on transcriptional regulation. At best, the model stops short at describing how low-levels of transcription (i.e, transcriptional arrest) leads to increased concentrations of hnRNPs not bound to nascent transcripts, and thus they will bind the abundant nuclear transcript of 7SK RNA. These binding events, in turn, lead to “trapping” of 7SK RNA into the 7SK-hnRNP RNPs, preventing P-TEFb association and transcriptional repression [151].

While this accurately describes the initial phase of 7SK-hnRNP RNP formation, it fails to address later events: namely the transition from the 7SK-hnRNP RNPs to the 7SK-P-TEFb RNP. I have found that the competition between RNA-binding proteins, aided in part through post-translational modifications, plays an integral role in maintaining the formation of RNPs (such as "trapping" the RNA at sufficiently high concentrations, see Chapter 2), and these competitions form the mechanistic basis by which they transition (Figure 4.1). I propose that the formation of the 7SK-hnRNP RNPs is not simply a mechanism by which the cell promotes P-TEFb release and activity. Rather, it is a sort of rheostat to measure transcriptional activity of a cell. When transcriptional activity is low, hnRNPs bind to 7SK RNA, preventing P-TEFb association and thus promoting transcription. However, as transcriptional activity increases, the formation of new transcripts provides increased binding substrates for the hnRNPs sequestered by 7SK RNA. As the hnRNPs dissociate to bind these targets, the local concentration of hnRNPs decreases. This allows for the reassembly of the 7SK-P-TEFb RNP, again effectively lowering transcriptional output. Once tapered, the system is ready to receive new signals to ramp up transcription upon stimulation.

Clearly, more work is required to elucidate the details underlying this model – both "big picture" (Sections 4.1.2 and 4.1.3) experiments and "small-scale" experiments to finish Chapters 2 and 3. However, I hope the data presented here can be used as a framework and springboard to build a holistic understanding of 7SK RNA, P-TEFb regulation, promoter proximal pausing, and, ultimately, transcription regulation in metazoans.

APPENDIX A

PRIMERS USED IN THESIS

A.1 PRIMERS FOR HNRNP A1 VS SRSF2 STUDY

Table A.1. List of primers used in the hnRNP A1 and SRSF2 study.

Name	Sequence (5' - 3')	Purpose
Stem III For	CGACTCACTATAGGAGCTGCGCTCCCC	Generation of Stem III DNA template for transcription reaction
Stem III Rev	AGCTTGACTACCCTACGTTCTCCTAC	
Stem III T7	GCGCGCGAATTCTAATACGACTCACTATAGGAGCTGCG	
Stem III HDV Overlap	GGTGGAGATGCCATGCCGACCCAGCTTGACTACCCTACGTTCTC	
Stem III-SHAPE For	GGCCTTCGGGCCAAAGCTGCGCTCCCC	Generation of DMS probing-accessible Stem III DNA template for transcription reaction
Stem III-SHAPE Rev	GGCGAACCGGATCGAAGCTTGACTACCCTACGTTCTC	
Stem III-SHAPE T7	TAATACGACTCACTATAGGCCTTCGGGCCAA	
Stem III-SHAPE 3' HP	CGAAGCCCGATTTGGATCCGGCGAACCGGATCGA	
Stem III-SHAPE RT HP	GAACCGGACCGAAGCCCGATTTGGATCC	
SHAPE RT Primer	GAACCGGACCGAAGCCCG	Primer for reverse transcription
Stem III WT For	AGCTGCGCTCCCCTGCTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTG	Generation of mutant Stem III DNA templates for transcription
Stem III WT Rev	AGCTTGACTACCCTACGTTCTCCTACAAATGGACCTTGAGAGCTTG	

Stem III C209A For	AGCTGCGCTACCCTGCTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTG	reaction
Stem III C211G For	CGACTCACTATAGGAGCTGCGCTCCGCTGC	
Stem III DHM For	AGCTGCGCTCCAGTTCTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTG	
Stem III DHM Rev	AGCTTGACTGCAGTCCGTTCTCCTACAAATGGACCTTGAGAGCTTG	
Stem III C255G For	GCT GCG CTC CCC TGC TAG AAC CTC GAA AC	
Stem III Δ227 For	AGCTGCGCTCCCCTGCTAGAACCTCCAACAAGCTCTCAAGGTCCATTTG	
Stem III Δ227-8 For	AGCTGCGCTCCCCTGCTAGAACCTCCACAAGCTCTCAAGGTCCATTTG	
Stem III A251U Rev	AGCTTGACTACCCTACGTTCTCCAACAAATGGACCTTGAGAGCTTG	
Stem III Δ251 Rev	AGCTTGACTACCCTACGTTCTCCACAAATGGACCTTGAGAGCTTG	
Stem III A261U Rev	AGCTTGACTACCCAACGTTCTCCTACAAATGGACCTTGAGAGCTTG	
Stem III Δ261 Rev	AGCTTGACTACCCACGTTCTCCTACAAATGGACCTTGAGAGCTTG	
Stem III U265C Rev	AGCTTGACTGCCCTACGTTCTCCTACAAATGGACCTTGAGAGCTTG	
UP1 For	AGCTAGCATGTCTAAGTCAGAGTCTCCTAAAGAG	Cloning UP1 into pET28a
UP1 Rev	AATTGGATCCTTATCGACCTCTTTGGCTGG	
UP1 F17A For	GAACAGCTGAGGAAGCTCGCCATTGGAGGGTTGAGCTT	Generation of RNA-binding null mutants via site directed mutagenesis
UP1 F17A Rev	AAGCTCAACCCTCCAATGGCGAGCTTCCTCAGCTGTTC	
UP1 F108A For	CAGGTGCCCACTTAACTGTGAAAAAGATAGCTGTTGGTGGCATTAA	
UP1 F108A Rev	TTAATGCCACCAACAGCTATCTTTTTCACAGTTAAGTGGGCACCTG	
$SRSF2_{RRM}$ For	AAAGCTAGCATGAGCTACGGCCGC	Cloning $SRSF2_{RRM}$ into pET28a
$SRSF2_{RRM}$ Rev	AAAAGCGGCCGCTTAGCTGTGGTGTGAGTCCG	
GB1-SR Y44A For	ATCCCGCGGGATGGCCACGTCGCCGACG	Generation of RNA-binding null mutant via site directed mutagenesis
GB1-SR Y44A Rev	CGTCGGCGACGTGGCCATCCCGCGGGAT	
fl7SK For	CGACTCACTATAGGATGTGAGGGCGATC	RT-PCR for full length 7SK
fl7SK Rev	TTTGGATGTGTCTGGAGTCTTG	
7SK For	TCGGTCAAGGGTATACGAGTAG	RT-qPCR for 7SK RNA (NR_001445.2)
7SK Rev	TTTGGATGTGTCTGGAGTCTTG	
U2 For	CGTCCTCTATCCGAGGACAATA	RT-qPCR for U2 RNA (K02847.1)
U2 Rev	GTACTGCAATACCAGGTCGATG	
TBP For	CCTGCCGATAACTATCATCTGGC	RT-qPCR for

TBP Rev	GTTTCCACGGATGCTTTCTCG	TATA binding protein mRNA (NM_138572)
Actin Promoter For	CTCGCCTTTGCCGATCC	RT-qPCR for β actin (NM_001101.3)
Actin Promoter Rev	ATGCCGGAGCCGTTGTC	
Actin Body For	TGCCCTGAGGCACTCTT	RT-qPCR for β actin (NM_001101.3)
Actin Body Rev	GATGTCCACGTCACACTTCAT	
HSP Promoter For	CTGAGGAGCTGCTGCGA	RT-qPCR for heat shock protein 70 (NM_005346.4)
HSP Promoter Rev	CACAGGTTCGCTCTGGAAAG	
HSP Body For	ATCCTGATGGGGGACAAGTCCGAGAAC	RT-qPCR for heat shock protein 70 [123]
HSP Body Rev	CGGGTTGGTTGTCGGAGTAGGTGG	
Greb Promoter For	ACTTTGATGCCCATAGGAAGG	RT-qPCR for growth regulation by estrogen in breast cancer 1 (NM_033090.2)
Greb Promoter Rev	GCTCAGCAGAGACGAAGAAA	
Greb Body For	GCCATGGGAACTTCCCTTACCTCTG	RT-qPCR for growth regulation by estrogen in breast cancer 1 [123]
Greb Body Rev	CCAGCTGGACCAGGTAGTAGACGGT	

A.2 PRIMERS FOR HNRNP K VS HEXIM STUDY

Table A.2. List of primers used in the hnRNP K and HEXIM study.

Name	**Sequence (5' - 3')**	**Purpose**
5' Gen AAAbottom	GCGCGCGAATTCTAATACGACTCACTATAGG AAAGAAAGGCAGACTGCCACATGCAGCGCCTCATT TCCTATAGTGAGTCGTATT	Generation of Stem IV DNA template for transcription reaction
Stem IV HDV Overlap	GGTGGAGATGCCATGCCGACCCAAAGAAAGGCAGA CTGCCACATGC	
Stem I For	CGACTCACTATAGGAUGUGAGGGCGAUCUGG	Generation of Stem I DNA template for transcription reaction
Stem I Rev	GGAGCGGTGAGGGAGGAAGGG	
Stem I T7	GCGCGCGAATTCTAATACGACTCACTATAGGAUGU GAGGG	

hnRNP K For	ATATAGCTAGCATGGAAACTGAACAGCCAGAAG	Cloning hnRNP K into pHMG6
hnRNP K Rev	TAGAATTCTTAGAATCCTTCAACATCTGCATAC	
hnRNP J For	ATCCCTACCTTGGAAGAGTACCAACACTATAAAGGAAG	Generation of hnRNP J from hnRNP K via site directed mutagenesis
hnRNP J Rev	CTTCCTTTATAGTGTTGGTACTCTTCCAAGGTAGGGAT	
S193A For	CACGTCCTTTGATGGGAGCCTCAGATATAAGATCAAG	Generation of S to A or S to D mutants of hnRNP J via site directed mutagenesis
S193A Rev	CTTGATCTTATATCTGAGGCTCCCATCAAAGGACGTG	
S261A For	GGTCCTCGACGAGGGGCCATATCATCATAATCTCTTCTAGA	
S261A Rev	TCTAGAAGAGATTATGATGATATGGCCCCTCGTCGAGGACC	
S261D For	GATTATGATGATATGGATCCTCGTCGAGGACC	
S261D Rev	GGTCCTCGACGAGGATCCATATCATCATAATC	
Myc J For	AAAAAAGCGATCGCATGGAAACTGAACAGCCAGAAG	Cloning hnRNP J into pCMV6
Myc J Rev	AAAAGCGGCCGCGTGAATCCTTCAACATCTGCATACTGC	
MYC B PAS For	AAGTACATTTTGCTTTTTAAAGTTGATT	RT-qPCR for MYC [305]
MYC B PAS Rev	GGCTCAATGATATATTTGCCAGTTATTTTA	
MYC A PAS For	ATCGGGAAGGTGTTAGTCTGAATC	
MYC A PAS Rev	CACTCTCTCCTATTCTGAGGGCTT	
FRAT2 B PAS For	GTTCAAGGTCACGGTTTGCT	RT-qPCR for frequently rearranged in advanced T-cell lymphomas 2 [305]
FRAT2 B PAS Rev	CAACAGGGCTCTTCTTGGAG	
FRAT2 A PAS For	GAGGTGGGTTTTCATCTGGA	
FRAT2 A PAS Rev	CCAAGGCCATAGCTCAAGAC	
CCNB1 B PAS For	GTCAAGAACAAGTATGCCAC	RT-qPCR for cyclin B1 [305]
CCNB1 B PAS Rev	CACCTTTGCCACAGCCTTGG	
CCNB1 A PAS For	GACTCCAACTGGCAAAAGAT	
CCNB1 A PAS Rev	AGATCATGCCACTGCACTTC	

