The role of RNA Polymerase II-dependent transcription elongation in the cross-talk between mRNA synthesis and decay

Victoria Begley

Doctoral thesis

2018
## INDEX

1. **Introduction** .................................................................................................................. 1
   Gene expression .................................................................................................................. 3
   Transcription .................................................................................................................... 3
     1. Initiation: ....................................................................................................................... 5
     2. Elongation: .................................................................................................................... 6
     3. Termination: .................................................................................................................. 10
   mRNA degradation ............................................................................................................ 11
     The main deadenylator: the Ccr4-Not complex .............................................................. 13
     The RNA exonuclease Xrn1 .......................................................................................... 14
   From transcription to decay: mRNA imprinting .............................................................. 15
   From mRNA decay to transcription .................................................................................. 16
   Biological significance of the transcription/decay cross-talk .......................................... 18
2. **Objectives** .................................................................................................................... 21
3. **Results** ......................................................................................................................... 25

Chapter 1: Molecular characterization of mutants potentially affecting the transcription/mRNA decay crosstalk .................................................................................................................. 27
   Mutants show slight variation in gene expression that is exacerbated by longer gene
   length .................................................................................................................................... 31
   The selected mutants show altered Pol II occupancy and active Pol II .............................. 34
   Total mRNA concentrations are not dependent on any single variable studied ............. 43
   Growth rate correlates with transcription and degradation ............................................. 45
   Cell volume negatively correlates with transcription and degradation .......................... 46
   Pol II speed is important for the cross-talk mechanism ................................................. 49
   Degradation rates and synthesis rates correlate .............................................................. 51
   Materials and methods .................................................................................................... 55

Chapter 2: Mathematical Modelling and computational simulation of the cross-talk between mRNA transcription and decay .................................................................................. 61
   How to model such a complex system (ODD protocol) .................................................... 65
   Definition of the system and its agents ............................................................................ 65
   Programming the system .................................................................................................. 68
   Model variants .................................................................................................................. 71
   First results ......................................................................................................................... 73
Chapter 3: Genome-wide analysis of the contribution of Xrn1 to transcription elongation using CRAC .......................................................... 81

The deletion of XRN1 led to a reduction in transcription rates ......................... 84
RNAs labelled with 4-thiouracil cannot be efficiently degraded by the exosome in an xrn1Δ mutant ................................................................. 94
There is a global increase in arrested Pol II when Xrn1 is rapidly depleted from the cell ... 97
Forcing Xrn1 out of the nucleus led to changes in early Pol II elongation but not in late elongation ................................................................. 105
Materials and methods .................................................................................. 112

4. Discussion .................................................................................................. 119

The characterization of transcription by various different methods enables a better understanding of the transcriptional defects present in each mutant .................. 121
mRNA degradation is not only determined by function ..................................... 123
Correlations between mRNA turnover with growth rate and cell volume validate our results. .............................................................................. 124
Pol II elongation rate can influence mRNA degradation ..................................... 125
mRNA degradation can influence transcription elongation ............................... 126
Metagene analysis showed defects in transcription elongation in the absence of Xrn1 .... 127
The controversy of the role of Xrn1 in the cross-talk between mRNA decay and transcription. .............................................................................. 128
All results confirm a strong cross-talk between transcription and mRNA decay. ........ 130
Working model ............................................................................................... 131
The Cross-talk beyond transcription and mRNA decay ..................................... 135

5. Conclusions ............................................................................................. 137

6. Bibliography ............................................................................................. 141
1. Introduction.
Gene expression.

Gene expression is the process by which information contained in the DNA is converted into a functional gene product that dictates cell function. These products are often proteins, but in non-protein coding genes the product is a functional RNA. The most important molecule in gene expression is RNA, whose main job is to transfer the genetic code from the nucleus, where the DNA is, to the cytoplasm, where it can be translated into protein by ribosomes. The RNA levels in a cell are a result of the balance between its synthesis (transcription) and its degradation. Thus, these processes play a critical role in determining what proteins are present in a cell and in what amounts. Therefore, to understand how gene expression works, we need to understand how these two processes work.

Transcription.

Conventionally, transcription is regarded as one of the most important steps in the RNA life cycle since it is responsible for the synthesis of the transcript itself. In addition, by processing the 5’ and 3’ ends of the pre-mRNA, the transcription machinery prepares it for correct export, translation and degradation in the cytoplasm (Svetlov and Nudler, 2013).

The substrate of transcription is not naked DNA but chromatin. The nucleosome is the basic unit of DNA coiling in eukaryotic chromosomes that compacts the genome in the form of chromatin. It is composed of 145-147 base pairs of DNA wrapped around a histone octamer containing two copies of histones H2A, H2B, H3 and H4 (Luger et al., 1997). Chromatin contains a repeating array of nucleosomes that are spaced every 200 ± 40 base pairs throughout the genome. Following nucleosome disruption by multiprotein complexes, the transcription machinery can interact with the uncoiled DNA and begin the transcription process (Kornberg, 2001). Transcription of the eukaryotic genome is carried out by the nuclear RNA polymerase I (Pol I), Pol II and Pol III. Whereas Pol I transcribes the ribosomal RNA genes (rRNA); Pol II transcribes messenger RNA (mRNA) and a subset of non-coding RNAs (ncRNA); and Pol III transcribes transfer RNAs (tRNAs), 5S rRNA and the bulk of ncRNAs (Sainsbury et al., 2015). These three polymerases present a similar structure and share a 10 subunit core with specific subunits on the periphery (Cramer et al., 2008). Pol II transcription is the first step in gene expression, therefore we will focus on the study of this polymerase.

