Exploring a Link Between Stress Induced Termination Defects and the Phosphorylation State of RNA Polymerase II

Anoushka Singh

5 April 2021

Thesis Advisor:

Dr. Jennifer Kugel, Department of Biochemistry

Honors Council Representative:

Dr. Jeffrey Cameron, Department of Biochemistry

Committee Members:

Dr. Seth Myers, Program for Writing and Rhetoric

Dr. James Goodrich, Department of Biochemistry

Table of Contents

Abstract……………………………………………………………………3
Central Dogma………………………………………………………………4
DNA Replication…………………………………………………………….5
RNA Transcription…………………………………………………………..8
Protein Translation………………………………………………………...10
Post Translational Modifications………………………………………..13
Transcription Mechanism………………………………………………...15
The structure of Mammalian RNA Polymerase II……………………………15
Initiation of Transcription…………………………………………………..16
Promoter Proximal Pausing…………………………………………………19
Elongation of Transcription…………………………………………………..20
Termination of Transcription………………………………………………22
Pausing of Pol II……………………………………………………………..23
Poly(A) site dependent termination…………………………………………..25
Hypothesis……………………………………………………………………..28
Stress Induced Termination Defects……………………………………29
Hyperosmotic Stress………………………………………………………….30
HSV1 infection………………………………………………………………32
Influenza Virus Infection……………………………………………………34
Heat Shock……………………………………………………………………..36
Link Between Termination and CTD Modifications…………………...39
Serine 2 phosphorylation is linked to termination in eukaryotes………………40
Tyrosine 1 is linked to termination in eukaryotes………………………………...42
Threonine 4 is linked to termination in mammalian cells………………………..44
Connection between cell stress, kinases, and the CTD……………………...46
Significance and Outlook……………………………………………………….49
Acknowledgements……………………………………………………………….50
References………………………………………………………………………..51

Abstract

Termination of transcription occurs through intricate processes in which RNA polymerase II (Pol II) interacts with many proteins to ultimately detach from the DNA. These processes, however, are impacted by cell stresses. Cell stresses cause a significant impairment in termination through unknown mechanisms. Here, through a review of current scientific literature, I will explore how cell stresses influence phosphorylation of the C-terminal domain (CTD) of the large subunit of Pol II, and the impairment of termination. I hypothesize that stress induced CTD phosphorylation changes are linked to termination defects. I will begin by providing background on the mechanisms through which transcription termination occurs. Then, I will detail the patterns of defective termination caused by various stresses including hyperosmotic stress, HSV1 infection, influenza virus infection, and heat shock. Next, I will discuss the phosphorylation state of the CTD at the Serine 2, Tyrosine 1, and Threonine 4 residues and how they are linked to 3’ end processing and termination. I will show a connection between stresses and enzymes that are involved in the CTD phosphorylation state. Finally, I will discuss the significance of this research and how termination defects might be a survival mechanism for cells to respond to stress. Current literature is consistent with the model that stress induced termination defects can be caused by the phosphorylation state of the CTD.

Central Dogma

One of the most important processes in all living organisms is the ability of an organism to replicate its genetic information and convert it into functional units, a process known as the central dogma of biology (Zhou et al., 2010). DNA is the fundamental building block of life as it contains the genetic information that codes for proteins. The central dogma of biology can be narrowed down to three main steps: DNA replication in the nucleus, transcription into RNA in the nucleus, and translation into proteins (Figure 1) in the cytoplasm (Schneider-Poetsch and Yoshida, 2018). Transcription and translation are the two main steps in expressing a gene, which is important for carrying out functions in the cell (Schneider-Poetsch and Yoshida, 2018).

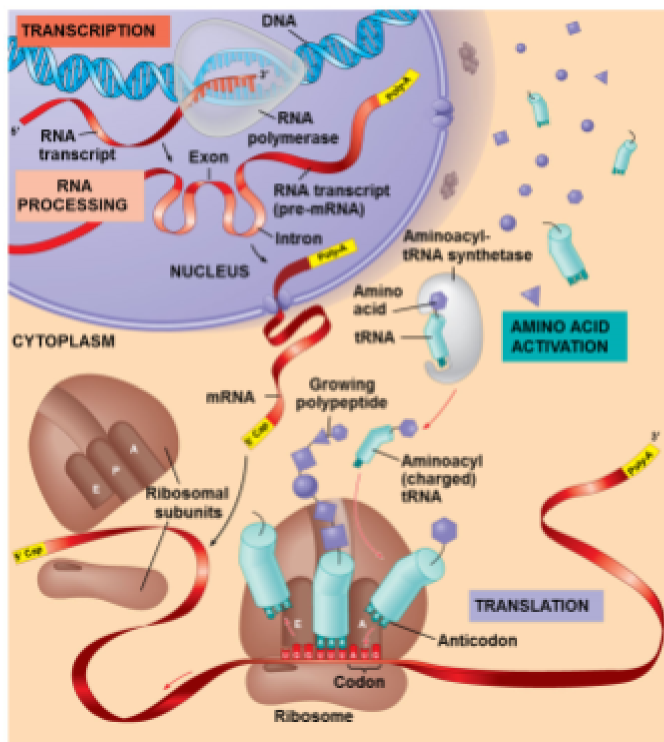

Figure 1. Schematic shows transcription occurring in the nucleus, movement of mRNA into the cytoplasm, and protein translation occurring on the ribosome. Figure taken from https://www.slideshare.net/kindarspirit/17-from-gene-to-protein-22899119.

DNA replication

DNA is a double-stranded helix, consisting of two strands of nucleotides. Each nucleotide consists of three main groups: a deoxyribose sugar, a phosphate group, and a nitrogenous base (Burgers and Kunkel, 2017). The four different bases in DNA include adenine, guanine, thymine, and cytosine (Figure 2). Both of the strands in DNA are connected by hydrogen bonds between nucleotide bases (Burgers and Kunkel, 2017). Adenine and thymine are connected by two hydrogen bonds while cytosine and guanine are connected by three hydrogen bonds (Burgers and Kunkel, 2017).

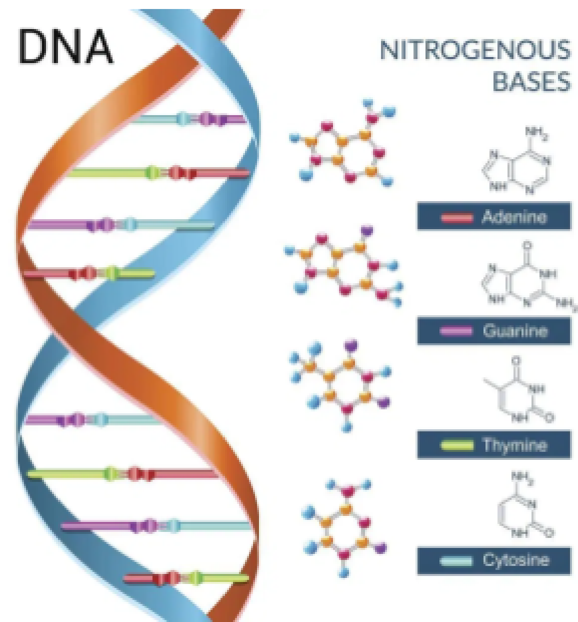

Figure 2. This figure represents the double helix structure of DNA with a sugar phosphate backbone on both strands. A and T are complementary base pairs bound by two hydrogen bonds. G and C are complementary base pairs with three hydrogen bonds. This figure is taken from https://www.medicalnewstoday.com/articles/319818.

When cells divide the genetic information in DNA needs to be replicated. DNA helicase binds to DNA on the origin of replication to begin forming the replication fork, which is the first step in DNA replication (Dewar and Walter, 2017). There are multiple origins of replication in eukaryotic cells (Ekundayo and Bleichert, 2019). The helicase separates the two strands by disrupting the hydrogen bonds between the bases.

Helicase uses an enormous amount of energy to break the hydrogen bonds, which comes from ATP hydrolysis (Burgers and Kunkel, 2017). After the strands have been separated (lagging strand (5' to 3') and leading strand (3' to 5'), the process of

replication begins with the building of RNA primers by DNA primase (Burgers and Kunkel, 2017). DNA primase is a DNA-dependent RNA polymerase responsible for making short RNA primers on each DNA strand in order for DNA polymerase III to begin replication. An RNA primer is needed as DNA polymerase III cannot initiate but adds nucleotides to 3' ends of DNA using the free 3' hydroxyl group on the sugar (Burgers and Kunkel, 2017). DNA pol III consists of three main subunits including α, ε, and θ (Burgers and Kunkel, 2017). The alpha subunit is responsible for the synthesis of new DNA while the epsilon and theta subunits act as proofreaders. DNA polymerase III can only synthesize DNA in the 5' to 3' direction and thus there is a slight difference in the replication mechanism on the leading and lagging strands. Along the leading strand, DNA polymerase III can continuously make the new strand as the replication fork moves in the same direction as the DNA pol III (Glover and McHenry, 2001). The lagging strand runs antiparallel to the leading strand (5' to 3' in the opposite direction). On the lagging strand, the DNA polymerase III synthesizes new DNA in fragments 5' to 3' away from the replication fork (Glover and McHenry, 2001). These fragments are called Okazaki fragments and are synthesized simultaneously with multiple primers being built at the same time. These fragments are only a couple hundred nucleotides long and take an extra step to be linked together using DNA ligase.

After the two new strands of DNA are formed, the RNA primers need to be removed (Frouin et al., 2003). The 5' to 3' exonuclease function of RNase removes the RNA primers from the leading strand at the beginning of each Okazaki fragment (Frouin et al., 2003). DNA polymerase I then fills in these gaps with their appropriate bases. Finally, DNA ligase joins the Okazaki fragments together to give rise to two new DNA

strands. While the replisome progresses, DNA can become supercoiled (Champoux, 2001). Topoisomerases are enzymes which are important in DNA replication as they help with underwinding and overwinding the coils of the DNA by either performing single or double strand cuts (Champoux, 2001). A typical eukaryotic replication fork is shown in Figure 3.

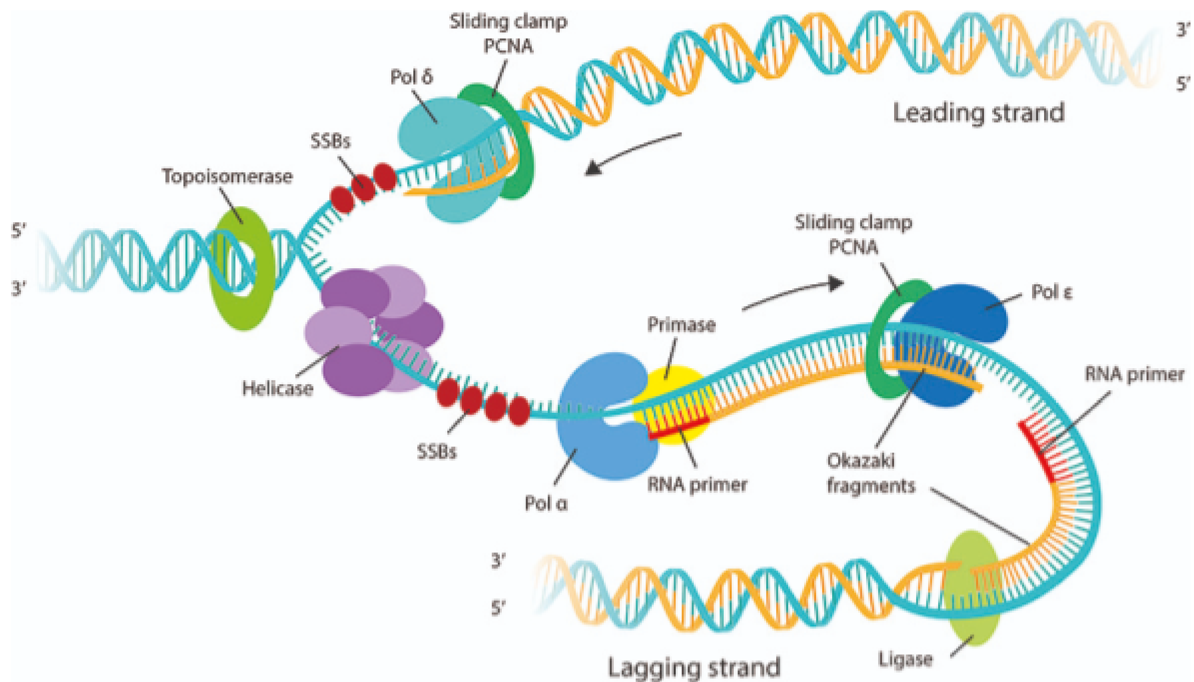

Figure 3. This figure shows a eukaryotic DNA replisome and the major components involved. The image is taken from https://touchnomore.org/home/

Replication is initiated at distinct origin spots along the DNA leading to multiple replication forks. DNA replication at one fork is terminated when it either comes to the end of the chromosome, or it meets with another replisome (“encounter”) which is traveling in the opposite direction (Yao and O’Donnell, 2018). However, if something goes wrong at the termination site of the replication fork, this can lead to severe consequences for the DNA. DNA rearrangement, amplification, or other major mutations may occur which can be harmful to the cell (Dewar and Walter, 2017).

RNA Transcription

Transcription of protein-encoding genes is the second step in the central dogma. This section provides a brief overview of transcription. Since the purpose of this paper is based on transcription, it will be discussed in detail later (Transcription Mechanism, p.15). Transcription is carried out in cells to make a messenger RNA (mRNA) copy of one strand of the DNA that encodes for a gene (Zhang et al., 2012). Being the first step in gene expression, transcription is highly regulated. The goal of transcription is to create an mRNA molecule that ultimately serves as the template for the production of functional proteins (Schneider-Poetsch and Yoshida, 2018).

Transcription, which is the conversion of DNA into RNA, is essential because some RNAs provide a code for the synthesis of proteins (mRNA), while others can carry out their own functions (tRNA, rRNA, and more) (Wu et al., 2014). Having RNA allows for multiple isoforms to be produced from the same gene due to differential splicing (Mehmood et al., 2020). It also allows for gene regulation where the levels of mRNA can influence the abundance of the protein (Baralle and Giudice, 2017). Cells can then more rapidly respond to changes in the environment by altering multiple points of gene expression and by degrading or sequestering mRNA (Mehmood et al., 2020). Therefore, a messenger RNA (mRNA) is needed to act as a middleman between genes and their respective protein. Although there are many types of RNA, this section will focus on mRNA which is the RNA responsible for converting DNA templates to their proteins. (Proudfoot, 2016)

Eukaryotes have three different RNA polymerases. RNA polymerase I (RNAP I) is mainly responsible for synthesizing rRNA which is integral in the structure of

ribosomes (Wu et al., 2014). RNA polymerase II (RNAPII or Pol II) serves to transcribe genes into mRNAs as well as transcribes lncRNAs. RNA polymerase III (RNAP III) functions to synthesize tRNA and other small nuclear and cytosolic RNA. This paper will focus on the structure and function of RNAP II (Pol II).