APPENDIX B

MEASURING THE AFFINITY OF AN INTERACTION – BASICS OF MODELING BIOCHEMICAL ASSOCIATIONS THROUGH CURVE-FITTING DATA

Biology is built upon intra- and inter-molecular interactions. Thus, one of the most common goals of biochemistry and biological physics is to detail, both qualitatively and quantitatively, the direct interactions between and among biological macromolecules. At the simplest level, interaction between two molecules is the culmination of two intrinsic physical properties: association and dissociation. The rate at which two molecules perform these functions gives us fundamental, and most importantly, comparable, insights into their respective functions, and it allows researchers to understand the individual "cogs" that must work together to produce the complicated "machinery" that is the cell. This appendix, adapted from [371], will cover the basics of bi-molecular interactions, as well as examine the equations used to model the interactions contained within this thesis.

B.1 MATHMATICAL BASIS OF INTER-MOLECULAR INTERACTIONS

As the initial stepping stone in elucidating the complex interaction networks that drive biological processes, bi-molecular interactions are the fundamental units in which biochemists describe the natural world. The binding of two molecules is not a static event, and indeed is driven through the competing processes of association and dissociation. We can describe the strength, or affinity, of an interaction between two molecules only when the reaction is at equilibrium. At this point, there is no change in the concentrations of free and bound molecules, and therefore the *rates* of association and dissociation are *constant*. This is a critical consideration, as our description of a binding event is expressed as a ratio between these constants, and is thus itself a constant that can be related to other events.

B.1.1 The dissociation constant (K_d)

Binding constants are traditionally expressed as the dissociation constant (K_d) due to the ease of understanding its unit (molar). Formally, the K_d of a reaction is the ratio of the rate of dissociation (expressed as s^{-1}) over the rate of association (expressed as $M^{-1}s^{-1}$), and is thus inversely proportional to the affinity of an interaction. As the K_d increases, the affinity between two molecules decreases. Because we measure these reactions at equilibrium, we can directly relate the rate constants to the concentration of products (association) and reactants (dissociation). Thus, in considering the simple reaction of

$$(1)\ A + B = AB$$

we can express the K_d as:

$$(2)\ K_d = ([A]*[B]) / [AB]$$

It is exceedingly difficult to precisely determine the exact concentrations of individual reactants and products at the same time. We can therefore take advantage of the fact that the concentration of free A ([A]) and B ([B]) in a reaction is but a fraction of the total molecules, such that

$$(3)\ [A] = [A]_{total} - [AB] \text{ and } [B]_{total} = [B] + [AB]$$

These identities allow us to substitute terms in the above equation to yield a more usable, and experimentally approachable, equation to express the dissociation constant.

When substituted and rearranged, we can describe a reaction as:

$$(4)\ [AB] / [A]_{total} = [B] / ([B] + K_d)$$

We can now calculate the K_d by measuring product formation relative to the concentration of a single reactant. By dividing the right side of the equation by the K_d, we can more clearly see the relationship between the K_d, free B, and product formed. Now,

$$(5)\ [AB] / [A]_{total} = ([B] / K_d) / (1 + ([B] / K_d))$$

When the concentration of free B is equal to the K_d, the right side of the equation becomes 0.5. The left side of the equation describes the amount of complex formed relative to the total concentration of A, and is typically described as the fraction bound. Together, this equation states that the K_d, or dissociation constant, is the concentration of B required to bind half the molecules of A in a reaction at equilibrium. Plotting the concentration of B against the fraction bound yields a hyperbolic curve on a linear scale that is transformed into a sigmoidal curve on a logarithmic scale. The sigmoidal representation is especially useful in calculations as the K_d is the inflection point of the curve.

B.1.2 Describing cooperativity

Sometimes the plot of a binding curve between A and B results in a sigmoidal curve on a linear scale. How can we account for this discrepancy mathematically? First, it is important to understand what a sigmoidal curve implies on a molecular scale. Initially, as B is titrated into reactions, the rate of association contributes less than the rate of dissociation, such that few molecules of B are able to bind A, and this results in the lag observed at the beginning of the curve. However, as more B binds A, a change occurs that makes it easier for more molecules of B to bind. This relates to the linear portion of the sigmoidal curve. Eventually, as the amount of B added into the reaction saturates, any molecule of AB that dissociates will quickly be bound by free B, and the curve asymptotes.

This concept – that initial binding events mediates further binding – is the basis for cooperativity. Cooperativity, then, defines the situation where *multiple* B's bind a single A, and therefore we are now in the regime of

$$(6)\ A + nB = AB_n$$

where "n" denotes the number of molecules of B. It is important to note that to write equation (6), we assume that *every* molecule of A is bound by n number of B's. Mathematically, we can add a term to the K_d equation that will transform the hyperbolic curve to a sigmoidal curve. The term, called the Hill coefficient, describes the slope of the linear portion of the sigmoidal curve when converted and rearranged into a logarithmic expression. When introduced in the linear expression, the term produces equation (7).

$$(7)\ [AB] / [A]_{total} = ([B] / K_d)^n / (1 + ([B] / Kd)^n)$$

Here, the "n" denotes the Hill coefficient. When n = 1, the equation will describe a hyperbolic curve, and thus indicates that there is no cooperativity in the system. As the degree of

cooperativity increases (more molecules of B are required to bind A), n becomes greater than 1, and the equation becomes more sigmoidal. This is known as positive cooperativity, as the first binding events of B help enhance the binding events of other molecules of B. While negative cooperativity, where $n < 1$, exists, it is incredibly rare in biology, and will thus not be discussed.

As mentioned in the preceding paragraph, use of the Hill coefficient assumes perfect cooperativity. However, this is never observed experimentally, and thus it is important to note the limitation of the Hill coefficient. It is a powerful term for describing the *degree* to which a system is cooperative. It cannot, however, fully describe a binding event: such as the strength of one event relative to the next, the order in which molecules bind, or even the number of molecules that are interacting within the system. Therefore, one must interpret with caution any information gained from the Hill coefficient.

B.1.3 How to describe the competition between two molecules binding another

It is indeed a rare event in biology where two molecules *only* interact with each other. Quite frequently, a single molecule will be subjected to a myriad of interactions between various molecular partners. Sometimes, these interactions are mutually exclusive, and we need to create a way to quantitatively compare these situations. Let us consider the interaction between A, B and C such that

$$A + B = AB \text{ and } A + C = AC \tag{8}$$

Note that in this scenario, B and C do not interact, so that

$$A + B + C = AB + AC \tag{9}$$

We can begin to understand this reaction by measuring the dissociation constant for each individual binding event to A. While this allows us to compare the *strength* of B and C binding

to A, it does not explain how these molecules behave when present in the same reaction. For that, we need a term that reflects the ability of B to compete with C for binding A.

If we again examine equation (8), we see that B and C must compete for binding to the same site on A, such that A can only form a complex with B or with C. Therefore, if we were to only look at the formation of AB, we can consider C to be an inhibitor of the reaction. As we add increasing amounts of C into reactions containing pre-formed AB complexes, C will compete with B upon dissociation to form the AC complex. At saturating concentrations of C, nearly every molecule of A will bind to C, and thus little-to-no AB complexes will be observed. This rationale forms the basis of the inhibitor concentration, or IC_{50}, which is formally defined as the concentration of C necessary to dissociate half of the pre-formed AB complex in a reaction. Expressed mathematically, the IC_{50} of a reaction may appear to be the same as the K_d; however, it is imperative to note that the IC_{50} **is not** the K_d of the AC reaction. As long as the concentration of B in the reaction is vastly greater than the K_d of the AB reaction, the following equation expresses the relationship between the concentration of C and the IC_{50}:

$$(10)\ [AB] / [A]_{total} = 1 - ([C] / ([C] + IC_{50}))$$

When the concentration of C is plotted against the fraction of AB bound, this equation will result in a negative hyperbolic curve (linear scale) or negative sigmoidal curve (logarithmic scale).

The IC_{50} of a reaction describes more than just the competition between B and C for binding to A. It also relates the individual affinities of the AB and AC reactions through the following equation:

$$(11)\ IC_{50} = K_{d(AC)} (1 + ([B] / K_{d(AB)})$$

Therefore, competition reactions in conjunction with affinity reactions can allow researchers to determine the K_d for a reaction that is difficult to assess independently. Additionally, this

equation states that the competition between two molecules is related to the affinity of their interactions as well as the local concentration of the initial molecule ("B" in this case).

B.1.4 Experimental considerations for measuring affinity

Now that I have examined the basic theories behind bi-molecular interactions, I will discuss how to properly set up and measure affinities experimentally. When designing experiments to measure the K_d of the reaction between A and B, we must consider the variables needed to fulfill equation (4). First, we must be able to differentiate the formation of AB from free A. A routine method involves labeling A and separating reactions through a native matrix to allow visualization of the different complexes. In this system, the AB complex will migrate differently than free A, and thus both species are easily quantifiable, allowing for calculation of the fraction bound. Secondly, we must be able to precisely approximate free B in the reaction. To accomplish this, most reactions will contain concentrations of A orders of magnitude lower than the K_d. Therefore, the formation of AB complexes represents a negligible portion of the B molecules present within the reaction, such that $[B]_{total} \approx [B]$. We can then plot $[B]_{total}$ on the x-axis without having to directly measure the concentration of free B. This consideration is also true of IC_{50} reactions.