Yeast Pol II is a multiprotein complex of 514 kDa consisting of twelve subunits (Cramer, 2004). These proteins are coded by the genes RPB1-12 and the two specific subunits of this polymerase are the heterodimer Rpb4/7. All RPB genes are essential for cell viability except for
**RPB4** and **RPB9** (Woychik and Young, 1989; Woychik et al., 1991). The crystal structure of transcribing Pol II (Gnatt et al., 2001) revealed that Rpb1 and Rpb2 form a centre cleft, with the smaller subunits arrayed around the periphery. In this cleft is the active site, marked by an Mg$^{2+}$ ion, where the biogenesis of nascent RNA takes place. Comparison with a higher resolution crystal structure later revealed that the polymerase is divided into four mobile modules, including a clamp that swings over the active centre (Cramer et al., 2014). It also revealed another metal ion at the active site that could possibly be exchangeable during RNA synthesis. Duplex DNA is thought to enter the main cleft or “jaw” and unwind before the active site. RNA synthesis then takes place, leading to a 9bp DNA-RNA hybrid that is held in place by the clamp. After passing through the active site, the DNA path is blocked by a protein “wall” forcing it to pass up the wall at nearly right angles to the incoming DNA in the cleft. The nascent RNA and the transcribed DNA abandons the polymerase through an exit channel. A pore in the floor of the cleft beneath the active site allows entry of substrate nucleoside triphosphates (NTPs) and also allows exit of RNA during retrograde movement of the polymerase on the DNA.

**Figure 1. Side view of the RNA polymerase II transcribing complex** (Klug, 2001).

In light grey the front of the enzyme is shown, and in dark grey the back part. The duplex DNA, in blue and green, can be seen entering the polymerase through the “jaw”. The template strand coding for RNA is in blue, the non-template strand is in green. The RNA in the DNA-RNA hybrid in the active centre region is in red. The 3’ (growing) end of the RNA is located adjacent to an active site Mg$^{2+}$ ion (pink sphere). Opposite to the entering site of DNA, a protein “wall” forces the DNA-RNA to make a nearly right angle turn to exit the enzyme. The NTPs enter through a funnel-shaped opening on the underside of the enzyme and gain access to the active centre through a pore.

One distinct feature of Pol II is the presence of a C-terminal extension on its largest subunit (Rpb1), referred to as the C-terminal domain (CTD). It consists of tandem repeats of the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, repeated 26 times in yeast. Each amino acid in the CTD can be post-translationally modified. The CTD is highly conserved and essential for viability in all organisms (Corden, 2013). During transcription, the CTD is
phosphorylated at specific amino acids depending on the progression along the gene. These phosphorylations serve as a signature of the different Pol II transcription steps, and they create binding sites for chromatin regulators and RNA processing factors that are required in specific stages of transcription (Suh et al., 2016). The combination of posttranslational modifications in the CTD is known as the CTD code (Buratowski, 2003).

The transcription process by Pol II can be divided into three steps: initiation, elongation and termination.

1. **Initiation:**

   For initiation, Pol II assembles with the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIIE, TFIIF and TFIIH at promoter DNA to form the pre-initiation complex (PIC) (Grünberg and Hahn, 2013). GTFs cooperate with Pol II to bind to and open promoter DNA, to initiate RNA synthesis and stimulate the escape of Pol II from the promoter (Sainsbury et al., 2015). The PIC assembles in nucleosome-free promoter regions that are flanked by an upstream −1 nucleosome and a downstream +1 nucleosome (Venters and Pugh, 2010). In yeast, the transcription start site (TSS) is located in the upstream border of the +1 nucleosome, suggesting that this nucleosome potentially regulates access to the TSS (Albert et al., 2007).

   During PIC assembly, a dephosphorylated Pol II is recruited to promoters through interactions with TFIIB and the Mediator Complex (Myers et al., 1998). This complex serves as a central scaffold within the PIC, and the different mediator subunits make contact with various transcription factors, including activators, co-activators, GTFs, and subunits of Pol II (Ansari and Morse, 2013). The first DNA contact of the GTFs occur at the TATA or TATA-like gene promoters. Both promoters are located at a fixed, short distance from the TSS that is recognized by the TATA-binding protein (TBP) (Venters and Pugh, 2010). However, on the one hand, the TATA promoter is made up of a consensus sequence (TATA(A/T)A(A/T)(A/G)), and on the other hand, the TATA-like promoters possess a sequence with two or less mismatches to the TATA box consensus (Sung Rhee and Franklin Pugh, 2012). Approximately 80-90% of all *Saccharomyces cerevisiae* genes are TATA-like, and in this case the TBP was thought to be a part of the TFIID complex which guided it to gene promoters. For the rest of genes (10-20%) the SAGA (Spt-Ada-Gcn5-acetyltransferase) complex was thought to direct TBP to TATA box containing Pol II promoters (Bhaumik and Green, 2002). Thus, the expression of most mRNAs was considered to be dominated by either the transcription factor TFIID or SAGA. However, recently this notion has been challenged. Two publications show that both transcription factors are necessary for nearly all yeast genes, and the differences observed before are just
differential dependence on SAGA or TFIID (Baptista et al., 2017; Warfield et al., 2017). Genome-wide analysis showed that TATA-containing promoters are highly regulated and generally stress responsive, with expression levels being extremely high or low. In contrast, TATA-like promoters are those genes referred to as housekeeping genes and are more constitutively active with less regulation requirements (Hahn and Young, 2011).