Eukaryotic mRNA transcription occurs in three different phases: initiation, elongation, and termination (Venters and Pugh, 2009). During initiation, RNA polymerase II (Pol II) binds the core promoter region of a gene, near the start site of transcription, along with other proteins known as general transcription factors (Venters and Pugh, 2009). This is known as the preinitiation complex. After initiation has successfully been completed, the RNA polymerase will begin synthesizing the new complementary RNA strand using the DNA template strand in the process called elongation (Venters and Pugh, 2009). The nascent RNA strand will be formed from its 5' to its 3' end. Each RNA base is complementary to the corresponding DNA base on the template strand, except thymine is replaced by uracil in RNA (Venters and Pugh, 2009). Once the RNA polymerase has

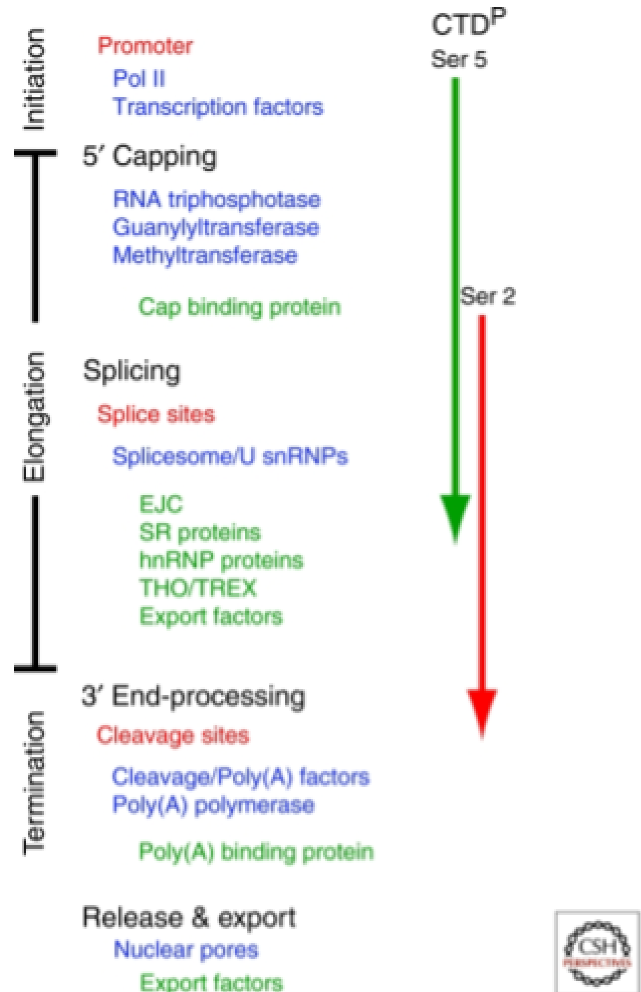

Figure 4. Shows all the steps for transcription and how cotranscriptional processes take place with the progression of the Pol II. Figure taken from *Hocine et al. 2010.*

finished synthesizing the nascent RNA strand, the termination of transcription will take place (Eaton and West, 2020). All these steps of transcription involve many regulatory factors, which help the RNA polymerase with the process. A summary of the transcriptional process, with co-transcriptional RNA processing, is shown in Figure 4.

Protein Translation

The second phase of gene expression is protein translation. Once the mRNA is transcribed, it will leave the nucleus in order to be translated into polypeptide chains by ribosomes (Merrick, 1992). These polypeptide chains fold into proteins and perform functions in cells. Proteins carry out the functions of nearly every cellular process and comprise nearly every cellular structure. Similar to transcription, translation also occurs in its own initiation, elongation, and termination phases. Initiation involves the leader sequence of the mRNA, which is located on the 5’ end between the first nucleotide of the mRNA and the first AUG codon (Pain, 1996). This segment of the mRNA is also known as the 5’ untranslated region (5’UTR) which includes the kozak sequence which directs translation initiation and contains the start codon (Pain, 1996). Translation relies on ribosomes. Although ribosomes have two subunits, they exist independently of each other in the cytoplasm and only combine together when they bind to an mRNA for translation (Hershey et al., 2019). The 5’UTR can also serve to impact the stability of the mRNA or provide binding sites for other proteins (Hershey et al., 2019).

During initiation, the translation initiation complex is formed around the mRNA (Pain, 1996). In eukaryotes, initiation factors such eIF3 or eIF4F attach to the smaller (40s) subunit of the ribosome as well as to the 5’ region of the mRNA to hold the

ribosome on the mRNA (Pain, 1996). Then, a tRNA will bind to the start codon (AUG) of the mRNA which codes for a methionine amino acid (Hershey et al., 2019). Each codon in the mRNA will code for a specific amino acid which will be read by the tRNA (Hong et al., 2018). tRNAs read the codon of the mRNA through specific base pairing of their anticodon to the mRNA. Each base-paired tRNA has a separate end that binds to a unique amino acid, specific to that anticodon. After this tRNA-amino acid complex is formed, the larger subunit of the ribosome attaches to it causing the initiation factors to release (Hershey et al., 2019). The subunits of the ribosome have three different sites (Hong et al., 2018). The A site is where the tRNA's anticodon pairs with the mRNA's codon. The P site is where the tRNA adds the amino acid and the polypeptide chain is elongated. The E site is where the tRNA exits the ribosome.

After all the components required for the initiation of translation are in place, the elongation process occurs. The elongation phase of translation takes place when the polypeptide chain gets longer (Dever et al., 2018). Different elongation factors are used to guide this process. The eukaryotic elongation factor eIEF2 helps guide the ribosome along the mRNA in a 5' to 3' direction from the A site to the P site of the ribosome by hydrolyzing GTP (Robison and Colbran, 2013). Then, a second tRNA binds to the A site of the larger subunit which is complementary to the codon on the mRNA (Dever et al., 2018). The amino acids in the polypeptide chain are joined together through a peptidyl transferase reaction catalyzed by rRNA within the ribosome (Dever et al., 2018). Each time another amino acid is added to the chain, the first tRNA moves to the E site, the second tRNA moves to the P site, and a third tRNA attaches to the A site. This cycle continues until the termination of translation.

The termination of translation is triggered by one of three stop codons: UAA, UAG, and UGA. These codons are not recognized by any tRNAs. Instead, they are recognized by proteins called release factors (Hershey et al., 2019). In eukaryotes, the release factor eRF1 is responsible for recognizing all three stop codons and signaling for the dissociation of the ribosome and the release of the newly synthesized polypeptide chain (Frolova et al., 2000). Release factor eRF3 helps with the release of the polypeptide (Frolova et al., 2000). A summary of protein translation is shown in Figure 5.

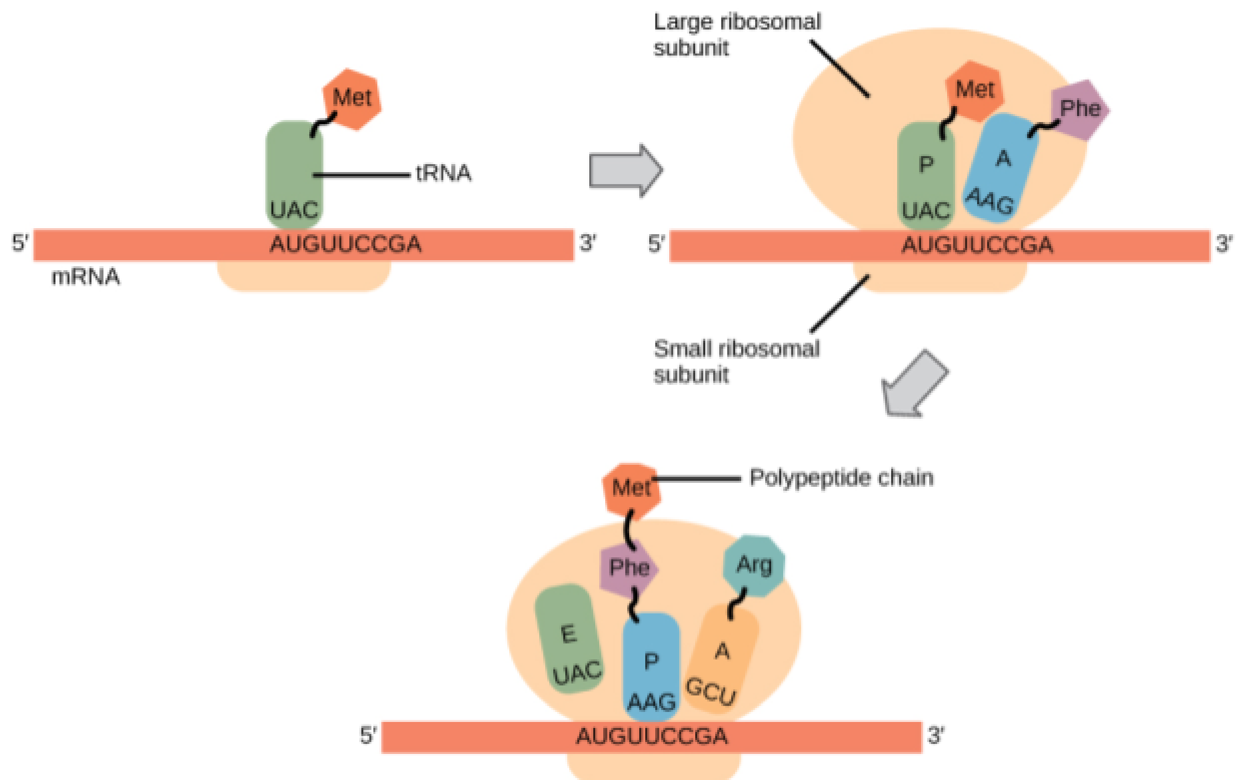

Figure 5. This schematic shows the small ribosomal subunit binding with the mRNA strand. The large ribosomal subunit then binds and translation starts at the AUG codon. The polypeptide chain begins to form with each new codon being translated.
This figure is taken from
https://courses.lumenlearning.com/wm-nmbiology1/chapter/reading-steps-of-translation/.

Post Translational Modifications

Additionally, polypeptides often undergo post-translational modifications (PTMs), which are changes in the chemical structure of the protein due to the addition of different molecules or the cleavage of the protein. PTMs can either be used to modulate the activity of the protein or to further specialize the protein for a specific function in the cell. These changes add to the diversity of proteins that can be expressed and help regulate protein function (Uversky, 2013). Some proteins undergo certain modifications post-synthesis. Post-translational modifications (PTMs) are important for protein function and proteome diversity. They help with protein specialization or regulation as well as modify proteins in response to stimuli (Herhaus and Dikic, 2015). Although there are numerous types of PTMs, a few of the most common ones are described below.

Protein phosphorylation is one of the most common PTMs, and is important for my hypothesis regarding how transcriptional termination is controlled during cell stress (Vilborg et al., 2017). Phosphorylation refers to the attachment of a phosphate group to any biomolecule, a reaction that is catalysed by kinases (Barber and Rinehart, 2018). Kinases are enzymes that transfer the terminal phosphate group from ATP to the substrate being phosphorylated. Phosphorylation of proteins occurs on serine, threonine, or tyrosine residues because they all have hydroxyl groups suitable for phosphorylation (Barber and Rinehart, 2018). Phosphorylating specific residues on a protein is very essential to regulating biological pathways in a cell because phosphorylation can act as a “switch” to turn a protein on or off (Pereira et al., 2011). With the addition of phosphate groups, proteins can change their interaction partners or their activities (Pereira et al., 2011). Phosphorylation is also reversible since phosphate

groups can also be detached from the molecule by phosphatases (Pereira et al., 2011). Thus, since phosphate groups can be added and removed phosphorylation can function as a switch for protein activity. For example, the phosphorylation of the protein p53 can initiate the transcription of genes which are responsible for inhibiting the cell cycle, DNA repair, an even cell death (Ardito et al., 2017).

Another common post-translational modification is called ubiquitination. Ubiquitin is a small protein found in the cell which attaches to other proteins as a signal for degradation by the proteasome (Herhaus and Dikic, 2015). Ubiquitination occurs in different steps (Pickart, 2001). The first step is the activation of the ubiquitin molecule by an E1 enzyme which consumes ATP to complete this step. The second step is the conjugation of the ubiquitin onto an E2 enzyme which causes the E2 enzyme to transfer its ubiquitin molecule onto the E3 enzyme. The ubiquitin is transferred multiple times so that there can be a few types of E1-2 molecules but many E3s that are specific for the target protein. For instance, the ubiquitination of RNA polymerase II is important in transcriptional arrest as it causes the proteasome to degrade the Pol II when necessary (Proudfoot, 2016). Along with marking proteins for degradation, ubiquitin can also be attached to proteins for other functions, including immune responses or signaling for DNA repair (Hu and Sun, 2016; Schwertman et al., 2016).

Protein methylation is a PTM that is essential for cellular processes like transcriptional regulation or signal transduction (Rahimi and Costello, 2015; Zhang, 2001). Methylation refers to the addition of a methyl group to certain amino acid residues in a protein. Arginine and lysine residues are the most common amino acids to undergo methylation. The methylation of proteins can inhibit or promote their

functionality based on the type of methylation because the addition of methyl groups impacts the binding partners of the protein (Zhang, 2001). The enzyme that carries out this reaction is a methyltransferase which is activated when S-adenosyl-l-methionine (SAM) donates a methyl group (Zhang, 2001). For example, methylation of histone H3 and H4 proteins in the chromatin are believed to be recognized and bound by other proteins responsible for regulating gene expression (Whetstine).

Transcription Mechanism

The Structure of mammalian RNA Polymerase II

Pol II is a multiprotein complex that is responsible for the synthesis of all messenger RNAs (mRNAs) as well as many long non-coding RNAs (lncRNAs) from DNA templates (Woychik and Hampsey, 2002). This DNA dependent RNA polymerase is a large asymmetrical molecule with a ten subunit core and two additional subunits (Woychik and Hampsey, 2002). The “clamp” of the Pol II molecule functions to bind the DNA and lead it to the cleft of the active site of Pol II (composed of the Rpb2 and Rpb1 subunits) so that the enzyme can then perform its polymerase activity. Rpb1 is the largest subunit on Pol II and it has a long unstructured carboxy terminal domain (CTD) that is important for regulating transcription (Woychik and Hampsey, 2002). The CTD has 52 heptad repeats of the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (McCracken et al., 1997). PTMs are added to the residues in these repeats to control Pol II. For example, the phosphorylation marks on several different residues of the CTD determine play a role in the elongation and termination process, as well as the co-transcriptional processing of the RNA transcript

(Hsin and Manley, 2012). At the 3' end of the nascent RNA Pol II will keep adding new nucleotides to elongate the transcript and build on the DNA:RNA hybrid (Fong and Zhou, 2001). The mRNA transcript will exit the enzyme at the exit channel (Woychik and Hampsey, 2002). Other smaller subunits in the pol II guide the holoenzyme in carrying out its function. Pol II, however, needs the assistance of other factors, known as general transcription factors, to recognize the target promoter, elongate the transcript, and even terminate transcription (Bentley, 2002).