But how do we know that the $[B]_{total}$ is truly representative of every molecule capable of binding? During purification, some proportion of B molecules will suffer irreversible damage, and this error should be corrected prior to establishing affinity. If uncorrected, it is customary to denote the observed affinity as "$K_{d,app}$," or the apparent dissociation constant. To determine the true affinity of an interaction, we can utilize an activity assay. In this experiment, the concentration of A is held in excess over the K_d. Then, B is titrated into the reaction and the

formation of AB is measured. When the fraction bound is plotted against $[B]_{total}$, the binding curve will begin with a linear portion prior to approaching the asymptote at saturation. The curve is linear because the reaction was conducted under a regime in which any molecule of B capable of binding A, will. We can use the linear portion to calculate a linear regression and obtain the concentration of B required for full binding at saturation. If we divide $[A]_{total}$ by the concentration of B required to reach saturation, we can establish the proportion of molecules of B that are capable of binding, or the "activity" of the molecule. We can then correct the $[B]_{total}$ used in affinity experiments to reflect the true proportion of active molecules and calculate an accurate K_d.

B.2 MEASURING AFFINITY IN THIS THESIS

Unless otherwise noted (Sections 2.2.9 and 3.3.6), binding curves were generated in KaleidaGraph software v4.5.2. The fraction shifted is calculated from pixel density measurements obtained through software listed in the respective methods section. Note that "pixel density" implies software-subtracted background values and only corresponds to intensities visibly darker than background (thus, positive values). "Fraction shifted" is calculated by dividing all intensities above the unshifted control (unless otherwise noted) by the total density value obtained from shifted and unshifted values. The fraction shifted is plotted on the y-axis against the respective concentration of protein titrated into that reaction on the x-axis. Note that the protein concentration is corrected for activity (Section B.1.4).

Graphs are generated as scatter plots. Curve-fitting is performed through KaleidaGraph using user-submitted multi-variable equations. Before fitting, initial estimates of all variables are

supplied by the user. Optimized values are then output from the software that induces "best fit" curves, along with calculated error for each variable. Accepted values generally do not contain errors larger than 1. Note that data not approaching "1" on the y-axis does not generate well-fit curves through the program, underscoring the importance of testing orders of magnitude when titrating protein. Calculated K_d values are cross-checked to raw data to assure validity.

To calculate the K_d, the following equation was used:

$$(12)\ y = (x^{m2}/m1^{m2}) / (1 + (x^{m2}/m1^{m2}))$$

where y is the fraction shifted, x is the protein concentration, m1 is the K_d in μM, and m2 is the Hill coefficient. m1 and m2 estimates are supplied by the user, usually with values of 1 and 2, respectively. To calculate the IC_{50}, I used a standard sigmoidal curve-fit equation supplied by KaleidaGraph, as the mid-point for the curve is explicit:

$$(13)\ y = m1 + (m2-m1) / (1 + (x/m3)^{m4})$$

where y is the fraction of pre-bound protein shifted, x is the competing protein concentration, m1 is the Y_{min}, m2 is the Y_{max}, m3 is the x-value at the mid-point of the curve (IC_{50}), and m4 is the slope of the curve at the mid-point. Again, initial estimates are supplied by the user.

BIBLIOGRAPHY

[1] Morris KV, Mattick JS. The rise of regulatory RNA. Nature reviews Genetics. 2014;15:423-37.
[2] Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Molecular cell. 2012;45:439-46.
[3] Goodfellow SJ, Zomerdijk JC. Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. Subcell Biochem. 2013;61:211-36.
[4] Shandilya J, Roberts SG. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. Biochimica et biophysica acta. 2012;1819:391-400.
[5] Turowski TW, Tollervey D. Transcription by RNA polymerase III: insights into mechanism and regulation. Biochemical Society transactions. 2016;44:1367-75.
[6] Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, et al. Structure of eukaryotic RNA polymerases. Annual review of biophysics. 2008;37:337-52.
[7] Arimbasseri AG, Rijal K, Maraia RJ. Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation. Transcription. 2014;5:e27639.
[8] Fuda NJ, Ardehali MB, Lis JT. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature. 2009;461:186-92.
[9] Sainsbury S, Bernecky C, Cramer P. Structural basis of transcription initiation by RNA polymerase II. Nature reviews Molecular cell biology. 2015;16:129-43.
[10] Corden JL. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. Chemical reviews. 2013;113:8423-55.
[11] Allison LA, Wong JK, Fitzpatrick VD, Moyle M, Ingles CJ. The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. Molecular and cellular biology. 1988;8:321-9.
[12] Chapman RD, Conrad M, Eick D. Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation. Molecular and cellular biology. 2005;25:7665-74.
[13] Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, et al. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Molecular cell. 2012;45:158-70.
[14] Eick D, Geyer M. The RNA polymerase II carboxy-terminal domain (CTD) code. Chemical reviews. 2013;113:8456-90.
[15] Li B, Carey M, Workman JL. The role of chromatin during transcription. Cell. 2007;128:707-19.

[16] Muller F, Tora L. Chromatin and DNA sequences in defining promoters for transcription initiation. Biochimica et biophysica acta. 2014;1839:118-28.
[17] Weil PA, Luse DS, Segall J, Roeder RG. Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. Cell. 1979;18:469-84.
[18] Horikoshi M, Wang CK, Fujii H, Cromlish JA, Weil PA, Roeder RG. Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. Nature. 1989;341:299-303.
[19] Kuras L, Kosa P, Mencia M, Struhl K. TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo. Science. 2000;288:1244-8.
[20] Ha I, Lane WS, Reinberg D. Cloning of a human gene encoding the general transcription initiation factor IIB. Nature. 1991;352:689-95.
[21] Pinto I, Ware DE, Hampsey M. The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo. Cell. 1992;68:977-88.
[22] Imbalzano AN, Zaret KS, Kingston RE. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. The Journal of biological chemistry. 1994;269:8280-6.
[23] Kuras L, Struhl K. Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. Nature. 1999;399:609-13.
[24] Li XY, Virbasius A, Zhu X, Green MR. Enhancement of TBP binding by activators and general transcription factors. Nature. 1999;399:605-9.
[25] Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nature reviews Molecular cell biology. 2015;16:167-77.
[26] Allen BL, Taatjes DJ. The Mediator complex: a central integrator of transcription. Nature reviews Molecular cell biology. 2015;16:155-66.
[27] Wong KH, Jin Y, Struhl K. TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Molecular cell. 2014;54:601-12.
[28] Guzman E, Lis JT. Transcription factor TFIIH is required for promoter melting in vivo. Molecular and cellular biology. 1999;19:5652-8.
[29] Luse DS, Kochel T, Kuempel ED, Coppola JA, Cai H. Transcription initiation by RNA polymerase II in vitro. At least two nucleotides must be added to form a stable ternary complex. The Journal of biological chemistry. 1987;262:289-97.
[30] Liu X, Bushnell DA, Kornberg RD. RNA polymerase II transcription: structure and mechanism. Biochimica et biophysica acta. 2013;1829:2-8.
[31] Dvir A. Promoter escape by RNA polymerase II. Biochimica et biophysica acta. 2002;1577:208-23.
[32] Guo J, Price DH. RNA polymerase II transcription elongation control. Chemical reviews. 2013;113:8583-603.
[33] Ramanathan A, Robb GB, Chan SH. mRNA capping: biological functions and applications. Nucleic acids research. 2016;44:7511-26.
[34] Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annual review of biochemistry. 2015;84:291-323.
[35] Orphanides G, Reinberg D. RNA polymerase II elongation through chromatin. Nature. 2000;407:471-5.

[36] Skalska L, Beltran-Nebot M, Ule J, Jenner RG. Regulatory feedback from nascent RNA to chromatin and transcription. Nature reviews Molecular cell biology. 2017;18:331-7.
[37] Tanny JC. Chromatin modification by the RNA Polymerase II elongation complex. Transcription. 2014;5:e988093.
[38] Harlen KM, Churchman LS. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nature reviews Molecular cell biology. 2017;18:263-73.
[39] Jonkers I, Kwak H, Lis JT. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife. 2014;3:e02407.
[40] Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, et al. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome research. 2014;24:896-905.
[41] Fuchs G, Voichek Y, Benjamin S, Gilad S, Amit I, Oren M. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. Genome Biol. 2014;15:R69.
[42] Porrua O, Libri D. Transcription termination and the control of the transcriptome: why, where and how to stop. Nature reviews Molecular cell biology. 2015;16:190-202.
[43] West S, Gromak N, Proudfoot NJ. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature. 2004;432:522-5.
[44] Zhang Z, Fu J, Gilmour DS. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. Genes & development. 2005;19:1572-80.
[45] Zhang Z, Gilmour DS. Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Molecular cell. 2006;21:65-74.
[46] Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. Genome research. 2000;10:1001-10.
[47] Hu J, Lutz CS, Wilusz J, Tian B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. Rna. 2005;11:1485-93.
[48] Guhaniyogi J, Brewer G. Regulation of mRNA stability in mammalian cells. Gene. 2001;265:11-23.
[49] Xiang K, Tong L, Manley JL. Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. Molecular and cellular biology. 2014;34:1894-910.
[50] Luo W, Johnson AW, Bentley DL. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes & development. 2006;20:954-65.
[51] Jensen TH, Jacquier A, Libri D. Dealing with pervasive transcription. Molecular cell. 2013;52:473-84.
[52] Smolle M, Workman JL. Transcription-associated histone modifications and cryptic transcription. Biochimica et biophysica acta. 2013;1829:84-97.
[53] Plant KE, Dye MJ, Lafaille C, Proudfoot NJ. Strong polyadenylation and weak pausing combine to cause efficient termination of transcription in the human Ggamma-globin gene. Molecular and cellular biology. 2005;25:3276-85.
[54] Gromak N, West S, Proudfoot NJ. Pause sites promote transcriptional termination of mammalian RNA polymerase II. Molecular and cellular biology. 2006;26:3986-96.