The first step in canonical PIC assembly is the sequence-specific binding of TBP to promoter DNA, inducing a bend. Then, the GTFs TFIIA and TFIIB stabilize the TBP-DNA complex by flanking TBP on both sides. Subsequently, Pol II associated to TFIIF joins the resulting complex and finally, TFIIE and TFIIH bind and complete the PIC. Pol II cannot melt DNA itself, although it can maintain an open transcription bubble. Thus, TFIIE and TFIIH are required to melt the DNA in the presence of ATP (figure 2). Once the template DNA is opened and melted, RNA synthesis begins (Sainsbury et al., 2015). In early transcription, abortive initiation occurs with frequency, resulting in short unstable transcripts. When the transcript reaches 10 nucleotides, promoter escape is favoured over abortive initiation and the complex stabilizes. However, a transcript length of a minimum of 25 nucleotides is necessary to achieve productive initiation and accordingly, transcription elongation can commence (Saunders et al., 2006).

![Figure 2. Scheme of the PIC assembly and initiation of transcription.](image)

The initiation competent Pol II-TFIIF complex joins the initial PIC formed by TFIID, who contains TBP, TFIIA and B. TBP recognizes the promoter of the gene and induces a bend in the DNA. Next, TFIIE and TFIIH bind to complete the PIC and melt DNA forming the transcription bubble. Once this occurs, transcription can begin.

2. **Elongation:**

The Kin28 subunit of TFIIH mediates the transition from transcription initiation to elongation through phosphorylation of serines in the 5th position of the CTD (Valay et al., 1995). When approximately 25 nucleotides of nascent RNA have been synthesized, the cap structure is attached to its 5’ end. The capping enzyme (guanylyltransferase) specifically recognizes the phosphorylated Ser-5 CTD repeats, which then catalyzes the addition of a methylguanosine cap to the 5’ end of nascent mRNA, which is a signal for productive transcription initiation.
(Komarnitsky et al., 2000). The cap protects the nascent RNA from degradation, and marks the mRNA for transport to the cytoplasm (Venters and Pugh, 2010). At this point, all GTFs, except TFIIF, are released (Liu et al., 2013).

Once Ser-5 is phosphorylated and the emerging mRNA is capped, further signals are required to promote the transition to productive elongation. Stimulation of early elongation involves P-TEFb (Positive Transcription Elongation Factor b) mediated counteraction of the negative effects of DSIF (DRB sensitivity inducing factor) and NELF (negative elongation factor) in mammals (Yamaguchi et al., 1999). In *Saccharomyces cerevisiae*, the DSIF complex is formed by Spt4 and Spt5, which can physically interact with Pol II (Hartzog et al., 1998). However, the NELF complex seems to be absent in yeast. A definite yeast homologue of P-TEFb has not yet been identified but two candidate complexes have emerged. The closest relative, based on a phylogenetic computer analysis, is the Bur1p/Bur2p CDK/cyclin complex. Bur1p phosphorylates Ser-5 of the CTD in an *in vitro* kinase assay and does not phosphorylate Ser-2 (Murray et al., 2001). The second putative homologue of P-TEFb in yeast, CTDK-I (Carboxy-Terminal domain Kinase), consists of three subunits, Ctk1, Ctk2 and Ctk3, and can also phosphorylate the CTD in vitro (Sterner et al., 1995).

As Pol II escapes the promoter and elongation proceeds from the 5′ end of the gene towards the 3′ end, a phosphorylation switch occurs on the CTD from Ser-5 to Ser-2, although it is not clear to what extent one event triggers the other (Venters and Pugh, 2010). Ctk1 is the main kinase involved in Ser-2 phosphorylation (Ahn et al., 2009), which begins to saturate around 600-1000 nucleotides downstream from the TSS and sharply decreases around 100-200 nucleotides downstream from the poly(A) addition site (Mayer et al., 2010). The CTD Ser-5 phosphorylation mark is erased primarily by the Ssu72 (Suppressor of Sua72) phosphatase and the Ser-2 phosphorylation mark by Fcp1, a major CTD phosphatase conserved in eukaryotes (Krishnamurthy et al., 2004). Dephosphorylation of the CTD is necessary to regenerate initiation competent Pol II (Kobor and Greenblatt, 2002).

Productive transcription elongation also relies on chromatin remodelers and histone chaperones to slide or evict histones/nucleosomes from the DNA. Chromatin remodelers, like RSC, destabilize nucleosomes during transcription (Smolle et al., 2012) or prevent histone exchange, like Isw1 and Chd1 (Smolle et al., 2012). Histone chaperones, like FACT, on the other hand, contribute to the disassembly and reassembly of nucleosomes during transcription elongation (Formosa, 2012). Another histone chaperone discovered in our laboratory in yeast is the prefoldin complex (Millán-Zambrano et al., 2013). This study showed evidence of a role
of prefoldin in histone eviction after the co-transcriptional destabilization of nucleosomes. In vitro Pol II has been found to transcribe through hexasomal nucleosomes following the eviction of a single H2A/H2B dimer, while the H3/H4 tetramer is retained on the DNA. Accordingly, H2A/H2B dimers have been found to be rapidly exchanged in and out of existing nucleosomes over all transcribed regions, and, in contrast, histone exchange of H3/H4 tetramers is limited to highly transcribed genes (Thiriet and Hayes, 2005). Complete dissociation of histone octamers from the DNA over coding regions is predominant in highly transcribed genes such as TATA genes, which are subsequently reassembled in the wake of Pol II passage (Smolle and Workman, 2013).