Initiation of Transcription

The transcription of DNA to RNA begins with a preinitiation complex (PIC) assembling on the promoter region of the DNA (Haberle and Stark, 2018). Initiation is the first step in transcription in which the DNA is unwound, and Pol II begins to synthesize the RNA chain by binding the first few nucleotides. This complex consists of many general transcription factors (GTFs, TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH) binding to the promoter region and recruiting the Pol II to bind the DNA duplex (Haberle and Stark, 2018). The promoter region is marked with several core promoter elements, including the TATA box (consensus sequence is TATAAAAG) and a pyrimidine-rich initiator element near the start site of transcription (Down, 2002). On average, the transcription start site (TSS) is located about 30 bp downstream from the TATA box. In addition to these core promoter elements, distal enhancer elements regulate the assembly of PICs at the core promoters of genes. The distal enhancer region can be located kilobases upstream or downstream of the TSS (Haberle and Stark, 2018). Enhancers are regions upstream of the TSS where combinations of sequence-specific

transcription factors bind in order to activate or repress transcription in a gene specific manner.

The PIC begins to form with the binding of TFIID/TFIIA (Ranish et al., 1999). TFIID has a subunit called the TATA binding protein (TBP), which binds to the TATA box through an induced-fit mechanism. The TFIID also contains TBP-associated factors (TAFs) that assist in the binding of TBP with the promoter and interact with regulatory proteins. The chromatin is slightly remodeled to expose the TATA box so it becomes sterically favorable for the TFIID to bind to it (Drew and Travers, 1985) and form a stable interaction with the help of TFIIA. Next, the TFIIB factor will bind to the TFIID/TFIIA. The C terminal core of the TFIIB can bind to TBP and alsoDNA upstream and downstream regions of the TATA box (Zhu et al., 1996). Upon recruitment of TFIIB to the DNA, the next transcription factor, TFIIF, joins the PIC along with Pol II. TFIIE is another basal transcription factor that binds to the PIC to recruit the TFIIH (Ohkuma et al., 1995). The TFIIH is an important

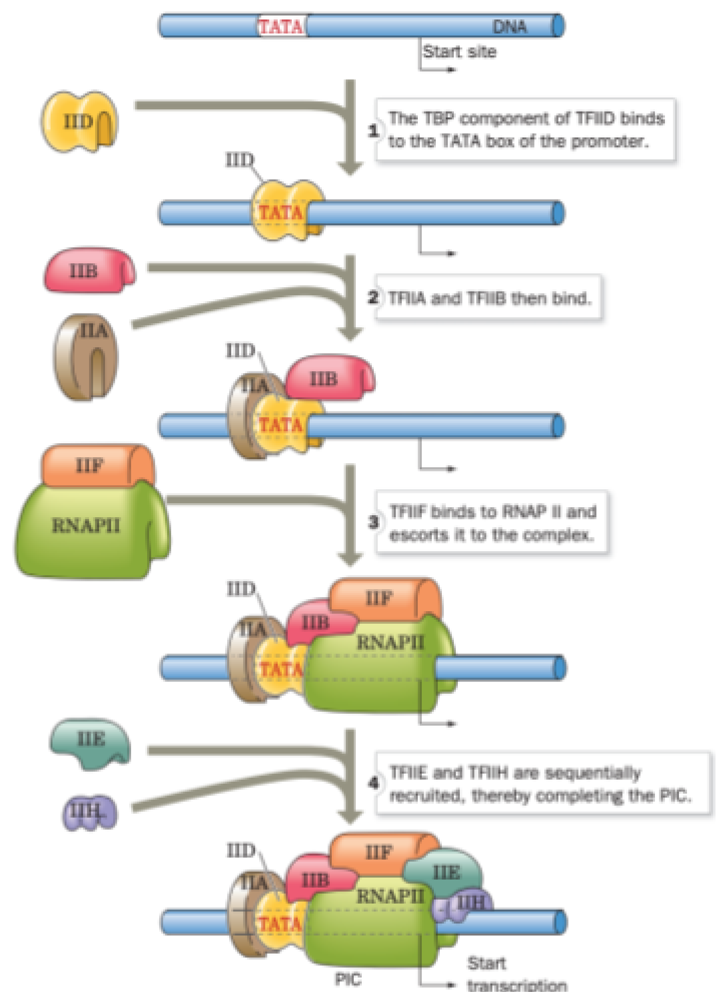

Figure 6. Schematic showing the order of which the GTFs bind to the promoter region to recruit the Pol II and initiate transcription. Figure taken from Voett et al. *Fundamentals of Biochemistry: Life at the molecular level*, 2013, p. 939.

transcription factor because it carries out catalytic functions including being an ATP-dependent DNA helicase, and a kinase important for phosphorylation of certain CTD residues (Nikolov and Burley, 1997). A step by step binding of the transcription factors is shown in Figure 6. The initiation of transcription relies on the stability of the PIC (Ranish et al., 1999). Once the PIC properly guides and attaches the Pol II onto the DNA to form a stable complex transcription initiates. During the early stage of transcription TFIIB, TFIIE, and TFIIH disassemble from the transcribing complex while TFIIF is thought to track with pol II during transcription (Ranish et al., 1999). Abortive initiation can occur in which the pol II unwinds short sections of DNA to create short mRNA transcripts before it continues into promoter escape (Hsu, 2009).

Enhancers bind activators and repressors, which are transcription factor proteins that regulate formation of the PIC (Knoll et al., 2018). The mediator complex is a multisubunit complex that aids in the association of the general transcription factors with the activators/repressors bound at enhancers, as well as pol II, and stabilizes the entire complex (Knoll et al., 2018). The intervening DNA between the core promoter and enhancer DNA is looped using DNA bending proteins which bring the distal enhancer elements closer to the PIC (Vámosi and Rueda, 2018).

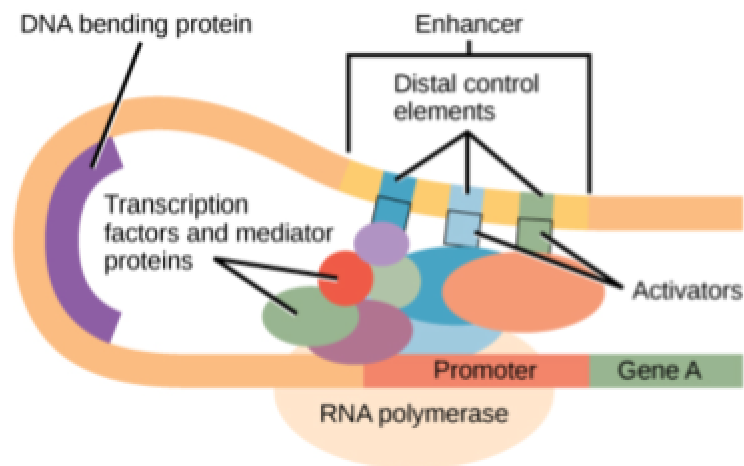

Figure 7. This schematic shows the promoter region of the gene and how the Pol II has multiple other factors that are involved in the pre-initiation complex. It also shows the DNA bending to bring the enhancers closer to the promoter region to enhance transcription. Image taken from https://chem.libretexts.org/@go/page/13361.

These enhancers contain activators which bind to transcription factors to stabilize the PIC, while repressors are proteins that bind to enhancers to interfere with the PIC and repress transcription (Kolovos et al., 2012). Figure 7 represents the transcription complex formed at the promoter before Pol II goes into elongation.

Promoter Proximal Pausing

After initiation and prior to transcriptional elongation is promoter proximal pausing. Promoter proximal pausing is a regulation mechanism which pauses the Pol II prior to the elongation stage of transcription (Adelman and Lis, 2012). Once Pol II transcribes ~20-90 nucleotides of nascent RNA, it is paused and will only continue elongation following pause release (Adelman and Lis, 2012). Pol II pausing is established by two different proteins, DRB sensitivity-inducing factor (DSIF) and the Negative elongation factor (NELF), binding to the Pol II after promoter escape (Yamaguchi et al., 2013). More recent studies show that other factors, such as Gdown1, can also be involved in stabilizing the complex while it is in the promoter proximal pause (Guo et al., 2013). Promoter proximal pausing is released with the dissociation of NELF triggered by the positive transcription elongation factor b (P-TEFb) (Adelman and Lis, 2012). P-TEFb phosphorylates the NELF, DSIF and the CTD (on Ser2) using its Cdk9 (cyclin-dependent kinase 9) subunit (Adelman and Lis, 2012). This makes the NELF disassociate from the complex and converts the DSIF to a positive elongation factor (Yamaguchi et al., 2013). Pol II will then continue to elongate the RNA.

Elongation of Transcription

Elongation of the pre-mRNA occurs in the 5' to 3' direction (Pufall and Kaplan, 2013). Elongation takes place in the transcription bubble, an unwound region of DNA of about 8-9 base pairs, which is maintained as the DNA unwinds downstream and rewinds upstream of the bubble (Pufall and Kaplan, 2013). Elongation can only occur if the serine 5 residue of the CTD is phosphorylated during PIC formation by the TFIIH (Pufall and Kaplan, 2013). When TFIIH phosphorylates the serine 5 residue of the CTD and P-TEFb phosphorylates the serine 2 residue, pol II can continue elongation. These phosphorylation marks recruit other molecules necessary for elongation including the mRNA capping enzyme and other polymerase II associated factors including the positive transcription elongation factor PTEFb and negative elongation factors such as NELF, and DRB sensitivity inducing factor (DSIF) (Yamaguchi et al., 2013). mRNAs undergo 5' capping, splicing, and 3' end processing. Capping and splicing occur cotranscriptionally (Fong et al., 2017). In general, Pol II moves along the DNA template at a rate of 1-1.5 kb/minute (Neugebauer, 2002). The slowness of transcription might occur to accommodate for the co-transcriptional processes.

When the newly synthesized mRNA transcripts are about 20-40 base pairs long a 7-methylguanosine cap is added to its 5' end via a triphosphate bridge (Furuichi, 2015). The alpha phosphate group of the 5' end is removed by a phosphohydrolase enzyme. Then, a guanylyltransferase enzyme attaches a GMP group from GTP onto the beta phosphate group of the 5' end. Finally, the guanine 7 methyl transferase takes a methyl group from the S-Ado-Met and attaches it onto the guanosine (Furuichi, 2015). The function of this cap is diverse including the protection of the nascent transcript from

degradation by the action of the 5' to 3' exonucleases. The cap also serves as a site for the binding of the cap-binding complex which is responsible for the splicing of the first intron (Neugebauer, 2002). It can also bind translation factors in the cytoplasm to mark it for translation initiation (Furuichi, 2015).

Splicing of the pre-mRNA is another important co-transcriptional process that involves the removal of introns to leave the remaining exons as part of the protein coding transcript (Chathoth et al., 2014). The spliceosome is a 60S ribonucleoprotein biomolecule which is responsible for the process of splicing (Chathoth et al., 2014). Splicing takes place in two steps (Neugebauer, 2002). The first step of splicing refers to the 2'hydroxyl group of the adenosine

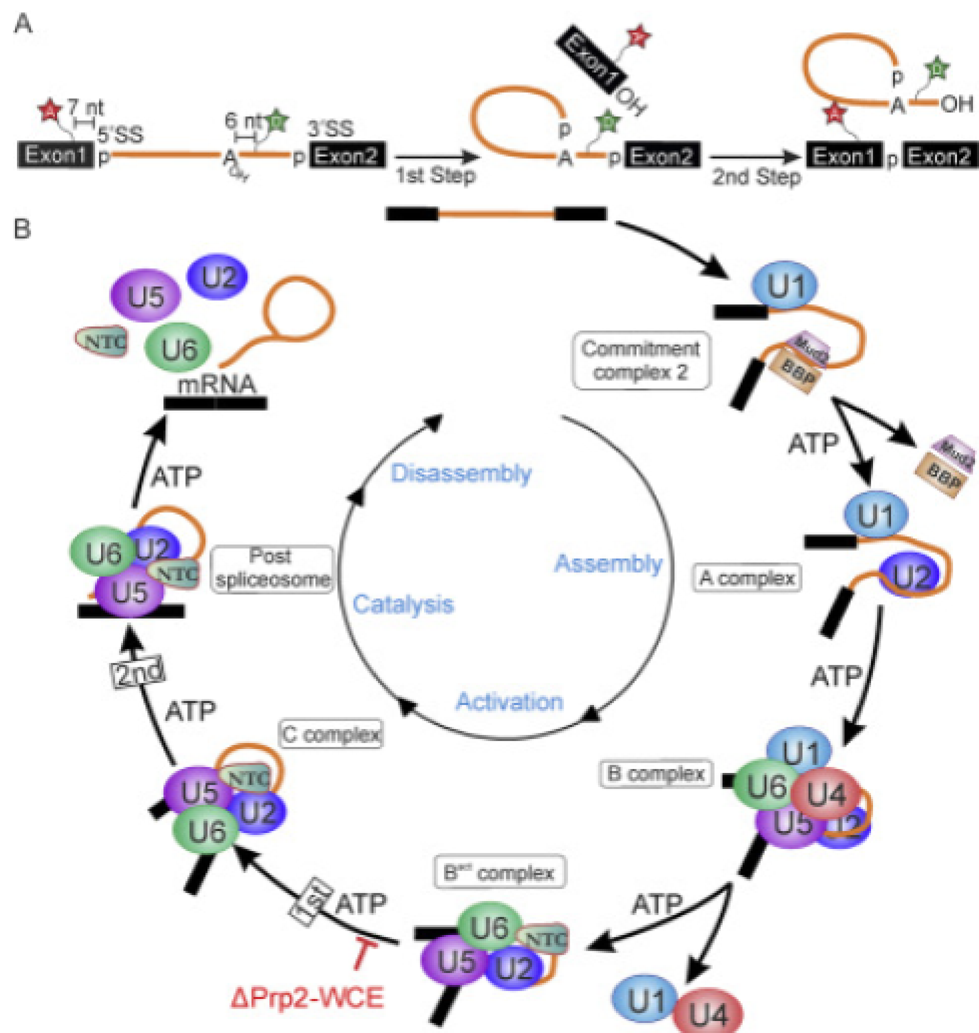

Figure 8. The canonical mechanism of pre-mRNA splicing by the spliceosome. (A) shows where in the gene the introns are being spliced and (B) shows the stepwise fashion of formation of the spliceosome by the binding of the different enzymes. Figure taken from *Kahlscheuer et al, 2015.*

attacking the 5’ splice site (Gehring and Roignant, 2021). In the second step, the 3’ hydroxyl group in the 5’ splice site attaches to the phosphate group in the 3’ splice site. As a result, the two exons are ligated together. The splice site is recognized by different snRNPs (small nuclear ribonucleoprotein) on the two ends of the splice sites and is thought to be recruited by phosphorylated CTD (Fong and Zhou, 2001). Introns are not spliced on the basis of when they are formed (LeMaire and Thummel, 1990; Wetterberg et al., 1996). Splicing kinetics are dependent on the rate of transcription. Splicing can occur within 30 seconds of the transcription of the 3’ splice site (Bentley, 2002). Splicing causes different combinations of exons to be formed to create diversity in the proteome and create a chain of translatable mRNA (Gehring and Roignant, 2021). The enzymes responsible for splicing are shown in Figure 8.