[55] Nag A, Narsinh K, Martinson HG. The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nature structural & molecular biology. 2007;14:662-9.
[56] Bentley DL, Groudine M. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature. 1986;321:702-6.
[57] Rougvie AE, Lis JT. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell. 1988;54:795-804.
[58] Rougvie AE, Lis JT. Postinitiation transcriptional control in Drosophila melanogaster. Molecular and cellular biology. 1990;10:6041-5.
[59] Marshall NF, Price DH. Control of formation of two distinct classes of RNA polymerase II elongation complexes. Molecular and cellular biology. 1992;12:2078-90.
[60] Chodosh LA, Fire A, Samuels M, Sharp PA. 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. The Journal of biological chemistry. 1989;264:2250-7.
[61] Sehgal PB, Darnell JE, Jr., Tamm I. The inhibition by DRB (5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole) of hnRNA and mRNA production in HeLa cells. Cell. 1976;9:473-80.
[62] Izban MG, Luse DS. Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. Genes & development. 1991;5:683-96.
[63] Adamson TE, Price DH. Cotranscriptional processing of Drosophila histone mRNAs. Molecular and cellular biology. 2003;23:4046-55.
[64] Glover-Cutter K, Kim S, Espinosa J, Bentley DL. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nature structural & molecular biology. 2008;15:71-8.
[65] Gilchrist DA, Nechaev S, Lee C, Ghosh SK, Collins JB, Li L, et al. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes & development. 2008;22:1921-33.
[66] Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, et al. c-Myc regulates transcriptional pause release. Cell. 2010;141:432-45.
[67] Saponaro M, Kantidakis T, Mitter R, Kelly GP, Heron M, Williams H, et al. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. Cell. 2014;157:1037-49.
[68] Peng J, Liu M, Marion J, Zhu Y, Price DH. RNA polymerase II elongation control. Cold Spring Harb Symp Quant Biol. 1998;63:365-70.
[69] Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes & development. 1998;12:343-56.
[70] Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell. 1999;97:41-51.
[71] Yamaguchi Y, Wada T, Watanabe D, Takagi T, Hasegawa J, Handa H. Structure and function of the human transcription elongation factor DSIF. The Journal of biological chemistry. 1999;274:8085-92.

[72] Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. The EMBO journal. 2011;30:1302-10.
[73] Narita T, Yamaguchi Y, Yano K, Sugimoto S, Chanarat S, Wada T, et al. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Molecular and cellular biology. 2003;23:1863-73.
[74] Missra A, Gilmour DS. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. Proceedings of the National Academy of Sciences of the United States of America. 2010;107:11301-6.
[75] Kwak H, Lis JT. Control of transcriptional elongation. Annu Rev Genet. 2013;47:483-508.
[76] Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH. A highly purified RNA polymerase II elongation control system. The Journal of biological chemistry. 2001;276:42601-9.
[77] Yamaguchi Y, Shibata H, Handa H. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. Biochimica et biophysica acta. 2013;1829:98-104.
[78] Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, et al. RNA polymerase is poised for activation across the genome. Nature genetics. 2007;39:1507-11.
[79] Core LJ, Waterfall JJ, Gilchrist DA, Fargo DC, Kwak H, Adelman K, et al. Defining the status of RNA polymerase at promoters. Cell reports. 2012;2:1025-35.
[80] Hu X, Malik S, Negroiu CC, Hubbard K, Velalar CN, Hampton B, et al. A Mediator-responsive form of metazoan RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America. 2006;103:9506-11.
[81] Cheng B, Li T, Rahl PB, Adamson TE, Loudas NB, Guo J, et al. Functional association of Gdown1 with RNA polymerase II poised on human genes. Molecular cell. 2012;45:38-50.
[82] Li J, Gilmour DS. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. The EMBO journal. 2013;32:1829-41.
[83] Weber CM, Ramachandran S, Henikoff S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Molecular cell. 2014;53:819-30.
[84] Paparidis NF, Durvale MC, Canduri F. The emerging picture of CDK9/P-TEFb: more than 20 years of advances since PITALRE. Mol Biosyst. 2017;13:246-76.
[85] Peng J, Marshall NF, Price DH. Identification of a cyclin subunit required for the function of Drosophila P-TEFb. The Journal of biological chemistry. 1998;273:13855-60.
[86] Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell. 1998;92:451-62.
[87] Peng J, Zhu Y, Milton JT, Price DH. Identification of multiple cyclin subunits of human P-TEFb. Genes & development. 1998;12:755-62.
[88] Fu TJ, Peng J, Lee G, Price DH, Flores O. Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. The Journal of biological chemistry. 1999;274:34527-30.
[89] Kohoutek J, Blazek D. Cyclin K goes with Cdk12 and Cdk13. Cell Div. 2012;7:12.
[90] Kim JB, Sharp PA. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. The Journal of biological chemistry. 2001;276:12317-23.

[91] Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Molecular and cellular biology. 2004;24:787-95.
[92] Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Molecular cell. 2006;21:227-37.
[93] Egyhazi E, Ossoinak A, Pigon A, Holmgren C, Lee JM, Greenleaf AL. Phosphorylation dependence of the initiation of productive transcription of Balbiani ring 2 genes in living cells. Chromosoma. 1996;104:422-33.
[94] Jiang Y, Liu M, Spencer CA, Price DH. Involvement of transcription termination factor 2 in mitotic repression of transcription elongation. Molecular cell. 2004;14:375-85.
[95] Lin C, Smith ER, Takahashi H, Lai KC, Martin-Brown S, Florens L, et al. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Molecular cell. 2010;37:429-37.
[96] Luo Z, Lin C, Shilatifard A. The super elongation complex (SEC) family in transcriptional control. Nature reviews Molecular cell biology. 2012;13:543-7.
[97] Shore SM, Byers SA, Maury W, Price DH. Identification of a novel isoform of Cdk9. Gene. 2003;307:175-82.
[98] Baumli S, Lolli G, Lowe ED, Troiani S, Rusconi L, Bullock AN, et al. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. The EMBO journal. 2008;27:1907-18.
[99] Marshall NF, Peng J, Xie Z, Price DH. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. The Journal of biological chemistry. 1996;271:27176-83.
[100] Hsin JP, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes & development. 2012;26:2119-37.
[101] Bartkowiak B, Liu P, Phatnani HP, Fuda NJ, Cooper JJ, Price DH, et al. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes & development. 2010;24:2303-16.
[102] Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, et al. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes & development. 2011;25:2158-72.
[103] Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. Genes & development. 2006;20:2922-36.
[104] Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. Nature reviews Molecular cell biology. 2015;16:178-89.
[105] Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature reviews Genetics. 2012;13:720-31.
[106] Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008;322:1845-8.
[107] Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science. 2010;327:335-8.

[108] Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes & development. 2011;25:742-54.
[109] Kwak H, Fuda NJ, Core LJ, Lis JT. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science. 2013;339:950-3.
[110] Henriques T, Gilchrist DA, Nechaev S, Bern M, Muse GW, Burkholder A, et al. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. Molecular cell. 2013;52:517-28.
[111] Boettiger AN, Levine M. Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science. 2009;325:471-3.
[112] Lagha M, Bothma JP, Esposito E, Ng S, Stefanik L, Tsui C, et al. Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. Cell. 2013;153:976-87.
[113] Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao Y, Li L, et al. Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell. 2010;143:540-51.
[114] Flynn RA, Almada AE, Zamudio JR, Sharp PA. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:10460-5.
[115] Marshall NF, Price DH. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. The Journal of biological chemistry. 1995;270:12335-8.
[116] McNamara RP, McCann JL, Gudipaty SA, D'Orso I. Transcription factors mediate the enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. Cell reports. 2013;5:1256-68.
[117] Jang MK, Mochizuki K, Zhou M, Jeong HS, Brady JN, Ozato K. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Molecular cell. 2005;19:523-34.
[118] Yang Z, Yik JH, Chen R, He N, Jang MK, Ozato K, et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Molecular cell. 2005;19:535-45.
[119] Liu W, Ma Q, Wong K, Li W, Ohgi K, Zhang J, et al. Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell. 2013;155:1581-95.
[120] Barboric M, Nissen RM, Kanazawa S, Jabrane-Ferrat N, Peterlin BM. NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. Molecular cell. 2001;8:327-37.
[121] Simone C, Stiegler P, Bagella L, Pucci B, Bellan C, De Falco G, et al. Activation of MyoD-dependent transcription by cdk9/cyclin T2. Oncogene. 2002;21:4137-48.
[122] Eberhardy SR, Farnham PJ. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. The Journal of biological chemistry. 2002;277:40156-62.
[123] McNamara RP, Reeder JE, McMillan EA, Bacon CW, McCann JL, D'Orso I. KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Molecular cell. 2016;61:39-53.
[124] Kanazawa S, Okamoto T, Peterlin BM. Tat competes with CIITA for the binding to P-TEFb and blocks the expression of MHC class II genes in HIV infection. Immunity. 2000;12:61-70.

[125] Gomes NP, Bjerke G, Llorente B, Szostek SA, Emerson BM, Espinosa JM. Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. Genes & development. 2006;20:601-12.
[126] Kurosu T, Peterlin BM. VP16 and ubiquitin; binding of P-TEFb via its activation domain and ubiquitin facilitates elongation of transcription of target genes. Current biology : CB. 2004;14:1112-6.
[127] Bisgrove DA, Mahmoudi T, Henklein P, Verdin E. Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. Proceedings of the National Academy of Sciences of the United States of America. 2007;104:13690-5.
[128] Schulte A, Czudnochowski N, Barboric M, Schonichen A, Blazek D, Peterlin BM, et al. Identification of a cyclin T-binding domain in Hexim1 and biochemical analysis of its binding competition with HIV-1 Tat. The Journal of biological chemistry. 2005;280:24968-77.
[129] AJ CQ, Bugai A, Barboric M. Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. Nucleic acids research. 2016;44:7527-39.
[130] Taube R, Lin X, Irwin D, Fujinaga K, Peterlin BM. Interaction between P-TEFb and the C-terminal domain of RNA polymerase II activates transcriptional elongation from sites upstream or downstream of target genes. Molecular and cellular biology. 2002;22:321-31.
[131] Dey A, Chitsaz F, Abbasi A, Misteli T, Ozato K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. Proceedings of the National Academy of Sciences of the United States of America. 2003;100:8758-63.
[132] Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, et al. Enhancer loops appear stable during development and are associated with paused polymerase. Nature. 2014;512:96-100.
[133] Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature. 2013;494:497-501.
[134] Takahashi H, Parmely TJ, Sato S, Tomomori-Sato C, Banks CA, Kong SE, et al. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell. 2011;146:92-104.
[135] Shore SM, Byers SA, Dent P, Price DH. Characterization of Cdk9(55) and differential regulation of two Cdk9 isoforms. Gene. 2005;350:51-8.
[136] Herrmann CH, Carroll RG, Wei P, Jones KA, Rice AP. Tat-associated kinase, TAK, activity is regulated by distinct mechanisms in peripheral blood lymphocytes and promonocytic cell lines. Journal of virology. 1998;72:9881-8.
[137] Cho S, Schroeder S, Kaehlcke K, Kwon HS, Pedal A, Herker E, et al. Acetylation of cyclin T1 regulates the equilibrium between active and inactive P-TEFb in cells. The EMBO journal. 2009;28:1407-17.
[138] Larochelle S, Merrick KA, Terret ME, Wohlbold L, Barboza NM, Zhang C, et al. Requirements for Cdk7 in the assembly of Cdk1/cyclin B and activation of Cdk2 revealed by chemical genetics in human cells. Molecular cell. 2007;25:839-50.
[139] Chen R, Yang Z, Zhou Q. Phosphorylated positive transcription elongation factor b (P-TEFb) is tagged for inhibition through association with 7SK snRNA. The Journal of biological chemistry. 2004;279:4153-60.
[140] Gudipaty SA, McNamara RP, Morton EL, D'Orso I. PPM1G Binds 7SK RNA and Hexim1 To Block P-TEFb Assembly into the 7SK snRNP and Sustain Transcription Elongation. Molecular and cellular biology. 2015;35:3810-28.