Posttranslational modifications (PTMs) of histone tails are also important for transcription elongation. Histone tails consist of 20-40 amino-terminal amino acids, which extend from the globular nucleosome core. They are subjected to a great number of posttranslational modifications (PTMs), including acetylation, methylation and ubiquitination. These modifications provide docking sites for other proteins that contribute to the structural dynamics of chromatin. For each modification, there is a corresponding complex that is able to erase it. This is consistent with the dynamic nature of chromatin modifications (Venters and Pugh, 2010).

Pol II is believed to oscillate by Brownian motion between no translocation, a forward translocation (in the direction of transcription) and a backward translocation (backtracking) at every step of transcription. When there is no translocation, the polymerase remains the same and a nucleotide is added to the growing RNA chain. Forward translocation clears the active site for the entry of the next NTP and backtracking extrudes the nucleotide just added from the active centre (Liu et al., 2013).

Transcriptional pausing is a critical feature of gene regulation (Landick, 2006). It is a frequent phenomenon that in 75% of cases (Churchman and Weissman, 2011) involves Pol II backtracking (figure 3A), where the catalytic site becomes disengaged from the 3’ end of RNA, rendering the elongation complex inactive, but stable (Nudler et al., 1997). The displaced backtracked RNA exits the Pol II out the funnel domain, also called the backtracked site, where it becomes trapped (Cheung and Cramer, 2011). This state can be resolved by the elongation factor TFIIS (figure 3B), encoded by the gene DST1 in S. cerevisiae (Izban and Luse, 1992). TFIIS stimulates intrinsic hydrolysing activity of the polymerase, which cleaves the backtracked 3’ stretch of the nascent RNA, allowing transcription to restart from an upstream point. TFIIS has been shown to relieve promoter proximal pausing (Adelman et al., 2005) which usually occurs
within the first ~50 nucleotides of transcribed sequences. Pausing not only occurs near promoters, but throughout the body of genes (Churchman and Weissman, 2011). This pausing along genes have been associated with nucleosome positioning (Bintu et al., 2012), regulatory factors (Churchman and Weissman, 2011), the interaction between nucleic acids and Pol II (Cheung and Cramer, 2011), and the structure of nascent RNA (Hein et al., 2014).

Figure 3. Scheme showing Pol II backtracking and its subsequent reactivation. A. The polymerase moves forward during transcription elongation. Due to the Brownian movement of the polymerase, it can move backwards and the 3’ end of the RNA becomes misaligned with the active site, exiting into the backtracked site in the funnel domain of Pol II. In this event the polymerase becomes paused as it cannot continue transcribing. B. TFIIS can rescue this arrested situation by stimulating the hydrolytic activity of the polymerase and releasing the 3’ end of RNA. C. Ccr4-Not was also found to relieve Pol II backtracking in a different manner to TFIIS. Three possible modes of action has been suggested for this complex: 1. By binding to the polymerase and transcript and preventing Pol II backtracking. 2. By binding to the backtracked polymerase and transcript and realigning the 3’ end of the transcript and the active site. 3. By cooperating with TFIIS. The possibility that Ccr4-Not performs all three modes simultaneously cannot be ruled out as they are not mutually exclusive.

Recently, the Ccr4-Not complex was identified as a direct regulator of backtracked Pol II (Kruk et al., 2011). The mechanism by which Ccr4-Not affects Pol II backtracking was shown to be different from that of TFIIS (figure 3C), with which it might cooperate (Dutta et al., 2015). One possibility is that it rescues backtracked Pol II by binding the polymerase and the transcript, restricting its movement through the exit channel, and realigning the 3’ end into the active centre. Another possibility is that the binding of the complex to the polymerase and transcript causes it to be less susceptible to backtracking. Perhaps the complex can do both, as these functions are not mutually exclusive.
Another factor that has been suggested to be implicated in Pol II backtracking is Sfp1, a global regulator of genes encoding ribosomal components (Gómez-Herreros et al., 2012a). In our lab, we attributed a possible transcriptional function to Spf1 upstream of TFIIS that, contrary to TFIIS, would favour Pol II backtracking. This role for Sfp1 was identified thanks to a genetic screening for suppressors of the lethality of dst1Δ in the presence of drugs that reduced the pool of NTPs. Pol II pausing and backtracking has also been associated with transcription termination because this mechanism provides a free RNA 3’ end for the core exosome to degrade (Lemay et al., 2014).

3. Termination:

Transcription termination is a complicated process whose correct execution is not only important for the gene in question, but also for the transcription of downstream genes (Greger and Proudfoot, 1998). Transcriptional termination differs from transcriptional initiation in that it does not have a consensus sequence on which to occur, but rather it happens in a zone downstream from fairly loosely defined terminator elements in the DNA template (Svejstrup, 2004).