Termination of Transcription

The termination of transcription is the dissociation of the Pol II from the DNA template. This occurs a few hundred base pairs downstream of the annotated 3’ end of the mRNA (Proudfoot, 2016). Termination can be signalled by various mechanisms. One of the most studied mechanisms of termination is dependent upon the polyadenylation signal (PAS), a sequence in the mRNA (AAUAAA) which helps trigger the cleavage and polyadenylation of the nascent transcript (Proudfoot, 2016). Proper termination is important for preventing Pol II from continuing into downstream regions of the DNA and potentially into other genes, causing consequences to the cells (Proudfoot, 2016). Pol II is vulnerable to termination throughout transcription of the gene, not just at the termination site. This section will cover all the types of termination mechanisms.

Pausing of Pol II

Pol II is capable of transcribing RNAs of diverse lengths, including snRNA genes of a couple hundred base pairs and all protein coding genes, which can be a couple hundred kilobases in length (Proudfoot, 2016). So, the termination of Pol II requires a strong mechanism capable of stopping this highly processive biomolecule. Pol II slows down once it passes the PAS as the cleavage and polyadenylation (CPA) complex is being recruited, which is called transcriptional pausing (see Figure 9, (1)). The CPA complex consists of the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulatory factor (CstF), cleavage factor I and IIm subcomplexes (Proudfoot, 2016). In vitro experiments show that the interaction between the CPSF, CstF, and Pol II pauses elongation and can gradually release the Poll II from the DNA template independent of the PAS-dependent cleavage of the mRNA (Zhang et al., 2015). This suggests that a simple change in the confirmation of the Pol II

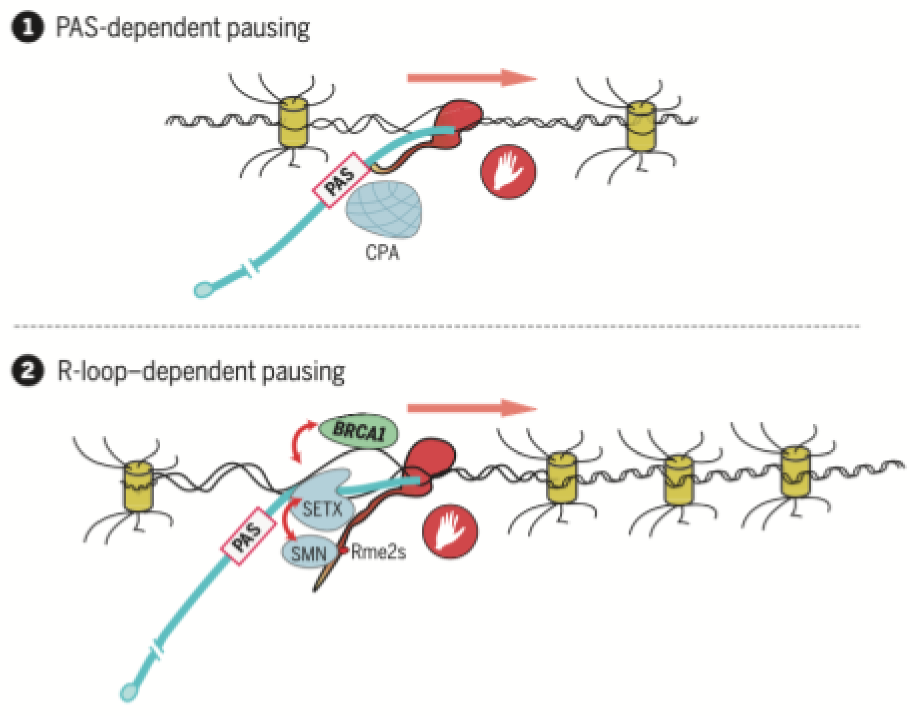

Figure 9. Shows two different types of Pausing on Pol II upon CPA recognition of the PAS. Pol II is shown in red, the nascent RNA is shown in the light blue, the barrels are nucleosomes with the red dots denoting methylation. Figure taken from *Proudfoot, 2016.*

is enough to cause termination (i.e. dissociation of Pol II). The longer the Pol II is paused over the PAS, the better it can interact with the CPA and cause eventual termination (Proudfoot, 2016).

Pol II can also be paused during elongation when it encounters nucleosomal barriers, as nucleosome-free DNA is transcribed more easily (Proudfoot, 2016). This pausing can lead to premature termination. Sometimes during transcription the newly synthesized mRNA can interact with the unwounded nontemplate strand of DNA to make DNA:RNA hybrids located outside of the elongation complex (Aguilera and García-Muse, 2012). This structure is called an R loop, and it can promote termination by slowing down Pol II transcription (see Figure 9, (2)). The R loop is a stable structure since it is more favorable than a DNA duplex, and forms at G rich sequences because the template strand can form stable G quadruplex structures (Skourti-Stathaki and Proudfoot, 2014). Splicing defects can cause R loop structures to form because this leads to a build up of the nascent transcript close to the unwound downstream DNA template (Proudfoot, 2016). An abnormally high number of R loops can cause DNA breakage and recombination leading to high amounts of DNA damage (Aguilera and García-Muse, 2012).

A helicase, Senataxin, is important in the removal of these structures and terminating transcription (Skourti-Stathaki et al., 2011). Senataxin can be recruited to resolve R loop structures in different ways. One is that a phosphorylation mark on the Arg1810 residue on the CTD recruits SMN (survival motor neuron protein) which recruits the Senataxin. In the second way BRCA1 (a DNA repair factor) is recruited to R

loops which rapidly recruits Senataxin for immediate termination and prevents DNA damage (Skourti-Stathaki et al., 2011).

Poly(A) site dependent termination

Poly(A) site termination involves specific cleavage of the nascent mRNA (Proudfoot, 2016). It is marked by a U-rich sequence upstream of the AAUAAA motif in the PAS, followed by a GU-rich sequence downstream. The CPA complex is recruited to the PAS by the CTD of Pol II. The CPA is assembled onto the nascent pre-mRNA as it moves outside of the exit channel of the active site of the Pol II (Proudfoot, 2016). The AAUAAA motif is recognized by the CPSF30 and the WDR33 components of the CPSF subcomplex. The endonuclease activity of the CPSF73 subunit then cleaves between the AAUAA and the downstream GU-rich element (Clerici et al., 2017). The cleaved nascent transcript is then polyadenylated (Proudfoot, 2016). There are two historical models that elaborate on the mechanism of the PAS-dependent termination of Pol II, and more recently a combined unified model has been proposed (West et al., 2004)

The first model that explains how termination occurs via the PAS is called the **allosteric** or **antiterminator model** (Figure 10) (West et al., 2004). The basis of this model is that the conformational changes within the active site of

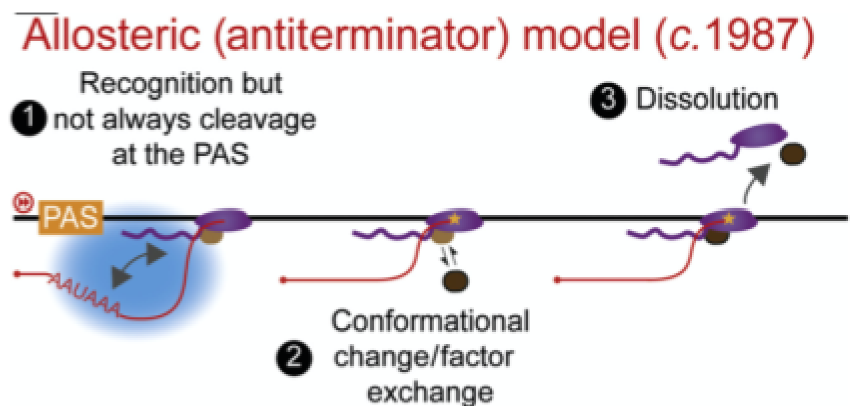

Figure 10. Allosteric model showing conformation change or factor exchange required for Pol II termination. Figure taken from *Eaton and West, 2020.*

the Pol II or the dissociation of proteins that prevent termination cause Pol II pausing and a release from the DNA template. The Pol II recognizes the PAS because of the CPA complex binding to the CTD. Since most genes in the eukaryotic genome have multiple PAS, antitermination factors like SCAF4 and SCAF8 prevent termination until the correct PAS has been transcribed (West et al., 2004). If these factors are not present, then premature cleavage and polyadenylation can take place in which a PAS meant to be a part of the gene body can be used as a termination site (Gregersen et al., 2019). Premature termination can also be caused by the depletion of other biomolecules including cyclin-dependent kinase (CDK) 12, U1 small nuclear RNA, or the nuclear poly(A) binding protein (PABPN1) (Eaton and West, 2020). High levels of PCF11 (a factor that can loosely be associated with the CPA machinery) can cause premature cleavage and polyadenylation as well (Kamieniarz-Gdula et al., 2019). Any proteins disassembled from the termination prevention machinery can cause premature termination as proposed by the antiterminator model of PAS-dependent termination (Proudfoot, 2016).

The second model that explains the PAS-dependent termination is called the **torpedo model** (Figure 11) (West et al., 2004). This model claims that Pol II continues to transcribe for a short period of time even after the cleavage of the nascent transcript. This

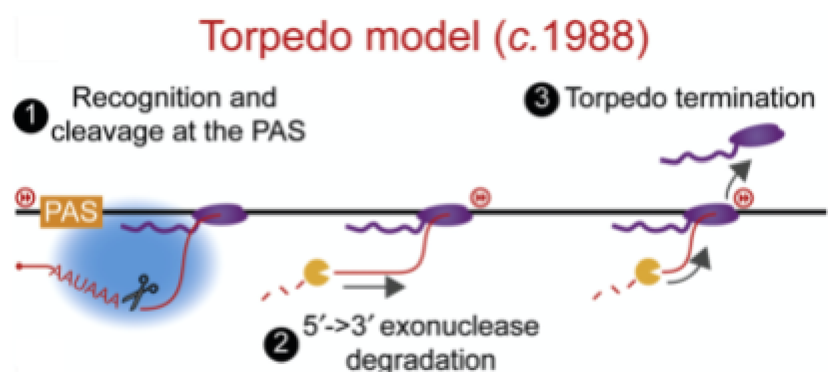

Figure 11. Shows PAS cleavage exonuclease 5' to 3' degradation of RNA leading to termination of Pol II. Figure taken from *Eaton and West, 2020.*

elongating transcript can invade the DNA duplex and produce R loops that decrease the rate of elongation as they cause torsional stress in the structure (Crossley et al., 2019). In addition, post cleavage, while the Pol II is continuing to elongate a the transcript, the 5' to 3' exonuclease Xrn2 is recruited to degrade the nascent RNA (West et al., 2004). Then, in kinetic competition with Pol II, Xrn2 reaches Pol II and causes its detachment from the DNA template triggering termination (Connelly and Manley). This kinetic competition determines where exactly the termination site will be. When Xrn2 was deleted from mammalian cells, there was a substantial loss of termination in these cells (West et al., 2004). Varying degrees of termination defects were observed with the removal of Xrn2 by RNAi technology (West et al., 2004). This suggests that other factors must also be involved in the removal of the Pol II from the DNA template.

These models can be unified to explain how antitermination factors/conformational changes and Xrn2 are both critical for PAS-dependent termination (Figure 12) (West et al., 2004). After passing through the PAS, Pol II's rate of transcription slows down significantly through the allosteric model, making it easier for the Xrn2 to torpedo it (Cortazar et al., 2019). The SPT5 elongation factor associates with Pol II and is phosphorylated when the polymerase is near the promoter

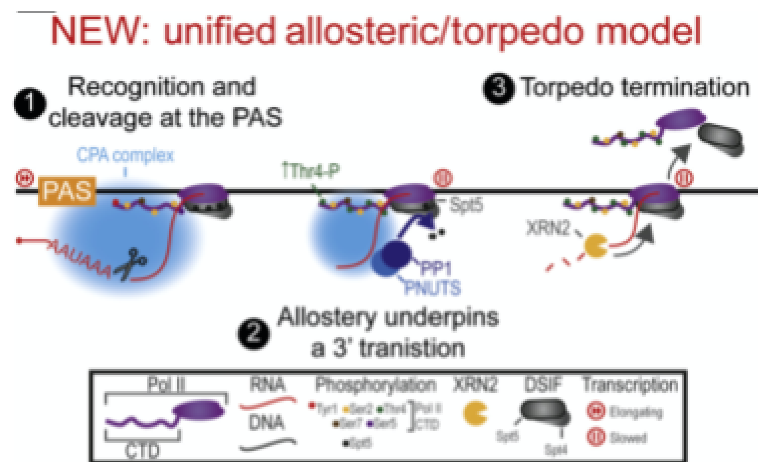

Figure 12. A combined allosteric/torpedo model caused by PAS cleavage of CPA assembly causing a dephosphorylation of elongation factors including Spt5 by PNUTS/PP1.

region by CDK9 to initiate efficient elongation (Cortazar et al., 2019). However, SPT5 is dephosphorylated just downstream of the PAS by protein phosphatase 1 (PP1) (Cortazar et al., 2019). PP1 and some of its nuclear targeting factors, like PNUTS, are present in the CPA complexes, which provides evidence that phosphorylation plays an important role in 3' end processing and termination (Shi et al., 2009). PP1 is also responsible for dephosphorylating p54nrb which is shown to help recruit Xrn2 to DNA (Kaneko et al., 2007). However, it is also shown that compared to the deletion of Xrn2, the deletion of CPA subunits causes an increase in the length of transcriptional read through of the proper 3' end. This suggests that the CPA complex would be more integral to proper termination and Xrn2 simply supports it (Eaton and West, 2020).

Hypothesis

In the past few years several studies have shown that in response to different types of cellular stress, termination by Pol II is disrupted (Bauer et al., 2018; Cardiello et al., 2018; Hennig et al., 2018; Vilborg et al., 2015). Pol II reads through the termination signals at the 3' ends of genes, and the mechanism isn't known. I hypothesize that **stress-induced termination defects in eukaryotic cells are linked to the phosphorylation state of the CTD in Pol II**. It is already known the phosphorylation of the CTD plays a role in recruitment of transcriptional regulatory proteins, and phosphorylation of some residues of the CTD (Tyr1, Ser2, and Thr4) have links to RNA processing and termination, although precise mechanisms are not clear. I will go over stress-induced termination defects in osmotic stress, viral infection, and heat shock.

Then, I will link CTD modifications, CTD kinases, and CTD phosphatases with termination, and finally, link how stresses can impact CTD kinases.