[141] Chen R, Liu M, Li H, Xue Y, Ramey WN, He N, et al. PP2B and PP1alpha cooperatively disrupt 7SK snRNP to release P-TEFb for transcription in response to Ca2+ signaling. Genes & development. 2008;22:1356-68.
[142] Contreras X, Barboric M, Lenasi T, Peterlin BM. HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription. PLoS pathogens. 2007;3:1459-69.
[143] Kim YK, Mbonye U, Hokello J, Karn J. T-cell receptor signaling enhances transcriptional elongation from latent HIV proviruses by activating P-TEFb through an ERK-dependent pathway. Journal of molecular biology. 2011;410:896-916.
[144] Mbonye UR, Wang B, Gokulrangan G, Chance MR, Karn J. Phosphorylation of HEXIM1 at Tyr271 and Tyr274 Promotes Release of P-TEFb from the 7SK snRNP Complex and Enhances Proviral HIV Gene Expression. Proteomics. 2015;15:2078-86.
[145] Fujinaga K, Barboric M, Li Q, Luo Z, Price DH, Peterlin BM. PKC phosphorylates HEXIM1 and regulates P-TEFb activity. Nucleic acids research. 2012;40:9160-70.
[146] Nguyen VT, Kiss T, Michels AA, Bensaude O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature. 2001;414:322-5.
[147] Yang Z, Zhu Q, Luo K, Zhou Q. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature. 2001;414:317-22.
[148] Barboric M, Yik JH, Czudnochowski N, Yang Z, Chen R, Contreras X, et al. Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. Nucleic acids research. 2007;35:2003-12.
[149] Garriga J, Peng J, Parreno M, Price DH, Henderson EE, Grana X. Upregulation of cyclin T1/CDK9 complexes during T cell activation. Oncogene. 1998;17:3093-102.
[150] Haaland RE, Herrmann CH, Rice AP. Increased association of 7SK snRNA with Tat cofactor P-TEFb following activation of peripheral blood lymphocytes. Aids. 2003;17:2429-36.
[151] Diribarne G, Bensaude O. 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. RNA biology. 2009;6:122-8.
[152] Prasanth KV, Camiolo M, Chan G, Tripathi V, Denis L, Nakamura T, et al. Nuclear organization and dynamics of 7SK RNA in regulating gene expression. Molecular biology of the cell. 2010;21:4184-96.
[153] Murphy S, Altruda F, Ullu E, Tripodi M, Silengo L, Melli M. DNA sequences complementary to human 7 SK RNA show structural similarities to the short mobile elements of the mammalian genome. Journal of molecular biology. 1984;177:575-90.
[154] Humphries P, Russell SE, McWilliam P, McQuaid S, Pearson C, Humphries MM. Observations on the structure of two human 7SK pseudogenes and on homologous transcripts in vertebrate species. The Biochemical journal. 1987;245:281-4.
[155] Murphy S, Tripodi M, Melli M. A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes. Nucleic acids research. 1986;14:9243-60.
[156] Kleinert H, Bredow S, Benecke BJ. Expression of a human 7S K RNA gene in vivo requires a novel pol III upstream element. The EMBO journal. 1990;9:711-8.
[157] Boyd DC, Turner PC, Watkins NJ, Gerster T, Murphy S. Functional redundancy of promoter elements ensures efficient transcription of the human 7SK gene in vivo. Journal of molecular biology. 1995;253:677-90.
[158] Murphy S, Pierani A, Scheidereit C, Melli M, Roeder RG. Purified octamer binding transcription factors stimulate RNA polymerase III--mediated transcription of the 7SK RNA gene. Cell. 1989;59:1071-80.

[159] Yoon JB, Murphy S, Bai L, Wang Z, Roeder RG. Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. Molecular and cellular biology. 1995;15:2019-27.
[160] Schaub M, Myslinski E, Schuster C, Krol A, Carbon P. Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. The EMBO journal. 1997;16:173-81.
[161] Orioli A, Pascali C, Pagano A, Teichmann M, Dieci G. RNA polymerase III transcription control elements: themes and variations. Gene. 2012;493:185-94.
[162] Chambers JC, Kurilla MG, Keene JD. Association between the 7 S RNA and the lupus La protein varies among cell types. The Journal of biological chemistry. 1983;258:11438-41.
[163] Stefano JE. Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. Cell. 1984;36:145-54.
[164] Wassarman DA, Steitz JA. Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. Molecular and cellular biology. 1991;11:3432-45.
[165] Marz M, Donath A, Verstraete N, Nguyen VT, Stadler PF, Bensaude O. Evolution of 7SK RNA and its protein partners in metazoa. Molecular biology and evolution. 2009;26:2821-30.
[166] Peterlin BM, Brogie JE, Price DH. 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. Wiley interdisciplinary reviews RNA. 2012;3:92-103.
[167] Reddy R, Henning D, Subrahmanyam CS, Busch H. Primary and secondary structure of 7-3 (K) RNA of Novikoff hepatoma. The Journal of biological chemistry. 1984;259:12265-70.
[168] Sinha KM, Gu J, Chen Y, Reddy R. Adenylation of small RNAs in human cells. Development of a cell-free system for accurate adenylation on the 3'-end of human signal recognition particle RNA. The Journal of biological chemistry. 1998;273:6853-9.
[169] He N, Jahchan NS, Hong E, Li Q, Bayfield MA, Maraia RJ, et al. A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. Molecular cell. 2008;29:588-99.
[170] Krueger BJ, Jeronimo C, Roy BB, Bouchard A, Barrandon C, Byers SA, et al. LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic acids research. 2008;36:2219-29.
[171] Markert A, Grimm M, Martinez J, Wiesner J, Meyerhans A, Meyuhas O, et al. The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. EMBO reports. 2008;9:569-75.
[172] Bayfield MA, Yang R, Maraia RJ. Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). Biochimica et biophysica acta. 2010;1799:365-78.
[173] Jeronimo C, Forget D, Bouchard A, Li Q, Chua G, Poitras C, et al. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Molecular cell. 2007;27:262-74.
[174] Gupta S, Busch RK, Singh R, Reddy R. Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of gamma-monomethyl-GTP cap structure in 7SK and several other human small RNAs. The Journal of biological chemistry. 1990;265:19137-42.
[175] Shumyatsky GP, Tillib SV, Kramerov DA. B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end. Nucleic acids research. 1990;18:6347-51.

[176] Zhao Y, Karijolich J, Glaunsinger B, Zhou Q. Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress HIV-1 transcription and escape from latency. EMBO reports. 2016;17:1441-51.
[177] Xue Y, Yang Z, Chen R, Zhou Q. A capping-independent function of MePCE in stabilizing 7SK snRNA and facilitating the assembly of 7SK snRNP. Nucleic acids research. 2010;38:360-9.
[178] Muniz L, Egloff S, Kiss T. RNA elements directing in vivo assembly of the 7SK/MePCE/Larp7 transcriptional regulatory snRNP. Nucleic acids research. 2013;41:4686-98.
[179] Flynn RA, Do BT, Rubin AJ, Calo E, Lee B, Kuchelmeister H, et al. 7SK-BAF axis controls pervasive transcription at enhancers. Nature structural & molecular biology. 2016;23:231-8.
[180] Hainer SJ, Gu W, Carone BR, Landry BD, Rando OJ, Mello CC, et al. Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. Genes & development. 2015;29:362-78.
[181] Yik JH, Chen R, Nishimura R, Jennings JL, Link AJ, Zhou Q. Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. Molecular cell. 2003;12:971-82.
[182] Michels AA, Fraldi A, Li Q, Adamson TE, Bonnet F, Nguyen VT, et al. Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. The EMBO journal. 2004;23:2608-19.
[183] Yik JH, Chen R, Pezda AC, Zhou Q. Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. The Journal of biological chemistry. 2005;280:16368-76.
[184] Li Q, Price JP, Byers SA, Cheng D, Peng J, Price DH. Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. The Journal of biological chemistry. 2005;280:28819-26.
[185] Blazek D, Barboric M, Kohoutek J, Oven I, Peterlin BM. Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. Nucleic acids research. 2005;33:7000-10.
[186] Yik JH, Chen R, Pezda AC, Samford CS, Zhou Q. A human immunodeficiency virus type 1 Tat-like arginine-rich RNA-binding domain is essential for HEXIM1 to inhibit RNA polymerase II transcription through 7SK snRNA-mediated inactivation of P-TEFb. Molecular and cellular biology. 2004;24:5094-105.
[187] Barboric M, Kohoutek J, Price JP, Blazek D, Price DH, Peterlin BM. Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb. The EMBO journal. 2005;24:4291-303.
[188] Li Q, Cooper JJ, Altwerger GH, Feldkamp MD, Shea MA, Price DH. HEXIM1 is a promiscuous double-stranded RNA-binding protein and interacts with RNAs in addition to 7SK in cultured cells. Nucleic acids research. 2007;35:2503-12.
[189] Belanger F, Baigude H, Rana TM. U30 of 7SK RNA forms a specific photo-cross-link with Hexim1 in the context of both a minimal RNA-binding site and a fully reconstituted 7SK/Hexim1/P-TEFb ribonucleoprotein complex. Journal of molecular biology. 2009;386:1094-107.