There are two termination pathways for Pol II-dependent transcription in yeast: the CPF-CF (cleavage and polyadenylation factor-cleavage factor) pathway that is specifically assigned to mRNA, and the NNS (Nrd1-Nab3-Sen1) pathway that is specific to ncRNA termination (Porrua and Libri, 2015). In this first pathway, several components of the CPF–CF complex recognize termination and processing signals in the 3’ untranslated region (UTR) of the nascent RNA (figure 4). Subsequently, the RNA is cleaved at the poly(A) site, and adenosine nucleotides are added to the free hydroxyl group on the 3’ end (3’OH) (Mischo and Proudfoot, 2013). There are two models to explain how transcription terminates once the RNA is cleaved. The most accepted model is the “torpedo model” (Fong et al., 2015), where the exposed 5’ end still in the polymerase is rapidly degraded by Rat1, a 5’-3’ RNA exonuclease that triggers the release of Pol II. The newly formed poly(A) tail is bound by a poly(A)-binding protein (Pab1), which promotes nuclear export of the mRNA (Dunn et al., 2005). In addition to poly(A)-coupled termination, the dismantling of elongation complexes can be induced by road-block factors, like Reb1 (Colin et al., 2014). In the second pathway, the NNS complex is responsible for transcription termination for genes encoding snRNAs and snoRNAs, and for cryptic unstable transcripts (CUTs). These RNAs are degraded rapidly after transcription by the nuclear exosome. Timely termination and degradation of these ncRNAs is important for preventing their interference with the coding transcriptome (Porrua and Libri, 2015).
Figure 4. Transcription termination of mRNAs. The CPF-CF complex recognizes A-rich and U-rich stretches in the 3’ untranslated region (UTR) and subsequently cleaves the RNA at the poly(A) site. The resulting RNA is polyadenylated on the 3’ end. In the torpedo model, Rat1 degrades the exposed resulting 5’ end of the RNA still bound to the polymerase which leads to the release of Pol II from the template DNA.

Once the mRNA is transcribed and matured it is transported to the cytoplasm. In order for the transportation to occur, the nascent mRNA must be assembled with RNA binding proteins forming an mRNA ribonucleoparticle (mRNP) (Luna et al., 2005). The failure to form this particle leads to nuclear retention of the mRNA and can trigger its decay in the nucleus. Recruitment of the Mex67-Mtr2 heterodimer is crucial for targeting export-competent mRNPs to the nuclear pore complex (NPC) (Stewart, 2010). This heterodimer is recruited through adaptor proteins who ensure its stable interaction with the mRNPs. These adaptor proteins include the THO complex (Chávez et al., 2000), the RNA helicase Sub2 and the export adaptor protein Yra1 (Rondón et al., 2010). Many of them are also important factors of other mRNA biogenesis steps, allowing optimal coupling of transcription elongation with mRNP packaging and 3’ processing steps.

mRNA degradation.

Degradation is the main regulator of posttranscriptional gene expression because it eliminates mRNA from the cell. The process modulates the length of time an mRNA molecule is available for translation, and therefore, modulates protein production. Degradation, with the help of the surveillance machinery, also ensures that only correctly processed mRNAs are translated into protein. If the error in mRNA is not recognized by one of the cytoplasmic surveillance machineries (nonsense-mediated decay (NMD), no-go decay (NGD) or non-stop decay (NSG)), it could potentially be translated into a toxic protein which would be detrimental for the cell (Siwaszek et al., 2014).

Once the aberrant mRNAs have been filtered, correctly processed mRNAs bearing a 7-MeG cap at their 5’ end are recognized by eukaryotic initiation factor 4E (eIF4E). eIF4E forms a complex
with eIF4F and recruits the small 40S ribosome, initiator tRNA\textsuperscript{Met} and other initiation factors to begin translation (Altmann and Linder, 2010).

mRNA molecules synthesized in the nucleus by Pol II are relatively unstable, mRNA precursors are degraded by the Xrn1 paralog, Rat1, or the nuclear exosome complex. In the cytoplasm, they are protected from the attack of exonucleases by the 5’-cap structure and the 3’-poly(A) tail. The cytoplasmic bulk mRNA degradation pathway in eukaryotic cells starts with shortening of the poly(A) tail (figure 5). In yeast, deadenylation is catalyzed primarily by the Ccr4-Not with a lesser contribution of the Pan2–Pan3 deadenylase, however there is probably a significant redundancy and functional overlap (Parker and Song, 2004). Deadenylation is the first and therefore often rate-limiting step of mRNA decay (Wahle and Winkler, 2013), and its removal triggers mRNA degradation in yeast (Siwaszek et al., 2014). Following deadenylation, the mRNA can be degraded from the 3’ end by the exosome (Chlebowski et al., 2013), resulting in an oligonucleotide whose cap is removed by a salvage pathway catalyzed by Dcs1.

![Figure 5. mRNA decay pathways in the cytoplasm.](image)

There are two major cytoplasmic pathways to degrade mRNA that both start with the shortening of the poly(A) tail by one of the deadenylase complexes (Ccr4-Not or Pan2/3). The exosome can degrade the deadenylated mRNA in the 3’-to-5’ direction, and the scavenger decapping enzyme, Dcs1, degrades the remaining cap. The most prominent pathway involves decapping by Dcp1 and Dcp2, aided by the auxiliary factors Dhh1 and Pat1, and subsequent degradation by the 5’-to-3’ exonuclease, Xrn1 (Braun and Young, 2014).

Deadenylation also promotes decapping (Nissan et al., 2011). The Lsm 1–7 proteins bind to the 3’ end of the mRNA and recruit the Dcp1-Dcp2 decapping complex. The decapping enzymes, aided by auxiliary factors, in particular Dhh1 and Pat1, hydrolyze the 5’ cap, exposing the
mRNA to decay that is carried out by Xrn1, a 5’-3’ exoribonuclease. This is the major degradation pathway that occurs in the cytoplasm. Decapping and the 5’-3’ degradation are coupled, as Xrn1 nuclease interacts directly with Pat1. Some mRNAs can be degraded without the removal of the poly(A) tail. These mRNAs are aberrant and are subjected to the NMD surveillance system. Xrn1 degrades these aberrant mRNAs following 5’ cap removal by Dcp2 initiated by the NMD factors (Conti and Izaurralde, 2005). Deletions of the main enzymes of either of the degradation pathways do not result in the total accumulation of aberrant mRNA, which suggests that these enzymes work in cooperation (Siwaszek et al., 2014).