Stress Induced Termination Defects

In recent years, a lot of research has been done to characterize the correlation between transcriptional termination and environmental or cell stresses. In this section, I will summarize key findings from research done on four main types of stresses introduced in cells including HSV1 infection, influenza virus infection, hyperosmotic stress, and heat shock. Oxidative stresses, for example, have also shown to cause a readthrough of Pol II in many genes, though this will not be discussed in detail in this paper (Vilborg et al., 2017). Treatment of NIH 3T3 fibroblast cells with H_2O_2 showed an induction of widespread read through transcription (Figure 13, blue line). These stresses are often used to learn about the regulatory processes in central dogma. All of these studies show both a global loss of transcription and Pol II occupation along protein coding genes, and they show termination defects characterized by read through of the termination signals so transcription occurs downstream of genes (DoG). These read throughs

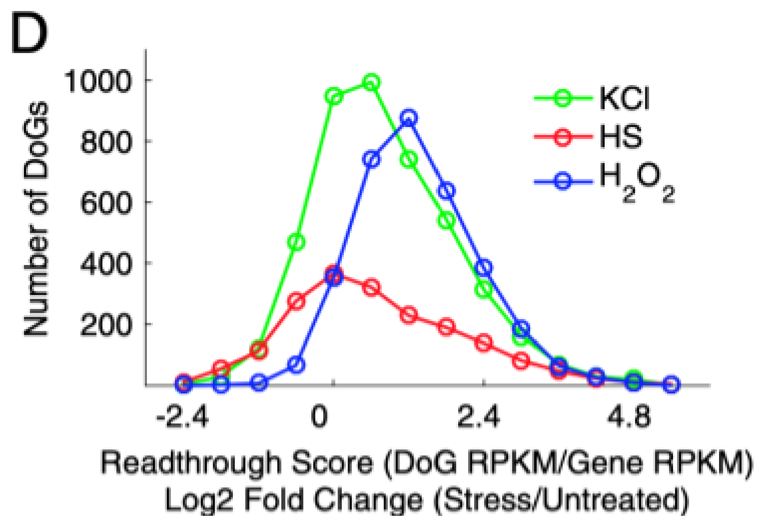

Figure 13. The blue line shows th relative number of readthroughs in oxidative stressed cells compared to KCl treated or heat shocked cells. RPKM gene's longest readthrough was used. Figure taken from *Vilborg et al. 2017.*

can cause the transcription of non-coding RNAs that might be fused to protein-coding RNA. Studies show that DoG transcripts usually remain in the nucleus close to the site of transcription (Hennig et al., 2018; Vilborg et al., 2015). This new area of research has taken advantage of chromatin immunoprecipitation sequencing (ChIP-seq), which shows a loss of Pol II peak after the PAS and Pol II in intergenic regions downstream of genes.

Hyperosmotic Stress

Hyperosmotic stress (salt treatment) is a common method to understand the mechanisms by which transcription is regulated due to stress. The first report of DoG transcripts was given by Vilborg et al. in 2015 in which they found that osmotic stress induced read through in more than 10% of protein coding genes. In a paper published by Rosa-Mercado et al. the effects of hyperosmotic stress were studied in relation to transcription termination. In this study, human embryonic kidney (HEK293T) cells were subjected to 80 mM KCl for 60 minutes. During the last 5 minutes of this salt exposure, 4-thiouridine was added to mark the RNA being actively transcribed (Schwalb et al., 2016). Then, transient transcriptome sequencing (TT-seq) coupled with TimeLapse (TL) chemistry was performed to understand the profiles of the nascent transcripts being transcribed. They found that transcription was globally repressed upon hyperosmotic stress and that DoG transcription takes place irrespective of the level of transcription in the upstream gene. This makes it hard to study whether transcription is being activated in a gene or if higher levels of RNA in a gene are due to read- in transcription from the upstream gene with a termination defect. This paper uses the term “clean gene” to

describe genes that do not overlap with readthrough transcripts and have a higher RNA levels in the gene body compared to 1 kb upstream of the TSS. 4,584 clean genes were identified (Rosa-Mercado et al., 2021).

Hyperosmotic stress led to a widespread repression of transcription. Quantitatively, Rosa-Mercado et al. observed a 3 fold decrease in transcription (Figure 14). Their studies showed that more than 88% of the clean genes were repressed and about 3% were activated. They also found that a termination defect took place independently of the upstream gene transcription level. About 13% of clean genes were DoG producing during hyperosmotic stress, with read counts in the downstream regions increased (Figure 15) (Rosa-Mercado et al., 2021).

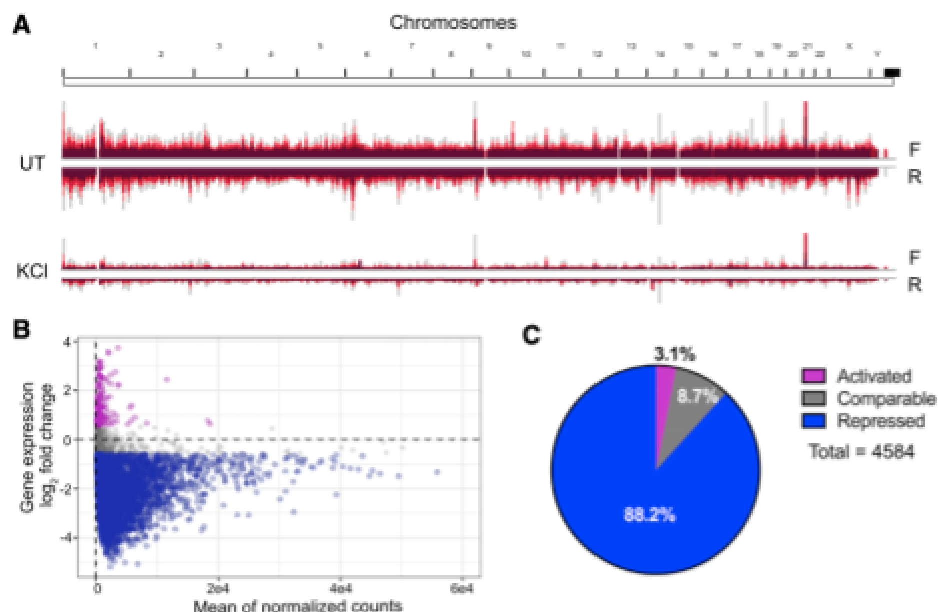

Figure 14. (A) Genome wide transcriptional repression caused by KCl treatment. (B) and (C) show that some genes are activated but most are repressed. Figure taken from *Rosa-Mercado et al. 2021*.

Another effect of salt stress is that it leads to the decline in the occupancy of the integrator complex along the genome (Rosa-Mercado et al., 2021). The integrator complex is important for 3’ end processing of non protein coding RNAs including snRNAs, enhancer RNAs, lncRNAs, and more (Lai et al., 2015). The catalytic subunit (Int11) interacts with Int9 to make the complex important for termination of these RNAs by inducing cleavage of the nascent RNA (Rosa-Mercado et al., 2021). However, Rosa-Mercado et al. showed that there is a decrease in interaction between Pol II and the integrator complex at mRNA genes. Also, a knockdown of Int11 through siRNA caused termination readthrough in many genes. This proves that hyperosmotic stress is linked to removal of the integrator complex from the DNA, which in turn causes a defect in transcription termination of hundreds of mRNA genes.

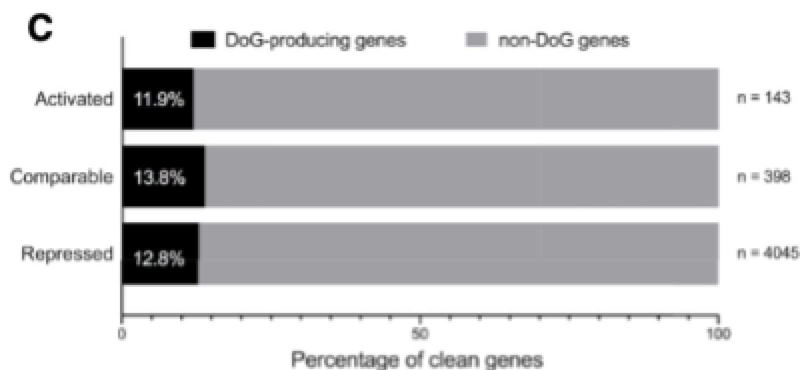

Figure 15. Bar graph shows the percent of DoG producing genes in each category of gene transcription. Figure taken from *Rosa-Mercado et al. 2021*.

HSV1 Infection

Herpes Simplex Virus 1 (HSV1) infection alters transcription and gene expression of the host cell through lytic infection. Henning et al. recently infected primary human

foreskin fibroblasts (HFF) with HSV1 and analyzed the host cell nascent RNA transcripts every hour within the first 8 hours of infection by labeling the newly transcribed RNA with 4sU and sequencing. Their results show that every hour after infection, 3' end read-through (a termination defect) takes place (Figure 16). HSV1 infected cells showed an increase in chromatin accessibility in the DoG regions after the affected poly(A) sites that correlated with leading to increased read-through transcription (Hennig et al., 2018).

Hennig et al. also compared HSV1 induced termination defects with termination defects induced by osmotic and heat stress in HFF cells using 4-thiouridine sequencing. They exposed the cells to either 80 mM KCl salt or heat shocked them for 1 and 2 hours at 44 degrees celsius. Figure 17 shows data for the SRSF3 gene and Figure 16 shows this trend genome-wide. Their results show transcription termination defects in all cases but a much more evident read-through was observed in the HSV1 infected cells. Intergenic read through transcription decreased as Pol II moved further downstream of the PAS (Hennig et al., 2018).

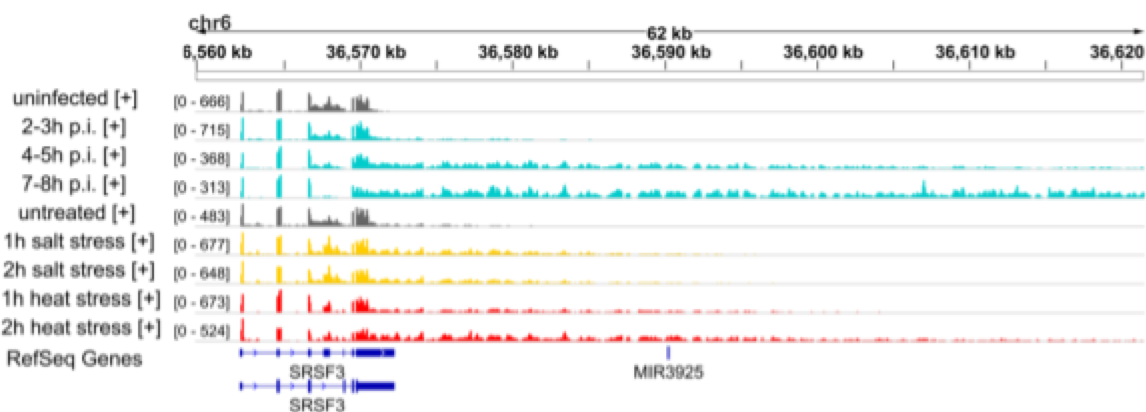

Figure 17. Shows DoG transcription in SRSF3 gene for HSV1 infected and heat stressed cells. As time of stress increases, DoG increases in both cases. Figure taken from *Hennig et al. 2018*.

Another interesting finding that Hennig et al. found was that the presence of the 6-mer consensus PAS (AAUAAA) at the 3' end of the gene did not correlate with read-through in all three different conditions of the cells. Having an AAUAAA 100 nucleotides upstream or downstream negatively correlated with readthrough of termination in HSV1 infected cells. The only 6-mer sequence that correlated with readthrough transcription termination was AUUUUU downstream of the 3' end. This sequence can bind some RNA binding proteins. Though the mechanism is not known, it is clear that HSV1 infected cells undergo a disruption in transcriptional termination.

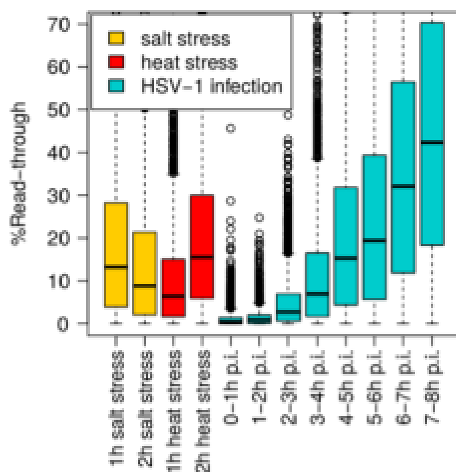

Figure 16. Boxplot showing readthrough genome wide in salt, heat, and HSV1 cells. Figure taken from *Hennig et al. 2018.*

Recent data suggest one of the HSV1 viral proteins (ICP27) might be causing the termination defect at the host cell genes (Wang et al., 2020).

Influenza Virus Infection

In an article published by Bauer et al. the effects of influenza virus infection on host transcription was studied. To study the behavior of Pol II during transcription in influenza virus infected cells, mammalian native elongating transcript sequencing (mNET-seq) was performed in human lung epithelial cells (A549 cells). They infected the cells with H1N1 influenza virus and performed mNET-seq to examine the result of the infection on Pol II across the genome (Figure 18 (ii)). Their results showed two major impacts of H1N1 infection on transcription. Firstly, they saw that there is a widespread loss of Pol II occupancy downstream of the TSS, consistent with

transcription repression.

Secondly, they observed a defect in transcription termination and interference in 3' end processing causing transcription to occur tens of kilobases downstream of the termination site (Bauer et al., 2018).

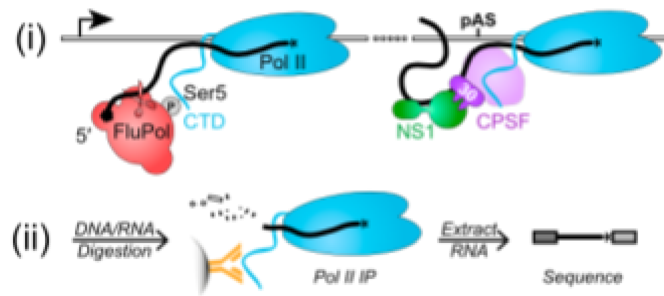

Figure 18. (i) shows the interference of FluPol in the host genome in the 5' end of the gene by bidning to Ser5P for its own viral RNA synthesis and near the PAS where it inhibits 3' end processing by binding to CPSF30. (ii) shows how mNET-seq works in this case genome wide. Figure taken from *Bauer et al. 2018.*

Influenza virus has an RNA-dependent RNA polymerase (FluPol) that carries out its transcription and replication in the host nucleus of the viral RNA (te Velthuis and Fodor, 2016). FluPol associates closely with the host Pol II as it is dependent on Pol II for the supply of 5' capped RNA transcripts which are processed to be used as primers in the virus by FluPol (Krug et al., 1979). FluPol is said to be linked to phosphorylated Ser5 (Ser5P) in the Pol II CTD, as it is the CTD modification which is marks initiating transcription and is in direct contact with the 5' capping enzymes (Lukarska et al., Martinez-Alonso et al., 2016). This allows the virus to take over the mRNA pool of the host and interfere with transcription.

In their study, Bauer et al. studied which genes underwent a failure in termination of host transcription upon infection by the virus. They studied the ratio of mNET-seq signal 2 kb before and after the PAS of each gene. This ratio increased after infection, showing termination defects in transcription of most protein coding genes besides histone coding genes. Since 3' ends of histone genes are not processed by the CPA complex (Kolev, 2005), it was concluded that influenza virus likely impacts termination of

protein coding genes by interfering with the CPA complex and 3' end processing (Figure 19). Another mechanism through which the virus is shown to cause termination defects is by expressing its nonstructural 1 (NS1) protein. NS1 of some strains of the influenza virus can interact with CPSF30 to interfere with 3' end processing (Nemeroff et al., 1998). By expressing NS1 protein in HEK293 cells, Bauer et al. observed that there was a similar effect on Pol II termination as the virus infection. They concluded that the two effects of H1N1 infection are two independent events (Figure 18 (i)). However, their results also showed that NS1 protein's direct interaction with CPSF30 is not the sole cause of termination defect upon the infection of the virus (Bauer et al., 2018).