[190] Czudnochowski N, Vollmuth F, Baumann S, Vogel-Bachmayr K, Geyer M. Specificity of Hexim1 and Hexim2 complex formation with cyclin T1/T2, importin alpha and 7SK snRNA. Journal of molecular biology. 2010;395:28-41.
[191] Lebars I, Martinez-Zapien D, Durand A, Coutant J, Kieffer B, Dock-Bregeon AC. HEXIM1 targets a repeated GAUC motif in the riboregulator of transcription 7SK and promotes base pair rearrangements. Nucleic acids research. 2010;38:7749-63.
[192] Muniz L, Egloff S, Ughy B, Jady BE, Kiss T. Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator Tat. PLoS pathogens. 2010;6:e1001152.
[193] Dames SA, Schonichen A, Schulte A, Barboric M, Peterlin BM, Grzesiek S, et al. Structure of the Cyclin T binding domain of Hexim1 and molecular basis for its recognition of P-TEFb. Proceedings of the National Academy of Sciences of the United States of America. 2007;104:14312-7.
[194] Bigalke JM, Dames SA, Blankenfeldt W, Grzesiek S, Geyer M. Structure and dynamics of a stabilized coiled-coil domain in the P-TEFb regulator Hexim1. Journal of molecular biology. 2011;414:639-53.
[195] Egloff S, Van Herreweghe E, Kiss T. Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. Molecular and cellular biology. 2006;26:630-42.
[196] Lu H, Li Z, Xue Y, Schulze-Gahmen U, Johnson JR, Krogan NJ, et al. AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation. Proceedings of the National Academy of Sciences of the United States of America. 2014;111:E15-24.
[197] Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell. 2013;153:855-68.
[198] Cherrier T, Le Douce V, Eilebrecht S, Riclet R, Marban C, Dequiedt F, et al. CTIP2 is a negative regulator of P-TEFb. Proceedings of the National Academy of Sciences of the United States of America. 2013;110:12655-60.
[199] Krueger BJ, Varzavand K, Cooper JJ, Price DH. The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. PloS one. 2010;5:e12335.
[200] Calo E, Flynn RA, Martin L, Spitale RC, Chang HY, Wysocka J. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature. 2015;518:249-53.
[201] Elagib KE, Rubinstein JD, Delehanty LL, Ngoh VS, Greer PA, Li S, et al. Calpain 2 activation of P-TEFb drives megakaryocyte morphogenesis and is disrupted by leukemogenic GATA1 mutation. Dev Cell. 2013;27:607-20.
[202] Barrandon C, Bonnet F, Nguyen VT, Labas V, Bensaude O. The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. Molecular and cellular biology. 2007;27:6996-7006.
[203] Hogg JR, Collins K. RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. Rna. 2007;13:868-80.
[204] Van Herreweghe E, Egloff S, Goiffon I, Jady BE, Froment C, Monsarrat B, et al. Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. The EMBO journal. 2007;26:3570-80.
[205] Chaudhury A, Chander P, Howe PH. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. Rna. 2010;16:1449-62.

[206] Barboric M, Lenasi T, Chen H, Johansen EB, Guo S, Peterlin BM. 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. Proceedings of the National Academy of Sciences of the United States of America. 2009;106:7798-803.
[207] Oqani RK, Lin T, Lee JE, Choi KM, Shin HY, Jin DI. P-TEFb Kinase Activity Is Essential for Global Transcription, Resumption of Meiosis and Embryonic Genome Activation in Pig. PloS one. 2016;11:e0152254.
[208] Ji X, Lu H, Zhou Q, Luo K. LARP7 suppresses P-TEFb activity to inhibit breast cancer progression and metastasis. eLife. 2014;3:e02907.
[209] Expert-Bezancon A, Sureau A, Durosay P, Salesse R, Groeneveld H, Lecaer JP, et al. hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. The Journal of biological chemistry. 2004;279:38249-59.
[210] Zhu J, Mayeda A, Krainer AR. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Molecular cell. 2001;8:1351-61.
[211] Zahler AM, Damgaard CK, Kjems J, Caputi M. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. The Journal of biological chemistry. 2004;279:10077-84.
[212] Hallay H, Locker N, Ayadi L, Ropers D, Guittet E, Branlant C. Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site. The Journal of biological chemistry. 2006;281:37159-74.
[213] Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. The Biochemical journal. 2009;417:15-27.
[214] Cazalla D, Zhu J, Manche L, Huber E, Krainer AR, Caceres JF. Nuclear export and retention signals in the RS domain of SR proteins. Molecular and cellular biology. 2002;22:6871-82.
[215] Bourgeois CF, Lejeune F, Stevenin J. Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. Prog Nucleic Acid Res Mol Biol. 2004;78:37-88.
[216] Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD. The splicing factor SC35 has an active role in transcriptional elongation. Nature structural & molecular biology. 2008;15:819-26.
[217] Shen H, Kan JL, Green MR. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Molecular cell. 2004;13:367-76.
[218] Wu JY, Maniatis T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell. 1993;75:1061-70.
[219] Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, et al. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. The EMBO journal. 1996;15:265-75.
[220] Wang HY, Lin W, Dyck JA, Yeakley JM, Songyang Z, Cantley LC, et al. SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. The Journal of cell biology. 1998;140:737-50.
[221] Cavaloc Y, Bourgeois CF, Kister L, Stevenin J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. Rna. 1999;5:468-83.

[222] Phelan MM, Goult BT, Clayton JC, Hautbergue GM, Wilson SA, Lian LY. The structure and selectivity of the SR protein SRSF2 RRM domain with RNA. Nucleic acids research. 2012;40:3232-44.
[223] Daubner GM, Clery A, Jayne S, Stevenin J, Allain FH. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. The EMBO journal. 2012;31:162-74.
[224] Biamonti G, Bassi MT, Cartegni L, Mechta F, Buvoli M, Cobianchi F, et al. Human hnRNP protein A1 gene expression. Structural and functional characterization of the promoter. Journal of molecular biology. 1993;230:77-89.
[225] Kamma H, Portman DS, Dreyfuss G. Cell type-specific expression of hnRNP proteins. Experimental cell research. 1995;221:187-96.
[226] Jean-Philippe J, Paz S, Caputi M. hnRNP A1: the Swiss army knife of gene expression. International journal of molecular sciences. 2013;14:18999-9024.
[227] He Y, Smith R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. Cellular and molecular life sciences : CMLS. 2009;66:1239-56.
[228] Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes & development. 1999;13:1102-15.
[229] Kiledjian M, Dreyfuss G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. The EMBO journal. 1992;11:2655-64.
[230] Cartegni L, Maconi M, Morandi E, Cobianchi F, Riva S, Biamonti G. hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. Journal of molecular biology. 1996;259:337-48.
[231] Siomi H, Dreyfuss G. A nuclear localization domain in the hnRNP A1 protein. The Journal of cell biology. 1995;129:551-60.
[232] Herrick G, Alberts B. Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. The Journal of biological chemistry. 1976;251:2124-32.
[233] Kumar A, Williams KR, Szer W. Purification and domain structure of core hnRNP proteins A1 and A2 and their relationship to single-stranded DNA-binding proteins. The Journal of biological chemistry. 1986;261:11266-73.
[234] Shamoo Y, Krueger U, Rice LM, Williams KR, Steitz TA. Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. Nature structural biology. 1997;4:215-22.
[235] Morgan CE, Meagher JL, Levengood JD, Delproposto J, Rollins C, Stuckey JA, et al. The First Crystal Structure of the UP1 Domain of hnRNP A1 Bound to RNA Reveals a New Look for an Old RNA Binding Protein. Journal of molecular biology. 2015;427:3241-57.
[236] Burd CG, Dreyfuss G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. The EMBO journal. 1994;13:1197-204.
[237] Mayeda A, Munroe SH, Xu RM, Krainer AR. Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. Rna. 1998;4:1111-23.
[238] Jain N, Lin HC, Morgan CE, Harris ME, Tolbert BS. Rules of RNA specificity of hnRNP A1 revealed by global and quantitative analysis of its affinity distribution. Proceedings of the National Academy of Sciences of the United States of America. 2017;114:2206-11.
[239] Schurer H, Lang K, Schuster J, Morl M. A universal method to produce in vitro transcripts with homogeneous 3' ends. Nucleic acids research. 2002;30:e56.

[240] Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). Journal of the American Chemical Society. 2005;127:4223-31.
[241] Abramochkin G, Shrader TE. The leucyl/phenylalanyl-tRNA-protein transferase. Overexpression and characterization of substrate recognition, domain structure, and secondary structure. The Journal of biological chemistry. 1995;270:20621-8.
[242] Zheng L, Baumann U, Reymond JL. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic acids research. 2004;32:e115.
[243] Laederach A, Das R, Vicens Q, Pearlman SM, Brenowitz M, Herschlag D, et al. Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. Nature protocols. 2008;3:1395-401.
[244] Bartholomeeusen K, Xiang Y, Fujinaga K, Peterlin BM. Bromodomain and extra-terminal (BET) bromodomain inhibition activate transcription via transient release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein. The Journal of biological chemistry. 2012;287:36609-16.
[245] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9:671-5.
[246] Hertel KJ, Graveley BR. RS domains contact the pre-mRNA throughout spliceosome assembly. Trends in biochemical sciences. 2005;30:115-8.
[247] Ziehler WA, Engelke DR. Probing RNA structure with chemical reagents and enzymes. Current protocols in nucleic acid chemistry / edited by Serge L Beaucage [et al]. 2001;Chapter 6:Unit 6 1.
[248] Regulski EE, Breaker RR. In-line probing analysis of riboswitches. Methods in molecular biology. 2008;419:53-67.
[249] Tijerina P, Mohr S, Russell R. DMS footprinting of structured RNAs and RNA-protein complexes. Nature protocols. 2007;2:2608-23.
[250] Wells SE, Hughes JM, Igel AH, Ares M, Jr. Use of dimethyl sulfate to probe RNA structure in vivo. Methods in enzymology. 2000;318:479-93.
[251] Moazed D, Noller HF. Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. Cell. 1986;47:985-94.
[252] Talkish J, May G, Lin Y, Woolford JL, Jr., McManus CJ. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. Rna. 2014.
[253] Merrill BM, Stone KL, Cobianchi F, Wilson SH, Williams KR. Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. The Journal of biological chemistry. 1988;263:3307-13.
[254] Robb GB, Brown KM, Khurana J, Rana TM. Specific and potent RNAi in the nucleus of human cells. Nature structural & molecular biology. 2005;12:133-7.
[255] Haaland RE, Herrmann CH, Rice AP. siRNA depletion of 7SK snRNA induces apoptosis but does not affect expression of the HIV-1 LTR or P-TEFb-dependent cellular genes. Journal of cellular physiology. 2005;205:463-70.
[256] Xiong Y, Yuan J, Zhang C, Zhu Y, Kuang X, Lan L, et al. The STAT3-regulated long non-coding RNA Lethe promote the HCV replication. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2015;72:165-71.