**The main deadenylator: the Ccr4-Not complex.**

The Ccr4-Not complex is essential and present in all eukaryotes. Structurally, it consists of 9 core subunits in yeast: Ccr4, Caf1, Caf40, Caf130, and Not1-5 (Collart and Panasenko, 2012). This core complex is between 0.9-1.2 MDa in size. It has been described, however, that the complex can also contain a combination of Dhh1, Dbf2, Caf4, Caf16, and Btt1 forming a larger complex of >1.9 MDa (Miller and Reese, 2012). In 2001, the core complex was purified by the Not1 subunit and viewed by electron microscopy (Nasertorabi et al., 2011). This allowed visualization of the global shape of Ccr4-Not that revealed an L-shaped complex with two arms of similar length (figure 6).

Ccr4, an endonuclease–exonuclease–phophatase (EEP), is the main deadenylase in the complex. Ccr4 is incorporated into the whole complex through Caf1 via a leucine-rich repeat motif (LRR) (Siwaszek et al., 2014). Caf1 possess an exonuclease motif (RNase D) and its deletion shows defects in deadenylation in vivo, similar to that of ccr4Δ (Tucker et al., 2001). However, it is dispensable for catalysis in vitro (Tucker et al., 2002). In this context, two different roles have been proposed for the Caf1 protein: first, it may be responsible for deadenylation under some circumstances given the conservation of some of the key residues in the active site. And second, it may enhance the function of Ccr4 by stabilizing the complex
or providing additional interactions with the mRNA. This was proposed because in the absence of Caf1 a defect in deadenylation was found, however, strains containing a full-length but catalytically inactive Caf1 showed no phenotype (Parker and Song, 2004). The Pan2-Pan3 complex is also involved in deadenylation in yeast and it seems that the activity of both deadenylating complexes is partially redundant. Deletion of ccr4 gives slower deadenylation rates, but only the ccr4Δ pan2Δ double mutant accumulates poly(A) tailed mRNAs and has severe growth defects (Tucker et al., 2002).

Not1 is the scaffold of the complex and in S. cerevisiae it is the only subunit essential for viability (Maillet et al., 2000). It is the largest subunit of the complex and does not possess enzymatic activity. Caf130 and Caf40, which are also incorporated into the yeast complex, possess no clear function. Not2, Not3, Not4 and Not5 form a Not module interacting with the C-terminus of Not1 (Bhaskar et al., 2013). Not2, Not3 and Not5 are closely associated, and together they create a platform that may contribute to the functions of the Ccr4-Not complex, including deadenylation. Finally, Not4 is an active E3 ubiquitin ligase. It is involved in protein turnover regulation and has been found to have a specific role in protein quality control that extends beyond regulation of deadenylation (Halter et al., 2014).

**The RNA exonuclease Xrn1.**

Yeast Xrn1 is a large, nonessential, highly conserved cytoplasmic 5’-3’ exoribonuclease whose well defined biochemical function is to degrade mRNA containing a free 5’ monophosphate end (Nagarajan et al., 2013a). Xrn1 degrades a wide range of cytoplasmic RNAs, including NMD-substrates, noncoding RNAs and a novel class of long non-coding RNAs (IncRNAs) called XRN1-sensitive Unstable-Transcripts (XUTs) (Wery et al., 2016). Deletion of Xrn1 leads to stabilization of mRNAs on a global scale, consistent with its role in the major cytoplasmic mRNA decay pathway (Haimovich et al., 2013a; Sun et al., 2013; Wery et al., 2016).

Xrn1 possesses an essential paralog that degrades RNAs in the nucleus named Rat1. Xrn1 and Rat1 are highly conserved in their N-terminal nuclease domain and are functionally interchangeable: yeast Rat1 can substitute for the function of Xrn1 when localized to the cytoplasm, while Xrn1 tagged with a nuclear localization signal (NLS) can rescue the lethal phenotype of a temperature-sensitive Rat1 mutant, rat1-1ts (Johnson, 1997). However, both Xrn1 and Rat1 are involved in processing of nuclear structural RNA (snoRNA and pre-rRNA) (Petfalski et al., 1998). This suggests that Xrn1 may be able to locate to the nuclear or nucleolar compartment.
From transcription to decay: mRNA imprinting.

Two yeast Pol II proteins, Rpb4 and Rpb7, separate from the core and form a heterodimer that shuttles between the nucleus, where it functions in transcription, and the cytoplasm (Silitrennik et al., 2006), where it functions in the major mRNA decay pathway (Lotan et al., 2005, 2007). Rpb4/7 binds Pol II transcripts cotranscriptionally and is associated with the mRNA throughout its life (Goler-Baron et al., 2008), modulating its translation (Harel-Sharvit et al., 2010) and degradation. Efficient translation in the cytoplasm depends on association of Rpb4/7 with Pol II in the nucleus, therefore, Pol II can be considered a key regulator of the major stages of gene expression and acts as a “mRNA coordinator” (Haimovich et al., 2013b).

Another identified “mRNA coordinator” is Pab1 that, as mentioned previously, is involved in the 3′-end processing of mRNA. It then regulates the export of bound mRNA (Brune et al., 2005), translation initiation (Kahvejian et al., 2005), translation termination (Hosoda et al., 2003), deadenylation and decapping (Wiederhold and Passmore, 2010). Impairing the import of Pab1 into the nucleus delays mRNA export out of the nucleus (Brune et al., 2005). Thus, the nuclear and cytoplasmic roles of Pab1 are thought to be coupled.