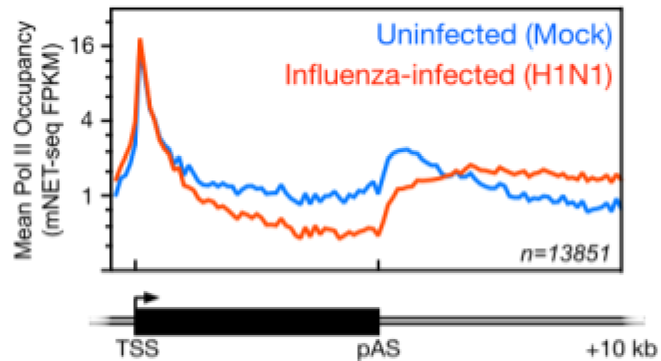

Figure 19. Pol II occupancy on all non-overlapping protein coding genes. Figure taken from *Bauer et al. 2018.*

Heat Shock

Cells contain mechanisms to respond to environmental stresses like heat shock, which are activated by stress-induced regulatory factors. The regulation of transcription in response to heat shock causes genes to either become activated or repressed (Vilborg et al., 2017). Genes that code for heat shock proteins are activated. These genes are upregulated by the transcriptional activator heat shock factor 1 (HSF1) (Vilborg et al., 2017). Studies done to determine the effects of heat stress in Drosophila

have shown a global decrease in transcription (Duarte et al., 2016). Similarly, a general repression in transcription was observed in thousands of genes following heat shock in human cells (Vihervaara et al., 2017). However, the mechanisms controlling transcriptional repression are not as well understood as the activation of heat shock protein genes. How heat shock impacts mechanisms of termination and the recruitment of the CPA complexes and their interaction with Pol II are still unknown.

In a paper published by Cardiello et al. transcriptional changes due to heat shock were observed. Chromatin immunoprecipitation sequencing (ChIP-seq) was done on Pol II in mouse NIH 3T3 cells to determine how Pol II occupancy is altered by heat stress. Three different conditions were tested, including a 15 minute heat shock at 45 degrees C (HS), a 15 minute heat shock at 45 degrees C followed by a return to homeostasis at 37 degrees C for 60 mins (60R), and a control group which did not undergo any type of heat shock (NHS).

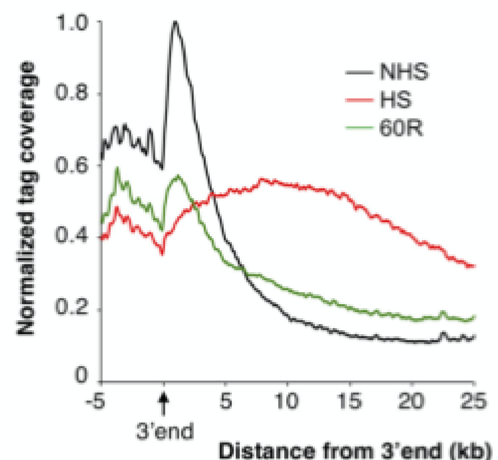

Figure 20. Averaged heat shock Pol II occupancy showed an increase in occupancy downstream of 3' end relative to NHS and 60R. A return to normal Pol II occupancy downstream of 3' end was observed in 60R. Figure taken from *Cardiello et al. 2018.*

Cardiello et al. showed that heat shock causes an alteration in transcription and the recovery from heat shock shows patterns of Pol II occupancy closer to the normal conditions. ChIP-seq data showed an alteration in Pol II occupancy across thousands of genes. Transcription was globally repressed due to heat shock and termination of transcription was strongly affected, observed by an increase in Pol II occupancy downstream of the 3' end of the

genes. However, there was a return to normal Pol II occupancy under the 60R condition as shown in Figure 20. ChIP-seq data showed that there was a global loss of termination as there was evident a broad peak downstream of the 3' end site. Upon quantitative analysis to see if heat shock causes termination defects 1,744 genes were considered filtered that did not have another gene 25 kb downstream (Cardiello et al., 2018). It was concluded that there is Pol II activity at least up to 25 kb downstream of the 3' end in these activated genes which is not present in NHS conditions. Finally, they observed that Pol II occupancy returns close to normal upon the recovery from HS for 60 minutes suggesting that this defect of termination by heat shock is a reversible process (Cardiello et al., 2018).

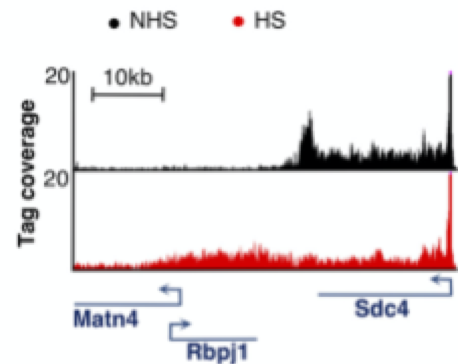

Figure 21. Transcription taking place from right to left shows that downstream genes of Sdc4 were falsely identified as activated because of readthrough of Sdc4. Figure taken from *Cardiello et al. 2018.*

Because of this read through of termination, there were over a thousand genes that were marked as activated because they were downstream of a termination defect gene (Cardiello et al., 2018). These trends are shown in the example of the Sdc4

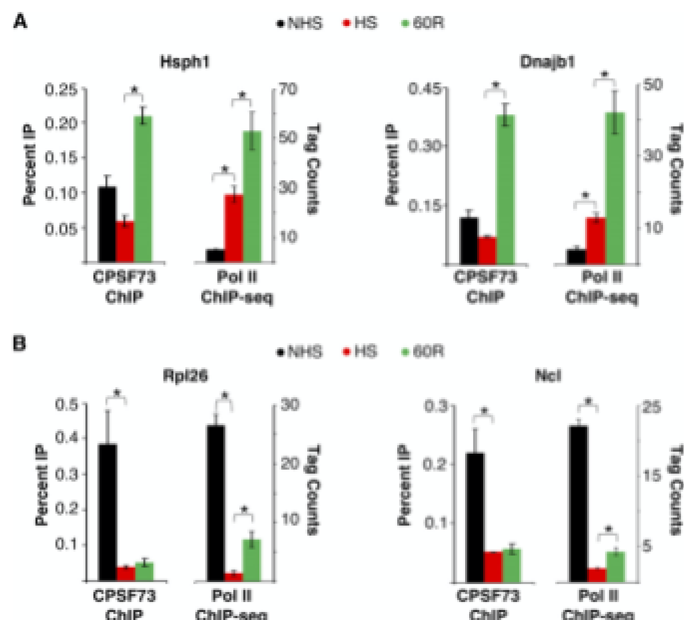

Figure 22. ChIP-qPCR data showing that CPSF73 recruitment to the 3' ends of 4 genes with termination defects is impaired after heat shock. Pol II ChIP signal is also shown, which reflects the activation or repression of each of these genes. Figure taken from *Cardiello et al. 2018.*

gene shown in Figure 21. Matn4 and Rbpj1 are neighboring genes of Sdc4. In the NHS condition, these genes did not have Pol II occupancy, however under the HS conditions, there was Pol II occupancy across these two neighboring genes as well (Figure 21). Termination could possibly be impaired in genes after HS due to the failure to recruit certain termination factors including CPSF73. Figure 22 shows a decrease in CPSF73 near the gene 3’ ends after heat shock.

Link Between Termination and CTD Modifications

The CTD is arguably the most important regulatory region of Pol II as it is proven that the modifications made on the CTD are linked to transcriptional regulation (Buratowski, 2009). Co-transcriptional processes such as 5’ capping, splicing, and 3’end processing are also all connected to the modifications on the CTD (Fong et al., 2017). The phosphorylation of specific residues of the CTD are shown to regulate the different stages of transcription and co-transcriptional processes. Of the seven repeated residues in the CTD, five are shown to be phosphorylated or dephosphorylated during transcription. Phosphorylation of Ser2 (Ser2P) and Ser5 (Ser5P) are the most studied. Studies have shown that Ser5P is linked primarily to transcription initiation (Komarnitsky, 2000). Ser5P is enriched near the transcription start site and in the promoter proximal peak. It is thought to play an important role in 5’ capping of the nascent RNA. Ser5P may be linked to splicing as well (Chathoth et al.). mNET-seq has also shown this for mammalian cells (Nojima et al., 2015). The mechanism is not yet clearly understood, however, initial studies have shown that parts of the spliceosome interact directly with the phosphorylated CTD (Morris and Greenleaf, 2000).

Termination of transcription, similarly, has shown some indication of being linked to CTD phosphorylation (Ahn et al., 2004; Davidson et al., 2014; Kuehner et al., 2011). Here, I will describe the phosphorylation of different residues of the CTD and how they are related to termination. I will focus on Serine 2, Tyrosine 1, and Threonine 4 phosphorylation (Ser2P, Tyr1P, and Thr4P, respectively) as comprehensive mNET-seq analyses have shown elevated levels of these phospho-marks at the 3' ends of genes (Nojima et al., 2015; Schlackow et al., 2017; Schüller et al., 2016).

Serine 2 Phosphorylation is Linked to Termination in Eukaryotes

Ser2P is present at higher levels as Pol II moves further away from the promoter (Buratowski, 2009). Levels Ser2P gradually increase as Pol II reaches the 3' end and peak near the PAS. Studies on the phosphorylation of Ser2 have been done in both yeast and mammalian cells. In a paper published by Ahn et al., they describe the importance of Ctk1, the catalytic subunit of CTDK1, which is a kinase important for phosphorylating the Ser2 residue in yeast cells. Ctk1 is required for the recruitment of 3' end processing factors. ChIP data showed that polyadenylation and cleavage factors were not localized near the 3' ends of the genes upon termination when Ser2P was disrupted by the knockdown of Ctk1 (Ahn et al., 2004). Ser2P is thus important in the recruitment of termination factors in yeast.

The human homolog of Ctk1 is Cdk9 (Rother and Strasser, 2007). Ser2P is phosphorylated by Cdk9 because the inhibition of Cdk9 severely decreases Ser2P in the CTD (Schüller et al., 2016). Cdk9, a subunit of the elongation factor P-TEFb, adds the Ser2P mark early in transcription, which later recruits the CPA complex needed for

termination (Ahn et al., 2004). Also important for Ser2 phosphorylation is Cdk12. This has been linked to recruitment of subunits of the CPA complex (Davidson et al., 2014; Tellier et al., 2020). Mutations introduced in Ser2 to replace it with an alanine have also shown an impairment in 3' end cleavage as well (Gu et al., 2013). Therefore current literature supports a strong connection between termination and proper levels of Ser2P.

Fcp1 is believed to be the phosphatase that dephosphorylates the Ser2P on the CTD (Mayfield et al., 2016). Newer research shows that another set of similar phosphatases (Scp1-3) could also play a role in dephosphorylating Ser2P, as they have an active site identical to that of Fcp1 (Mayfield et al., 2016). However, it is not well understood how dephosphorylation of Ser2P contributes to termination of transcription and 3' end formation.

It is also not well understood how the Ser2P recruits the CPA complex to carry out termination. Studies show that modifying levels of CPA complex subunits or the PAS changes Ser2P levels (Davidson et al., 2014). PAS mutations caused a lack of Pol II pausing at 3' ends and lack of hyperphosphorylation of Ser2 where it is normally located 1-2 kb downstream of the 3' end of the gene (Davidson et al., 2014; Kim et al., 2011). Knockdown of CPSF73 inhibits hyperphosphorylation of Ser2 suggesting a positive feedback mechanism of Ser2P recruitment of the CPA complex and the CPA complex triggering further phosphorylation of Ser2 (Fusby et al., 2015). The phosphorylation of the Ser2 residue in the CTD by Cdk12 or Cdk9 plays a major role in the recruitment of the termination machinery and the dephosphorylation of the Ser2 by Fcp1 begins to occur downstream of the 3' end site of the gene.

Tyrosine 1 is Linked to Termination in Eukaryotes

Tyr1P is mainly observed near the promoter region and at enhancer sites of genes in mammals (Descostes et al., 2014). However, Tyr1P is also present at the 3' ends of genes to a lesser extent (Shah et al., 2018). In yeast, the presence of Tyr1P found along the gene body was concluded to help prevent premature termination of transcription (Mayer et al., 2012). Mayer et. al found that phosphorylation of Tyr1 and its dephosphorylation occurs in the coding region of genes in yeast. It also showed that there is an increase in Tyr1P just downstream of the TSS and decreases about 180 nucleotides upstream of the PAS. Tyr1P could also enhance the interaction of the CTD with certain elongation factors. To find out whether Tyr1P impacts recruitment of factors to Pol II, proteins with CTD interacting domains (CID) were profiled for occupancy throughout the genome. It was found that the binding of the CID in Nrd1, Rtt103, and Pcf11 with the CTD was repressed by Tyr1P in vitro due to steric interactions. This signifies that Tyr1P could impair termination in vivo (Mayer et al., 2012). However, in humans Tyr1P does not appear to decrease near the PAS as in yeast (Shah et al., 2018), leaving the question of how does Tyr1P function in transcription termination in human cells?

The effects of phosphorylation of tyrosine 1 (Tyr1) are difficult to study because mutants of Tyr1 in mammalian cells resulted in the degradation of Rpb1 of Pol II (Mayer et al., 2012). Shah et al. recently overcame this hurdle by creating a mutant in which the last 75% of the repeats in the CTD had their Tyr1 residue replaced with a phenylalanine residue (YFFF mutant). Their experiments with this Pol II showed that Tyr1 is important in the recruitment of mediator and integrator complexes as well as in promoter-proximal

pausing and termination. The YFFF mutant showed an apparent transcriptional readthrough of 3' ends that was widespread and spanning up to hundreds of kilobases downstream of the PAS. It is not clear whether loss of Tyr1 itself or loss of the ability to control Tyr1P levels caused this effect. In wildtype (non mutant) cells, 3' readthrough has been shown to occur when various termination proteins such as Xrn2, CPSF73, and more are knocked down (Proudfoot, 2016). However, the YFFF mutant did not cause any interference in the association of these proteins with Pol II, so the mechanism of disrupted termination with this mutant is unknown.