[257] Chao SH, Fujinaga K, Marion JE, Taube R, Sausville EA, Senderowicz AM, et al. Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. The Journal of biological chemistry. 2000;275:28345-8.
[258] Chao SH, Price DH. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. The Journal of biological chemistry. 2001;276:31793-9.
[259] Shen H, Green MR. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. Molecular cell. 2004;16:363-73.
[260] Okunola HL, Krainer AR. Cooperative-binding and splicing-repressive properties of hnRNP A1. Molecular and cellular biology. 2009;29:5620-31.
[261] Hanamura A, Caceres JF, Mayeda A, Franza BR, Jr., Krainer AR. Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. Rna. 1998;4:430-44.
[262] Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics. 2015.
[263] Lemieux B, Blanchette M, Monette A, Mouland AJ, Wellinger RJ, Chabot B. A Function for the hnRNP A1/A2 Proteins in Transcription Elongation. PloS one. 2015;10:e0126654.
[264] Shnyreva M, Schullery DS, Suzuki H, Higaki Y, Bomsztyk K. Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein. The Journal of biological chemistry. 2000;275:15498-503.
[265] Expert-Bezancon A, Le Caer JP, Marie J. Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken beta-tropomyosin pre-mRNA. The Journal of biological chemistry. 2002;277:16614-23.
[266] Tsai PL, Chiou NT, Kuss S, Garcia-Sastre A, Lynch KW, Fontoura BM. Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing. PLoS pathogens. 2013;9:e1003460.
[267] Michelotti EF, Michelotti GA, Aronsohn AI, Levens D. Heterogeneous nuclear ribonucleoprotein K is a transcription factor. Molecular and cellular biology. 1996;16:2350-60.
[268] Samuel SK, Spencer VA, Bajno L, Sun JM, Holth LT, Oesterreich S, et al. In situ cross-linking by cisplatin of nuclear matrix-bound transcription factors to nuclear DNA of human breast cancer cells. Cancer research. 1998;58:3004-8.
[269] Denisenko ON, Bomsztyk K. The product of the murine homolog of the Drosophila extra sex combs gene displays transcriptional repressor activity. Molecular and cellular biology. 1997;17:4707-17.
[270] Ostareck DH, Ostareck-Lederer A, Wilm M, Thiele BJ, Mann M, Hentze MW. mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. Cell. 1997;89:597-606.
[271] Pelisch F, Pozzi B, Risso G, Munoz MJ, Srebrow A. DNA damage-induced heterogeneous nuclear ribonucleoprotein K sumoylation regulates p53 transcriptional activation. The Journal of biological chemistry. 2012;287:30789-99.
[272] Lee SW, Lee MH, Park JH, Kang SH, Yoo HM, Ka SH, et al. SUMOylation of hnRNP-K is required for p53-mediated cell-cycle arrest in response to DNA damage. The EMBO journal. 2012;31:4441-52.
[273] Bomsztyk K, Denisenko O, Ostrowski J. hnRNP K: one protein multiple processes. BioEssays : news and reviews in molecular, cellular and developmental biology. 2004;26:629-38.

[274] Charroux B, Angelats C, Fasano L, Kerridge S, Vola C. The levels of the bancal product, a Drosophila homologue of vertebrate hnRNP K protein, affect cell proliferation and apoptosis in imaginal disc cells. Molecular and cellular biology. 1999;19:7846-56.
[275] Gates J, Thummel CS. An enhancer trap screen for ecdysone-inducible genes required for Drosophila adult leg morphogenesis. Genetics. 2000;156:1765-76.
[276] Piano F, Schetter AJ, Morton DG, Gunsalus KC, Reinke V, Kim SK, et al. Gene clustering based on RNAi phenotypes of ovary-enriched genes in C. elegans. Current biology : CB. 2002;12:1959-64.
[277] Gao FH, Wu YL, Zhao M, Liu CX, Wang LS, Chen GQ. Protein kinase C-delta mediates down-regulation of heterogeneous nuclear ribonucleoprotein K protein: involvement in apoptosis induction. Experimental cell research. 2009;315:3250-8.
[278] van Domselaar R, Quadir R, van der Made AM, Broekhuizen R, Bovenschen N. All human granzymes target hnRNP K that is essential for tumor cell viability. The Journal of biological chemistry. 2012;287:22854-64.
[279] Xiao Z, Ko HL, Goh EH, Wang B, Ren EC. hnRNP K suppresses apoptosis independent of p53 status by maintaining high levels of endogenous caspase inhibitors. Carcinogenesis. 2013;34:1458-67.
[280] Lu J, Gao FH. Role and molecular mechanism of heterogeneous nuclear ribonucleoprotein K in tumor development and progression. Biomed Rep. 2016;4:657-63.
[281] Braddock DT, Baber JL, Levens D, Clore GM. Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. The EMBO journal. 2002;21:3476-85.
[282] Ostrowski J, Van Seuningen I, Seger R, Rauch CT, Sleath PR, McMullen BA, et al. Purification, cloning, and expression of a murine phosphoprotein that binds the kappa B motif in vitro identifies it as the homolog of the human heterogeneous nuclear ribonucleoprotein K protein. Description of a novel DNA-dependent phosphorylation process. The Journal of biological chemistry. 1994;269:17626-34.
[283] Backe PH, Messias AC, Ravelli RB, Sattler M, Cusack S. X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. Structure. 2005;13:1055-67.
[284] Valverde R, Edwards L, Regan L. Structure and function of KH domains. The FEBS journal. 2008;275:2712-26.
[285] Makeyev AV, Liebhaber SA. The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. Rna. 2002;8:265-78.
[286] Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. Nature reviews Molecular cell biology. 2007;8:479-90.
[287] Paziewska A, Wyrwicz LS, Bujnicki JM, Bomsztyk K, Ostrowski J. Cooperative binding of the hnRNP K three KH domains to mRNA targets. FEBS letters. 2004;577:134-40.
[288] Thandapani P, O'Connor TR, Bailey TL, Richard S. Defining the RGG/RG motif. Molecular cell. 2013;50:613-23.
[289] Ozdilek BA, Thompson VF, Ahmed NS, White CI, Batey RT, Schwartz JC. Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. Nucleic acids research. 2017.
[290] Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, et al. The BioGRID interaction database: 2013 update. Nucleic acids research. 2013;41:D816-23.

[291] Van Seuningen I, Ostrowski J, Bustelo XR, Sleath PR, Bomsztyk K. The K protein domain that recruits the interleukin 1-responsive K protein kinase lies adjacent to a cluster of c-Src and Vav SH3-binding sites. Implications that K protein acts as a docking platform. The Journal of biological chemistry. 1995;270:26976-85.
[292] Michael WM, Eder PS, Dreyfuss G. The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. The EMBO journal. 1997;16:3587-98.
[293] Habelhah H, Shah K, Huang L, Ostareck-Lederer A, Burlingame AL, Shokat KM, et al. ERK phosphorylation drives cytoplasmic accumulation of hnRNP-K and inhibition of mRNA translation. Nature cell biology. 2001;3:325-30.
[294] Gumireddy K, Li A, Yan J, Setoyama T, Johannes GJ, Orom UA, et al. Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. The EMBO journal. 2013;32:2672-84.
[295] Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, et al. Systematic discovery of Xist RNA binding proteins. Cell. 2015;161:404-16.
[296] Chi Y, Welcker M, Hizli AA, Posakony JJ, Aebersold R, Clurman BE. Identification of CDK2 substrates in human cell lysates. Genome Biol. 2008;9:R149.
[297] Tropea JE, Cherry S, Waugh DS. Expression and purification of soluble His(6)-tagged TEV protease. Methods in molecular biology. 2009;498:297-307.
[298] Klimek-Tomczak K, Wyrwicz LS, Jain S, Bomsztyk K, Ostrowski J. Characterization of hnRNP K protein-RNA interactions. Journal of molecular biology. 2004;342:1131-41.
[299] Dejgaard K, Leffers H, Rasmussen HH, Madsen P, Kruse TA, Gesser B, et al. Identification, molecular cloning, expression and chromosome mapping of a family of transformation upregulated hnRNP-K proteins derived by alternative splicing. Journal of molecular biology. 1994;236:33-48.
[300] UniProt C. UniProt: a hub for protein information. Nucleic acids research. 2015;43:D204-12.
[301] Xue Y, Ren J, Gao X, Jin C, Wen L, Yao X. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. Molecular & cellular proteomics : MCP. 2008;7:1598-608.
[302] Morris DP, Lee JM, Sterner DE, Brickey WJ, Greenleaf AL. Assaying CTD kinases in vitro and phosphorylation-modulated properties of RNA polymerase II in vivo. Methods. 1997;12:264-75.
[303] Keogh MC, Podolny V, Buratowski S. Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. Molecular and cellular biology. 2003;23:7005-18.
[304] Wood A, Shilatifard A. Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb. Cell cycle. 2006;5:1066-8.
[305] Sanso M, Levin RS, Lipp JJ, Wang VY, Greifenberg AK, Quezada EM, et al. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. Genes & development. 2016;30:117-31.
[306] Thisted T, Lyakhov DL, Liebhaber SA. Optimized RNA targets of two closely related triple KH domain proteins, heterogeneous nuclear ribonucleoprotein K and alphaCP-2KL, suggest Distinct modes of RNA recognition. The Journal of biological chemistry. 2001;276:17484-96.
[307] Ostrowski J, Wyrwicz L, Rychlewski L, Bomsztyk K. Heterogeneous nuclear ribonucleoprotein K protein associates with multiple mitochondrial transcripts within the organelle. The Journal of biological chemistry. 2002;277:6303-10.

[308] Habelhah H, Shah K, Huang L, Burlingame AL, Shokat KM, Ronai Z. Identification of new JNK substrate using ATP pocket mutant JNK and a corresponding ATP analogue. The Journal of biological chemistry. 2001;276:18090-5.
[309] Grassilli E, Pisano F, Cialdella A, Bonomo S, Missaglia C, Cerrito MG, et al. A novel oncogenic BTK isoform is overexpressed in colon cancers and required for RAS-mediated transformation. Oncogene. 2016;35:4368-78.
[310] Barboro P, Borzi L, Repaci E, Ferrari N, Balbi C. Androgen receptor activity is affected by both nuclear matrix localization and the phosphorylation status of the heterogeneous nuclear ribonucleoprotein K in anti-androgen-treated LNCaP cells. PloS one. 2013;8:e79212.
[311] Kay BK, Williamson MP, Sudol M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2000;14:231-41.
[312] Miki TS, Grosshans H. The multifunctional RNase XRN2. Biochemical Society transactions. 2013;41:825-30.
[313] Mikula M, Bomsztyk K, Goryca K, Chojnowski K, Ostrowski J. Heterogeneous nuclear ribonucleoprotein (HnRNP) K genome-wide binding survey reveals its role in regulating 3'-end RNA processing and transcription termination at the early growth response 1 (EGR1) gene through XRN2 exonuclease. The Journal of biological chemistry. 2013;288:24788-98.
[314] Moritz B, Lilie H, Naarmann-de Vries IS, Urlaub H, Wahle E, Ostareck-Lederer A, et al. Biophysical and biochemical analysis of hnRNP K: arginine methylation, reversible aggregation and combinatorial binding to nucleic acids. Biol Chem. 2014;395:837-53.
[315] Beuth B, Pennell S, Arnvig KB, Martin SR, Taylor IA. Structure of a Mycobacterium tuberculosis NusA-RNA complex. The EMBO journal. 2005;24:3576-87.
[316] Holmes JK, Solomon MJ. A predictive scale for evaluating cyclin-dependent kinase substrates. A comparison of p34cdc2 and p33cdk2. The Journal of biological chemistry. 1996;271:25240-6.
[317] Songyang Z, Blechner S, Hoagland N, Hoekstra MF, Piwnica-Worms H, Cantley LC. Use of an oriented peptide library to determine the optimal substrates of protein kinases. Current biology : CB. 1994;4:973-82.
[318] Kitagawa M, Higashi H, Jung HK, Suzuki-Takahashi I, Ikeda M, Tamai K, et al. The consensus motif for phosphorylation by cyclin D1-Cdk4 is different from that for phosphorylation by cyclin A/E-Cdk2. The EMBO journal. 1996;15:7060-9.
[319] Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. Nucleic acids research. 2006;34:D655-8.
[320] Li SS. Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. The Biochemical journal. 2005;390:641-53.
[321] Ostareck-Lederer A, Ostareck DH, Rucknagel KP, Schierhorn A, Moritz B, Huttelmaier S, et al. Asymmetric arginine dimethylation of heterogeneous nuclear ribonucleoprotein K by protein-arginine methyltransferase 1 inhibits its interaction with c-Src. The Journal of biological chemistry. 2006;281:11115-25.
[322] Yamagata K, Daitoku H, Takahashi Y, Namiki K, Hisatake K, Kako K, et al. Arginine methylation of FOXO transcription factors inhibits their phosphorylation by Akt. Molecular cell. 2008;32:221-31.