The yeast Ccr4-Not complex has also been found to function both in transcription and degradation (figure 7). The complex was initially identified as a transcription factor (Collart and Struhl, 1994; Draper et al., 1994). Later, it became apparent that the complex possessed many other functions such as nuclear RNA processing (Azzouz et al., 2009), mRNA export (Kerr et al., 2011), translation quality control and protein degradation (Collart and Panasenko, 2012) and, most importantly as one of the two major deadenylases (Tucker et al., 2002). Ccr4-Not mutants display many phenotypes observed in the rpb4 deleted strain, therefore, suspicions were formed as to whether the complex could also coordinate transcription and degradation.

In a recent paper, a connection was made between Rpb4/7 and Ccr4-Not (Babbarwal et al., 2014). They suggest that Rpb4/7 is required for Ccr4-Not to bind elongation complexes and rescue arrested Pol II, thus heightening these suspicions. Decisively, an imprinting mechanism through the Ccr4-Not complex was recently discovered (Gupta et al., 2016). In this model Not1 imprinting plays an important role in coupling different steps of gene expression, facilitating both translation of the imprinted mRNA, specifically ribosomal protein and mitochondrial genes, and its subsequent decay. This Not1 imprinting is both dependent on Not5 and transcription elongation, with more imprinting occurring in the presence of Not5 and with higher backtracking levels of Pol II. Therefore, they concluded that the complex is indeed linked to mRNA coordination, from transcription to mRNA decay.
Figure 7. The many functions of the Ccr4-Not complex connecting transcription, mRNA degradation and translation. 1. Ccr4-Not has been found to be involved in transcription initiation. 2. It is involved in transcription elongation. 3. It has been described to be involved in RNA processing and mRNA imprinting. 4. In mRNA export through the nuclear pore complex (NPC). 5. It is involved in translation associated to translating ribosomes. 6. In mRNA degradation by deadenylation. 7. In protein quality control. 8. And in the translation of Pol II.

From mRNA decay to transcription.

The Ccr4-Not complex has also been linked to coordinate mRNA decay to its transcription (figure 7). Once the complex has finished its role in deadenylation, it can shuttle into the nucleus where it has been implicated in various aspects of transcription. On one hand there is evidence indicating it is involved in transcription initiation. First of all, Ccr4-Not subunits are present at promoters (Venters et al., 2011) and tethering of Ccr4-Not complex subunits to promoters can activate or repress transcription (Benson et al., 1998). Second, Not4 can regulate the recruitment of certain transcription factors to promoters (Gulshan et al., 2012).
Finally, subunits of the Ccr4-Not complex show both genetic and physical interactions with TFIID and SAGA subunits (Miller and Reese, 2012). Gene expression analyses of ccr4-not mutants indicate that SAGA-dependent genes are mostly affected (Cui et al., 2008), and recently, it was shown that Not5 is important for the functional integrity of SAGA (Kassem et al., 2017).

On the other hand, it has also been found to be involved in transcription elongation. First, the deletion of Ccr4-Not subunits show synthetic lethal interactions with the deletion of other elongation factors: Paf1c, Spt16, TFIIIS and Spt5 (Denis et al., 2001). In addition, some Ccr4-Not mutants are sensitive to the elongation inhibitors 6-azauracil (6-AU) and mycophenolic acid. Second, specific transcription elongation assays show defects in several Ccr4-Not complex mutants (Gaillard et al., 2009; Kruk et al., 2011). Third, the complex was found to cross-link to the open reading frames (ORFs) of genes (Venters et al., 2011). Its recruitment to genes is likely due to its ability to interact directly with the Rpb4/7 module of Pol II (Babbarwal et al., 2014). Finally, Ccr4-Not was found to rescue backtracked polymerases (Dutta et al., 2015; Kruk et al., 2011), as mentioned previously. This might explain the synthetic lethal interaction with TFIIIS, another factor involved in rescuing backtracked Pol II.

Furthermore, the Ccr4-Not complex has been shown to be connected to the nuclear RNA degradation machinery with an impact on mRNA processing (Azzouz et al., 2009). It can also interact with nuclear ribonucleoproteins and the nuclear pore complex, suggesting a role in mRNA export (Kerr et al., 2011). Some of these factors may be the target of the Not4 ubiquitin ligase as ubiquitination has been implicated in the control of mRNA processing and export (Gwizdek et al., 2006).

In the cytoplasm, the Ccr4-Not complex possesses a role in translation. This role is supported by findings such as reduced levels of total polysomes in yeast cells lacking Not2, Not4 or Not5. Also, newly produced proteins have been found to aggregate massively in not2, not4, and not5 mutants (Panasenko and Collart, 2012). This role in translation is particularly important in the case of Not5 as it has been shown to contribute to the synthesis of Rpb1 (Villanyi et al., 2014). This connects translation to transcription.

Xrn1 is another interesting component of the degradation machinery for the study of cross-talk as it has been recently linked to transcription (Haimovich et al., 2013a; Sun et al., 2013). Haimovich et al. attributed a stimulatory role of Xrn1 in transcription. They show that Xrn1 was present upstream of the transcription start sites of the genes most affected by the loss of Xrn1. In contrast, the research by Sun et al. described a negative effect of Xrn1 on global
transcription rates. Despite this controversy, both studies concluded that Xrn1 plays a crucial role in maintaining mRNA homeostasis in response to alterations in either synthesis or mRNA decay. The implication of Xrn1 in cross-talk might be more intense in some gene categories, such as highly transcribed genes, mainly those encoding ribosome biogenesis and translation factors (Medina et al., 2014a).