Shah et al. showed that the Tyr1 mutations caused widespread termination defects in 3' ends of genes as well as antisense strand transcription at the 5' end of genes. Antisense, or divergent, transcription is caused by Pol II transcribing the opposite strand of DNA in the opposite direction from a gene's TSS. This phenomenon occurs at about half of the promoters in mammalian cells (Fenouil et al., 2012). Using RNA-seq, they quantified RNA levels and found a genome wide read through in divergent transcription as well (Shah et al., 2018). The PDCD6IP gene shows both these phenotypes of failed termination (Figure 23). This pattern was observed genome wide and a larger region

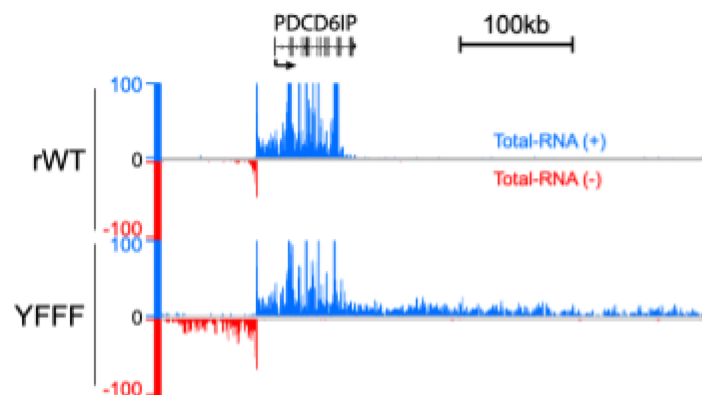

Figure 23. Shows an example of RNA-seq for a protein coding gene PDCD6IP showing 5' antisense (100 kb upstream) and 3' (300 kb downstream) readthrough. Figure taken from *Shah et al. 2018.*

of a

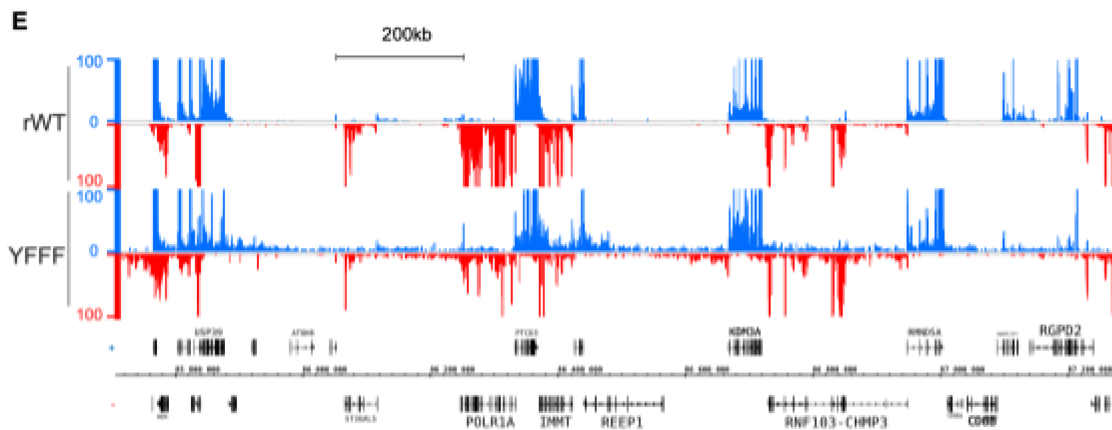

Figure 24. Chromosome 2 total RNA-seq showing massive readthrough generality at both 3' and 5' ends. Figure taken from *Shah et al. 2018.*

chromosome is shown as an example in Figure 24. Half of all mRNA genes showed a read through phenotype caused by a YFFF mutation (Shah et al., 2018).

Shah et al. data also show that polyadenylation occurs in these readthrough transcripts long after the occurrence of the PAS. Interestingly, however, despite the YFFF mutation, polyadenylation still occurs properly even if it occurs downstream of the PAS. So, polyadenylation and termination are uncoupled processes and the YFFF mutation does not interfere with the recruitment of the CPA complex downstream of the termination site. The mutation also impaired the interaction between Pol II and the integrator complex which can cause termination defects in many non protein coding RNAs (Shah et al., 2018).

Threonine 4 is Linked to Termination in Mammalian Cells

Threonine 4 (Thr4) is an essential residue of the CTD in mammalian cells. In yeast cells, replacing the Thr4 residue with a different amino acid, like alanine, did not prove to be lethal (Schwer and Shuman, 2011). However, the same mutation in human

cells proved to have severe consequences in cell growth. (Hintermair et al., 2012). Thereby indicating the phosphorylation of Thr4 (Thr4P) holds importance in the transcription cycle. Despite studies showing the importance of Thr4P, little is known about the mechanisms by which it controls transcription.

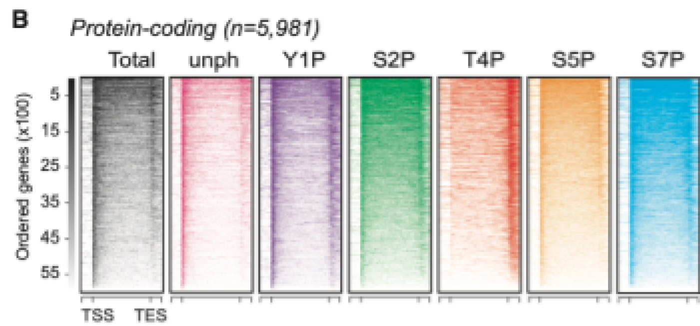

Figure 25. Color coded mNET-seq data showing phospho-CTD presence. Thr4P present strongly at the transcription end site. Figure taken from *Schlackow et al. 2017*.

Thr4 is important in elongation and termination. Mutating the threonine to alanine leads to a widespread defect in elongation (Harlen et al., 2016). Thr4P is known to be present in the gene body and at 3' ends of genes (Hintermair et al., 2012; Schlackow et al., 2017). This suggests that it could be important in termination of transcription (Figure 25). Interestingly, the signal for Thr4P increases at 3' ends and reaches a peak around 500 to 2000 nucleotides downstream of the 3' end of genes in human B-cells (Figure 26). This implies that Thr4P supports some structural changes in the Pol II downstream of the PAS. PLK3 phosphorylates Thr4 of the Pol II CTD in humans (Hintermair et al., 2012). However, Cdk9

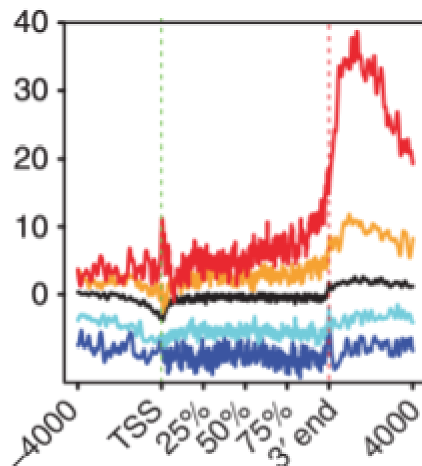

Figure 26. Genome wide Thr4P occupancy using average transcription unit analysis for genes > 4 kb in length (most of the protein coding genes). The different colors represent the various levels of Thr4P occupancy depending on the signal range. Figure is taken from *Hintermair et al. 2012*.

specific inhibitors have shown to also decrease Thr4P levels in humans (Hintermair et al., 2012). The exact mechanism of Thr4 phosphorylation and how it functions to mark different stages of transcription is still unknown.

Connection Between Cell Stress, Kinases, and the CTD

One of the main ways that the phosphorylation of CTD can be altered due to cell stress is if the cell stress impacts the role of the kinases responsible for phosphorylation. Not a lot of research has been done to understand the links between the cellular stresses that trigger termination defects and their impact on the phosphorylation state of the CTD. As discussed above, influenza and HSV1 infected cells, heat shocked cells, oxidative stressed cells, and osmotic stressed cells all show 3' read-through transcription throughout the genome with Pol II transcription reaching tens of kilobases downstream of the PAS. Termination defects have also been shown in cancer cells (Grosso et al., 2015), although the mechanisms of regulation could be quite different compared to cellular stress. Nonetheless, this implies that there might be a link between stress and cancer as well. It is possible that defects in termination might be caused by a stress-induced impact on the CTD modifying enzymes.

Cells have evolved to try to survive in any given environment and have many pathways through which they do so. There are also proteins that get activated by heat shock called heat shock proteins (Hsps) (Nadeau and Landry, 2007). Hsps are always expressed at a basal level, and specific hsps are bound to HSF prior to stress. Due to heat shock, Hsps dissociate from HSF, which allows it to translocate to the nucleus and activate transcription of Hsp genes that code to respond to heat shock (Nadeau and

Landry, 2007). In response to heat shock, in addition, mitogen-activated protein kinase (MAPK) pathways are activated in order to initiate signalling cascades that ultimately control transcription. This cascade has three layers which signals the activation of one pathway after the other (Nadeau and Landry, 2007). The purpose of this cascade is to ensure that transcription of genes upon exposure to heat shock occurs properly. So, cells have mechanisms which cause their survival, even under stressful conditions. Perhaps, activation of a kinase that could impact CTD phosphorylation and regulate the defect in termination is also one of these mechanisms which helps with cellular survival. Cdk7, which is a kinase subunit of TFIIH, showed a decrease in activity in heat shocked cells (Dubois et al., 1997). Cdk7 is responsible for phosphorylation of Ser5 on the CTD during early steps of transcription (Fisher, 2005). Unpublished data from our lab also shows that Cdk12 and cyclin K levels decrease at 3' end of the in Ncl gene in heat shock (Figure 27). So, heat shock alters co-transcriptional kinase activity, although the function and impact on transcription is not yet understood.

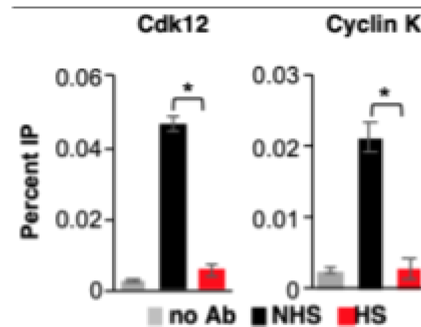

Figure 27. Unpublished Goodrich-Kugel data showing ChIP-qPCR of the 3' end of the Ncl gene indicating a loss of Cdk12 and cyclin K at 3' ends due to heat shock.

Furthermore, infection by influenza virus and HSV1 have mechanisms of controlling the CTD with the help of certain kinases. For example, Cdk9 can interact with the FluPol to enhance the FluPol's association with the Pol II complex (Zhang et al., 2010). NS1 of the virus can prevent termination by binding and disrupting the host CPSF30 interaction (Bauer et al., 2018). J Zhang et al. found that depleting human host

cells of cyclin T1 or Cdk9 completely disrupted the FluPol binding to the Ser2P CTD. This shows a direct link between the influenza virus and Cdk9 which is a kinase responsible for the phosphorylation of Ser2. So, Cdk9 serves as a mediator between the binding of FluPol and CTD. Similarly, Cdk9 is also involved in HSV1 transcription (Ou and Sandri-Goldin, 2013). They showed that inhibiting Cdk9 induced a global reduction in viral transcription levels as seen on microarray analysis. So, Ser2P is also required for HSV1 transcription, but how infection impacts Ser2P levels on the host genome isn't known. Studies have shown that HSV1 infection changes levels of CTD phosphorylation in general through an unknown mechanism (Wang et al., 2020).

It is known that oxidative stresses induce the Plk3 kinase to phosphorylate p53, a tumor suppressor protein (Xie et al., 2001). Hintermair et al. treated HeLa cells with hydrogen peroxide for 30 minutes to see whether there is an impact on Thr4P due to the upregulation of Plk3 by oxidative stress. They found that this treatment caused an increase in Thr4P levels in the CTD. They concluded that Plk3 can phosphorylate Thr4 in stress related occurrences (Figure 28). Additionally, there is a link between stress and the activation of c-Abl (Wang et al., 2006). In humans, the phosphorylation of Tyr1 is achieved by the kinase c-Abl which is potentially an important kinase for termination (Wang et al., 2006). Osmotic stress and ionizing radiation activate c-Abl (Kharbanda et al., 1995; Sun et al., 2000). Therefore some data supports links between cellular stresses and kinases that phosphorylate the CTD, however, much remains to be investigated regarding these relationships.

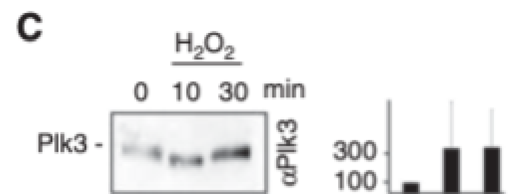

Figure 28. There is an increase in plk3 after a 30 minute treatment with H_2O_2. Figure is taken from *Hintermair et al. 2012*.

Stresses affect many kinases including those that phosphorylate the CTD. Therefore, it seems reasonable that this could be a link between cellular stress and termination defects.

Significance and Outlook

CTD phosphorylation state and its link to stress induced termination defects is an emerging area of research. The function, if any, of these termination defects is still to be unraveled. Since many factors influence transcription termination, there is likely a biological function for these defects. Termination defects may serve as a mechanism to help cells navigate stress. It might also be possible that cells reallocate their resources to fighting the stresses and transcription termination is not a top priority, and loss of termination does not create an insurmountable problem. Also, in many stress induced cells, global transcriptional repression and termination defects take place simultaneously. This could imply that a decrease in termination causes a reduction in Pol II recycling and thus contributes to repressing transcription globally. Termination defects have been found in cancer cells as well, implying that there might be certain cells that proliferate as a consequence of imparied termination, although they might occur through different mechanisms. Research to properly conclude whether or not the phosphorylation state of the CTD is influenced by stresses and in turn affects termination is currently being done. However, one likely cause of stress induced termination defects in mammalian cells is the phosphorylation of the CTD. There is so much more to learn about stress induced termination defect mechanisms but I hope to

illuminate this small area in order to contribute to the bigger picture of evolved cellular response to stresses.

Acknowledgements

I would sincerely like to thank Dr. James Goodrich and Dr. Jennifer Kugel for giving me the opportunity to be a part of their lab and assisting me with this thesis. I would also like to thank Kaitlyn Walsh for mentoring me through the process of writing my Honors Thesis. I would also like to thank Naiara Garcia, Tyler Brant, Tom Rivas, Georgiana Salant, Calvin Voong, Stephen Archuleta, and friends and family for their continued support.

References

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Aguilera, A., and García-Muse, T. (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. Mol. Cell *46*, 115–124.

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing. Mol. Cell *13*, 67–76.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., and Muzio, L.L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int. J. Mol. Med. 10.

Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. Nat. Rev. Mol. Cell Biol. *18*, 437–451.

Barber, K.W., and Rinehart, J. (2018). The ABCs of PTMs. Nat. Chem. Biol. *14*, 188–192.

Bauer, D.L.V., Tellier, M., Martínez-Alonso, M., Nojima, T., Proudfoot, N.J., Murphy, S., and Fodor, E. (2018). Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell Rep. *23*, 2119-2129.e3.

Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. Curr. Opin. Cell Biol. *14*, 336–342.

Buratowski, S. (2009). Progression through the RNA Polymerase II CTD Cycle. Mol. Cell *36*, 541–546.

Burgers, P.M.J., and Kunkel, T.A. (2017). Eukaryotic DNA Replication Fork. Annu. Rev. Biochem. *86*, 417–438.

Cardiello, J.F., Goodrich, J.A., and Kugel, J.F. (2018). Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. Mol. Cell. Biol. *38*, e00181-18, /mcb/38/18/e00181-18.atom.

Champoux, J.J. (2001). DNA Topoisomerases: Structure, Function, and Mechanism. Annu. Rev. Biochem. *70*, 369–413.

Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A Splicing-Dependent Transcriptional Checkpoint Associated with Prespliceosome Formation. Mol. Cell *53*, 779–790.