[323] Guo Z, Zheng L, Xu H, Dai H, Zhou M, Pascua MR, et al. Methylation of FEN1 suppresses nearby phosphorylation and facilitates PCNA binding. Nat Chem Biol. 2010;6:766-73.
[324] Hsu JM, Chen CT, Chou CK, Kuo HP, Li LY, Lin CY, et al. Crosstalk between Arg 1175 methylation and Tyr 1173 phosphorylation negatively modulates EGFR-mediated ERK activation. Nature cell biology. 2011;13:174-81.
[325] Yang JH, Chiou YY, Fu SL, Shih IY, Weng TH, Lin WJ, et al. Arginine methylation of hnRNPK negatively modulates apoptosis upon DNA damage through local regulation of phosphorylation. Nucleic acids research. 2014;42:9908-24.
[326] Bedford MT, Richard S. Arginine methylation an emerging regulator of protein function. Molecular cell. 2005;18:263-72.
[327] Michael WM, Siomi H, Choi M, Pinol-Roma S, Nakielny S, Liu Q, et al. Signal sequences that target nuclear import and nuclear export of pre-mRNA-binding proteins. Cold Spring Harb Symp Quant Biol. 1995;60:663-8.
[328] Pinol-Roma S, Dreyfuss G. Transcription-dependent and transcription-independent nuclear transport of hnRNP proteins. Science. 1991;253:312-4.
[329] Michael WM, Choi M, Dreyfuss G. A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. Cell. 1995;83:415-22.
[330] Kimura Y, Nagata K, Suzuki N, Yokoyama R, Yamanaka Y, Kitamura H, et al. Characterization of multiple alternative forms of heterogeneous nuclear ribonucleoprotein K by phosphate-affinity electrophoresis. Proteomics. 2010;10:3884-95.
[331] Liu P, Xiang Y, Fujinaga K, Bartholomeeusen K, Nilson KA, Price DH, et al. Release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein (snRNP) activates hexamethylene bisacetamide-inducible protein (HEXIM1) transcription. The Journal of biological chemistry. 2014;289:9918-25.
[332] Gruber AR, Kilgus C, Mosig A, Hofacker IL, Hennig W, Stadler PF. Arthropod 7SK RNA. Molecular biology and evolution. 2008;25:1923-30.
[333] Gruber AR, Koper-Emde D, Marz M, Tafer H, Bernhart S, Obernosterer G, et al. Invertebrate 7SK snRNAs. Journal of molecular evolution. 2008;66:107-15.
[334] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646-74.
[335] Sano M, Abdellatif M, Oh H, Xie M, Bagella L, Giordano A, et al. Activation and function of cyclin T-Cdk9 (positive transcription elongation factor-b) in cardiac muscle-cell hypertrophy. Nature medicine. 2002;8:1310-7.
[336] Huang F, Wagner M, Siddiqui MA. Ablation of the CLP-1 gene leads to down-regulation of the HAND1 gene and abnormality of the left ventricle of the heart and fetal death. Mechanisms of development. 2004;121:559-72.
[337] Mancebo HS, Lee G, Flygare J, Tomassini J, Luu P, Zhu Y, et al. P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. Genes & development. 1997;11:2633-44.
[338] Irie K, Nomoto S, Miyajima I, Matsumoto K. SGV1 encodes a CDC28/cdc2-related kinase required for a G alpha subunit-mediated adaptive response to pheromone in S. cerevisiae. Cell. 1991;65:785-95.
[339] Alazami AM, Al-Owain M, Alzahrani F, Shuaib T, Al-Shamrani H, Al-Falki YH, et al. Loss of function mutation in LARP7, chaperone of 7SK ncRNA, causes a syndrome of facial

dysmorphism, intellectual disability, and primordial dwarfism. Human mutation. 2012;33:1429-34.
[340] Cheng Y, Jin Z, Agarwal R, Ma K, Yang J, Ibrahim S, et al. LARP7 is a potential tumor suppressor gene in gastric cancer. Laboratory investigation; a journal of technical methods and pathology. 2012;92:1013-9.
[341] Biewenga P, Buist MR, Moerland PD, Ver Loren van Themaat E, van Kampen AH, ten Kate FJ, et al. Gene expression in early stage cervical cancer. Gynecologic oncology. 2008;108:520-6.
[342] Pino I, Pio R, Toledo G, Zabalegui N, Vicent S, Rey N, et al. Altered patterns of expression of members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family in lung cancer. Lung cancer. 2003;41:131-43.
[343] Ushigome M, Ubagai T, Fukuda H, Tsuchiya N, Sugimura T, Takatsuka J, et al. Up-regulation of hnRNP A1 gene in sporadic human colorectal cancers. International journal of oncology. 2005;26:635-40.
[344] Boukakis G, Patrinou-Georgoula M, Lekarakou M, Valavanis C, Guialis A. Deregulated expression of hnRNP A/B proteins in human non-small cell lung cancer: parallel assessment of protein and mRNA levels in paired tumour/non-tumour tissues. BMC cancer. 2010;10:434.
[345] Carpenter B, MacKay C, Alnabulsi A, MacKay M, Telfer C, Melvin WT, et al. The roles of heterogeneous nuclear ribonucleoproteins in tumour development and progression. Biochimica et biophysica acta. 2006;1765:85-100.
[346] Inoue A, Sawata SY, Taira K, Wadhwa R. Loss-of-function screening by randomized intracellular antibodies: identification of hnRNP-K as a potential target for metastasis. Proceedings of the National Academy of Sciences of the United States of America. 2007;104:8983-8.
[347] Hope NR, Murray GI. The expression profile of RNA-binding proteins in primary and metastatic colorectal cancer: relationship of heterogeneous nuclear ribonucleoproteins with prognosis. Hum Pathol. 2011;42:393-402.
[348] Ostrowski J, Kawata Y, Schullery DS, Denisenko ON, Bomsztyk K. Transient recruitment of the hnRNP K protein to inducibly transcribed gene loci. Nucleic acids research. 2003;31:3954-62.
[349] Quaresma AJ, Bressan GC, Gava LM, Lanza DC, Ramos CH, Kobarg J. Human hnRNP Q re-localizes to cytoplasmic granules upon PMA, thapsigargin, arsenite and heat-shock treatments. Experimental cell research. 2009;315:968-80.
[350] Muck F, Bracharz S, Marschalek R. DDX6 transfers P-TEFb kinase to the AF4/AF4N (AFF1) super elongation complex. Am J Blood Res. 2016;6:28-45.
[351] Roy BB, Hu J, Guo X, Russell RS, Guo F, Kleiman L, et al. Association of RNA helicase a with human immunodeficiency virus type 1 particles. The Journal of biological chemistry. 2006;281:12625-35.
[352] Tang H, Gaietta GM, Fischer WH, Ellisman MH, Wong-Staal F. A cellular cofactor for the constitutive transport element of type D retrovirus. Science. 1997;276:1412-5.
[353] Jarmoskaite I, Russell R. RNA helicase proteins as chaperones and remodelers. Annual review of biochemistry. 2014;83:697-725.
[354] Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nature protocols. 2006;1:1610-6.

[355] Uchikawa E, Natchiar KS, Han X, Proux F, Roblin P, Zhang E, et al. Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7. Nucleic acids research. 2015;43:3373-88.
[356] Baber JL, Libutti D, Levens D, Tjandra N. High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. Journal of molecular biology. 1999;289:949-62.
[357] Martinez-Zapien D, Legrand P, McEwen AG, Proux F, Cragnolini T, Pasquali S, et al. The crystal structure of the 5 functional domain of the transcription riboregulator 7SK. Nucleic acids research. 2017;45:3568-79.
[358] !!! INVALID CITATION !!! [255].
[359] Botstein D, Fink GR. Yeast: an experimental organism for 21st Century biology. Genetics. 2011;189:695-704.
[360] Kennedy BK. Mammalian transcription factors in yeast: strangers in a familiar land. Nature reviews Molecular cell biology. 2002;3:41-9.
[361] Bartkowiak B, Greenleaf AL. Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? Transcription. 2011;2:115-9.
[362] Zhu Y, Pe'ery T, Peng J, Ramanathan Y, Marshall N, Marshall T, et al. Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. Genes & development. 1997;11:2622-32.
[363] Palangat M, Renner DB, Price DH, Landick R. A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. Proceedings of the National Academy of Sciences of the United States of America. 2005;102:15036-41.
[364] Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods. 2009;48:240-8.
[365] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5:621-8.
[366] Chu Y, Sutton A, Sternglanz R, Prelich G. The BUR1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by SET2. Molecular and cellular biology. 2006;26:3029-38.
[367] Liu Y, Warfield L, Zhang C, Luo J, Allen J, Lang WH, et al. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. Molecular and cellular biology. 2009;29:4852-63.
[368] Lygerou Z, Conesa C, Lesage P, Swanson RN, Ruet A, Carlson M, et al. The yeast BDF1 gene encodes a transcription factor involved in the expression of a broad class of genes including snRNAs. Nucleic acids research. 1994;22:5332-40.
[369] Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nature reviews Molecular cell biology. 2013;14:153-65.
[370] Denisenko O, Bomsztyk K. Yeast hnRNP K-like genes are involved in regulation of the telomeric position effect and telomere length. Molecular and cellular biology. 2002;22:286-97.
[371] Goodrich JA, Kugel JF. Binding and kinetics for molecular biologists. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2007.