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

Connelly, S., and Manley, J.L. A functional .mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. 14.

Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D.L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. Mol. Cell *76*, 896-908.e4.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-Loops as Cellular Regulators and Genomic Threats. Mol. Cell *73*, 398–411.

Davidson, L., Muniz, L., and West, S. (2014). 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. Genes Dev. *28*, 342–356.

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. ELife *3*, e02105.

Dever, T.E., Dinman, J.D., and Green, R. (2018). Translation Elongation and Recoding in Eukaryotes. Cold Spring Harb. Perspect. Biol. *10*, a032649.

Dewar, J.M., and Walter, J.C. (2017). Mechanisms of DNA replication termination. Nat. Rev. Mol. Cell Biol. *18*, 507–516.

Down, T.A. (2002). Computational Detection and Location of Transcription Start Sites in Mammalian Genomic DNA. Genome Res. *12*, 458–461.

Drew, H.R., and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. J. Mol. Biol. *186*, 773–790.

Duarte, F.M., Fuda, N.J., Mahat, D.B., Core, L.J., Guertin, M.J., and Lis, J.T. (2016). Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes Dev. *30*, 1731–1746.

Dubois, M.-F., Vincent, M., Vigneron, M., Adamczewski, J., Egly, J.-M., and Bensaude, O. (1997). Heat-shock inactivation of the TFIIH-associated kinase and change in the phosphorylation sites on the C-terminal domain of RNA polymerase II. Nucleic Acids Res. *25*, 694–700.

Eaton, J.D., and West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! Trends Genet. *36*, 664–675.

Ekundayo, B., and Bleichert, F. (2019). Origins of DNA replication. PLOS Genet. *15*, e1008320.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I., et al. (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res. *22*, 2399–2408.

Fisher, R.P. (2005). Secrets of a double agent: CDK7 in cell-cycle control and transcription. J. Cell Sci. *118*, 5171–5180.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929–933.

Fong, N., Saldi, T., Sheridan, R.M., Cortazar, M.A., and Bentley, D.L. (2017). RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. Mol. Cell *66*, 546-557.e3.

Frolova, L.Yu., Merkulova, T.I., and Kisselev, L.L. (2000). Translation termination in eukaryotes:

Polypeptide release factor eRF1 is composed of functionally and structurally distinct domains. RNA *6*, 381–390.

Frouin, I., Montecucco, A., Spadari, S., and Maga, G. (2003). DNA replication: a complex matter. EMBO Rep. *4*, 666–670.

Furuichi, Y. (2015). Discovery of m7G-cap in eukaryotic mRNAs. Proc. Jpn. Acad. Ser. B *91*, 394–409.

Fusby, B., Kim, S., Erickson, B., Kim, H., Peterson, M.L., and Bentley, D.L. (2015). Coordination of RNA Polymerase II Pausing and 3' end processing factor recruitment with alternative polyadenylation. Mol. Cell. Biol. MCB.00898-15.

Gehring, N.H., and Roignant, J.-Y. (2021). Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. Trends Genet. *37*, 355–372.

Glover, B.P., and McHenry, C.S. The DNA Polymerase III Holoenzyme: An Asymmetric Dimeric Replicative Complex with Leading and Lagging Strand Polymerases. 10.

Gregersen, L.H., Mitter, R., Ugalde, A.P., Nojima, T., Proudfoot, N.J., Agami, R., Stewart, A., and Svejstrup, J.Q. (2019). SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. Cell *177*, 1797-1813.e18.

Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vítor, A.C., Desterro, J.M., Carmo-Fonseca, M., and de Almeida, S.F. (2015). Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. ELife *4*, e09214.

Gu, B., Eick, D., and Bensaude, O. (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic Acids Res. *41*, 1591–1603.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. *19*, 621–637.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep. *15*, 2147–2158.

Hennig, T., Michalski, M., Rutkowski, A.J., Djakovic, L., Whisnant, A.W., Friedl, M.-S., Jha, B.A., Baptista, M.A.P., L'Hernault, A., Erhard, F., et al. (2018). HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. PLOS Pathog. *14*, e1006954.

Herhaus, L., and Dikic, I. (2015). Expanding the ubiquitin code through post-translational modification. EMBO Rep. *16*, 1071–1083.

Hershey, J.W.B., Sonenberg, N., and Mathews, M.B. (2019). Principles of Translational Control. Cold Spring Harb. Perspect. Biol. *11*, a032607.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., et al. (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation: CTD Thr4 is required for transcription elongation. EMBO J. *31*, 2784–2797.

Hong, S., Sunita, S., Maehigashi, T., Hoffer, E.D., Dunkle, J.A., and Dunham, C.M. (2018). Mechanism of tRNA-mediated +1 ribosomal frameshifting. Proc. Natl. Acad. Sci. *115*, 11226–11231.

Hsin, J.-P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. *26*, 2119–2137.

Hsu, L.M. (2009). Monitoring abortive initiation. Methods *47*, 25–36.

Hu, H., and Sun, S.-C. (2016). Ubiquitin signaling in immune responses. Cell Res. *26*, 457–483.

Jiannan Guo, Tiandao Li, David H Price (2013). Runaway transcription. Genome Biol. *14*.

Kamieniarz-Gdula, K., Gdula, M.R., Panser, K., Nojima, T., Monks, J., Wiśniewski, J.R., Riepsaame, J., Brockdorff, N., Pauli, A., and Proudfoot, N.J. (2019). Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Mol. Cell *74*, 158-172.e9.

Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. (2007). The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. Genes Amp Dev. *21*, 1779–1789.

Kharbanda, S., Ren, R., Pandey, P., Shafmant, T.D., Feller, S.M., and Weichselbaum, R.R. (1995). Activation of the c-Abl tyrosine kinase in the stress response to DNA-damaging agents. *376*, 4.

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proc. Natl. Acad. Sci. *108*, 13564–13569.

Knoll, E.R., Zhu, Z.I., Sarkar, D., Landsman, D., and Morse, R.H. (2018). Role of the pre-initiation complex in Mediator recruitment and dynamics. ELife *7*, e39633.

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

Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R., and Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin *5*, 1.

Komarnitsky, P. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. *14*, 2452–2460.

Krug, R.M., Broni, B.A., and Bouloy, M. (1979). Are the 5′ ends of influenza viral mrnas synthesized in vivo donated by host mRNAs? Cell *18*, 329–334.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. Nat. Rev. Mol. Cell Biol. *12*, 283–294.

Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. Nature *525*, 399–403.

Lukarska, M., Fournier, G., Pflug, A., Resa-Infante, P., Reich, S., Naffakh, N., and Cusack, S. (2017). Structural basis of an essential interaction between influenza polymerase and Pol II CTD. Nature *541*, 117–121.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science *336*, 1723–1725.

Mayfield, J.E., Burkholder, N.T., and Zhang, Y.J. (2016). Dephosphorylating eukaryotic RNA polymerase II. Biochim. Biophys. Acta BBA - Proteins Proteomics *1864*, 372–387.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357–361.

Mehmood, A., Laiho, A., Venäläinen, M.S., McGlinchey, A.J., Wang, N., and Elo, L.L. (2020). Systematic evaluation of differential splicing tools for RNA-seq studies. Brief. Bioinform. *21*, 2052–2065.

Merrick, W.C. (1992). Mechanism and Regulation of Eukaryotic Protein Synthesis. MICROBIOL REV *56*, 25.

Morris, D.P., and Greenleaf, A.L. (2000). The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. J. Biol. Chem. *275*, 39935–39943.

Nadeau, S.I., and Landry, J. (2007). Mechanisms of Activation and Regulation of the Heat Shock-Sensitive Signaling Pathways. In Molecular Aspects of the Stress Response: Chaperones, Membranes and Networks, P. Csermely, and L. Vígh, eds. (New York, NY: Springer New York), pp. 100–113.

Nemeroff, M.E., Barabino, S.M.L., Li, Y., Keller, W., and Krug, R.M. (1998). Influenza Virus NS1 Protein Interacts with the Cellular 30 kDa Subunit of CPSF and Inhibits 3′ End Formation of Cellular Pre-mRNAs. Mol. Cell *1*, 991–1000.

Neugebauer, K.M. (2002). On the importance of being co-transcriptional. J. Cell Sci. *115*, 3865–3871.

Nikolov, D.B., and Burley, S.K. (1997). RNA polymerase II transcription initiation: A structural view. Proc. Natl. Acad. Sci. *94*, 15–22.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526–540.

Ohkuma, Y., Hashimoto, S., Wang, C.K., Horikoshi, M., and Roeder, R.G. (1995). Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha. Mol. Cell. Biol. *15*, 4856–4866.

Ou, M., and Sandri-Goldin, R.M. (2013). Inhibition of cdk9 during Herpes Simplex Virus 1 Infection Impedes Viral Transcription. PLoS ONE *8*, e79007.

Pain, V.M. (1996). Initiation of Protein Synthesis in Eukaryotic Cells. Eur. J. Biochem. *236*, 747–771.

Pereira, S.F.F., Goss, L., and Dworkin, J. (2011). Eukaryote-Like Serine/Threonine Kinases and Phosphatases in Bacteria. Microbiol. Mol. Biol. Rev. *75*, 192–212.

Pickart, C.M. (2001). Mechanisms Underlying Ubiquitination. Annu. Rev. Biochem. *70*, 503–533.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926–aad9926.

Pufall, M.A., and Kaplan, C.D. (2013). Mechanisms of eukaryotic transcription. Genome Biol. *14*, 311.

Rahimi, N., and Costello, C.E. (2015). Emerging roles of post-translational modifications in signal transduction and angiogenesis. PROTEOMICS *15*, 300–309.

Ranish, J.A., Yudkovsky, N., and Hahn, S. (1999). Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. Genes Dev. *13*, 49–63.

Robison, A.J., and Colbran, R.J. (2013). Calcium/Calmodulin-Dependent Protein Kinases. In Encyclopedia of Biological Chemistry, (Elsevier), pp. 304–309.

Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., and Steitz, J.A. (2021). Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. Mol. Cell *81*, 502-513.e4.

Rother, S., and Strasser, K. (2007). The RNA polymerase II CTD kinase Ctk1 functions in translation elongation. Genes Dev. *21*, 1409–1421.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Mol. Cell *65*, 25–38.

Schneider-Poetsch, T., and Yoshida, M. (2018). Along the Central Dogma—Controlling Gene Expression with Small Molecules. Annu. Rev. Biochem. *87*, 391–420.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol. Cell *61*, 305–314.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228.

Schwer, B., and Shuman, S. (2011). Deciphering the RNA Polymerase II CTD Code in Fission Yeast. Mol. Cell *43*, 311–318.

Schwertman, P., Bekker-Jensen, S., and Mailand, N. (2016). Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. Nat. Rev. Mol. Cell Biol. *17*, 379–394.

Shah, N., Maqbool, M.A., Yahia, Y., El Aabidine, A.Z., Esnault, C., Forné, I., Decker, T.-M., Martin, D., Schüller, R., Krebs, S., et al. (2018). Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. Mol. Cell *69*, 48-61.e6.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J., and Manley, J.L. (2009). Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. Mol. Cell *33*, 365–376.

Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to

genome integrity and powerful regulators of gene expression. Genes Dev. *28*, 1384–1396.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Mol. Cell *42*, 794–805.

Sun, X., Wu, F., Datta, R., Kharbanda, S., and Kufe, D. (2000). Interaction between Protein Kinase C δ and the c-Abl Tyrosine Kinase in the Cellular Response to Oxidative Stress. J. Biol. Chem. *275*, 7470–7473.

Tellier, M., Zaborowska, J., Caizzi, L., Mohammad, E., Velychko, T., Schwalb, B., Ferrer-Vicens, I., Blears, D., Nojima, T., Cramer, P., et al. (2020). CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. Nucleic Acids Res. *48*, 7712–7727.

Uversky, V.N. (2013). Posttranslational Modification. In Brenner’s Encyclopedia of Genetics, (Elsevier), pp. 425–430.

Vámosi, G., and Rueda, D. (2018). DNA Bends the Knee to Transcription Factors. Biophys. J. *114*, 2253–2254.

te Velthuis, A.J.W., and Fodor, E. (2016). Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. Nat. Rev. Microbiol. *14*, 479–493.

Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. Crit. Rev. Biochem. Mol. Biol. *44*, 117–141.

Vihervaara, A., Mahat, D.B., Guertin, M.J., Chu, T., Danko, C.G., Lis, J.T., and Sistonen, L. (2017). Transcriptional response to stress is pre-wired by promoter and enhancer architecture. Nat. Commun. *8*, 255.

Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2015). Widespread Inducible Transcription Downstream of Human Genes. Mol. Cell *59*, 449–461.

Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., and Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc. Natl. Acad. Sci. *114*, E8362–E8371.

Wang, J.Y.J., Minami, Y., and Zhu, J. (2006). Abl and Cell Death. In Abl Family Kinases in Development and Disease, (New York, NY: Springer New York), pp. 26–47.

Wang, X., Hennig, T., Whisnant, A.W., Erhard, F., Prusty, B.K., Friedel, C.C., Forouzmand, E., Hu, W., Erber, L., Chen, Y., et al. (2020). Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. Nat. Commun. *11*, 293.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522–525.

Whetstine, J.R. Histone Methylation. 9.

Woychik, N.A., and Hampsey, M. (2002). The RNA Polymerase II Machinery: Structure Illuminates Function. Cell *108*.

Wu, J., Xiao, J., Zhang, Z., Wang, X., Hu, S., and Yu, J. (2014). Ribogenomics: the Science and

Knowledge of RNA. Genomics Proteomics Bioinformatics *12*, 57–63.

Xie, S., Wang, Q., Wu, H., Jhanwar-Uniyal, M., Dai, W., Cogswell, J., and Lu, L. (2001). Reactive Oxygen Species-induced Phosphorylation of p53 on Serine 20 Is Mediated in Part by Polo-like Kinase-3. J. Biol. Chem. *276*, 36194–36199.

Yamaguchi, Y., Shibata, H., and Handa, H. (2013). Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829*, 98–104.

Yao, N.Y., and O'Donnell, M.E. (2018). Replication fork convergence at termination: A multistep process. Proc. Natl. Acad. Sci. *115*, 237–239.

Zhang, Y. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev. *15*, 2343–2360.

Zhang, H., Rigo, F., and Martinson, H.G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Mol. Cell *59*, 437–448.

Zhang, H.-M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H., and Guo, A.-Y. (2012). AnimalTFDB: a comprehensive animal transcription factor database. Nucleic Acids Res. *40*, D144–D149.

Zhang, J., Li, G., and Ye, X. (2010). Cyclin T1/CDK9 Interacts with Influenza A Virus Polymerase and Facilitates Its Association with Cellular RNA Polymerase II. J. Virol. *84*, 12619–12627.

Zhou, A.Q. Exploration of the Central Dogma at the Interface of Chemistry and Biology. 3.

Zhu, W., Zeng, Q., Colangelo, C.M., Lewis, L.M., Summers, M.F., and Scott, R.A. (1996). The N-terminal domain of TFIIB from Pyrococcus furiosus forms a zinc ribbon. Nat. Struct. Biol. *3*, 122–124.