**Biological significance of the transcription/decay cross-talk.**

Cross-talk between transcription and degradation of mRNAs is vital to coordinate hundreds to thousands of transcripts simultaneously, providing a rapid gene expression response during diverse cellular processes and stresses. Some of these processes are cell growth and cell division, cell cycle, response to stress or other environmental cues (Das et al., 2017; Haimovich et al., 2013b).

During exponential growth, cell volume increases over time until it undergoes cell division. During this time, mRNA levels are at risk of becoming diluted. Consequently, transcription and degradation must be regulated in order to maintain mRNA homeostasis since under or over-expression of factors could be deleterious to living cells. Supporting this idea, a correlation was found between both transcription and mRNA degradation with growth rates (García-Martínez et al., 2016a), and also with cell volume (Mena et al., 2017). It was proposed that a change in growth rate between generations could be damaging to cells if they are not accompanied by an adjustment in transcription and degradation rates (Chávez et al., 2016). For instance, if daughter cells inherit exceedingly stable mRNAs from their mothers, correct phenotypic response to stimuli could be inhibited. This correct adjustment becomes of great importance when it comes to faster proliferating cells.

More than 10% of the protein-coding genes are cell cycle regulated in *S. cerevisiae* (Spellman et al., 1998). These genes alternate between high and low expression periodically. It has been proposed that, at least for some of these genes, cross-talk between transcription and decay aids in the timely expression of genes necessary to enter into the different cell cycle phases (Trcek et al., 2011). Additionally, cross-talk is postulated to shape the expression of distinct gene clusters which undergo cell-cycle-dependent regulation, such as genes involved in DNA synthesis, cytokinesis, budding and other cell cycle-specific events (Wittenberg and Reed, 2005).

Various stresses, such as osmotic stress, heat shock or DNA damage require a rapid response by increasing the expression of some genes and repressing the expression of others. In these
cases, it is essential to reach the newly required steady state mRNA levels and maintain them. This sharp rise or decrease in gene expression can be achieved more efficiently thanks to a coordination between synthesis and decay (Romero-santacreu et al., 2009). Furthermore, this cross-talk provides an efficient way to maintain global mRNA concentrations within certain limits (±50%) in different physiological situations in S. cerevisiae (Benet et al., 2017; Sun et al., 2013). Even during stress, transcription and mRNA decay was shown to correlate with growth rate, thus cross-talk remains even in non-steady state conditions (García-Martínez et al., 2016b).

In a recent paper (Dori-Bachash et al., 2011), the evolution of transcription and degradation rates in two closely related yeast species was assessed. Remarkably, they found that around half of the evolutionary differences in mRNA degradation are coupled to evolutionary differences in transcription. These findings gave rise to the idea that coupling transcription with decay facilitates evolvability as it reduces the number of mutations required to evolve a new trait, such as a new regulatory mechanism (Haimovich et al., 2013b).

Transcription and mRNA degradation also appears to be functionally coupled in mammalian cells. In mice and humans, differences in mRNA degradation were found to compensate transcriptional differences. Additionally, this compensation found to be evolutionary conserved between the two species (Dori-Bachash et al., 2012). Furthermore, several families of mammalian proteins have been associated to cross-talk: BTG/Tob, TIS11, and KSRP. For example, in the first protein family, BTG/tob, the BTG domain is implicated in the interaction with components of transcription as well as mRNA decay such as CNOT7 and CNOT8 which are part of the Ccr4-Not complex. BTG/Tob are involved in the control of cell cycle progression in a variety of cell types and their loss of expression contributes to carcinogenesis and tumour growth (Winkler, 2010). Thus, cross-talk may be important for human diseases.

Another link of cross-talk between transcription and mRNA decay to human diseases is the exploit of the feedback mechanism by a gamma herpes virus. This virus encodes an endonuclease, SOX, that cleaves most cellular mRNAs. SOX generates many new fragments of RNA that require Xrn1 for their degradation. Consequently, Xrn1 is retained in the cytoplasm and cannot perform its function in transcription, leading to transcriptional repression in the nucleus. Indeed the virus uses this feedback mechanism to facilitate viral gene expression, as most of its genes escape this transcriptional repression, and infect its host (Abernathy et al., 2015). Finally, the pathogenic fungus Cryptococcus neoformans was also found to use cross-talk in its infecting process. In this case, the fungus uses the coupling between transcription
and degradation to adapt to the temperature of its host and to promote pathogenicity. Deletion of rpb4 uncoupled the two mechanisms and lead to impaired growth at host temperature and attenuated virulence (Bloom et al., 2013).

In conclusion, this work aims to better understand the cross-talk mechanism that could allow us to gain more insight into how cells adapt their mRNA levels to different environments and necessities, and understand human pathologies, both of a genetic and infective nature.
2. Objectives.
In this work we aimed to deepen our knowledge of the cross-talk between transcription and mRNA decay using the model yeast *Saccharomyces cerevisiae*. To achieve this, we addressed the following aims:

1. **Mutational analysis of the crosstalk between transcription and mRNA decay in a model gene.**

2. **Mathematical modelling and computational simulation of the cross-talk between transcription and mRNA degradation.**

3. **Genome-wide analyses of the transcriptional effects of Xrn1 perturbation.**


mRNA turnover in steady and non-steady states. RNA Biol. 13, 1175–1181.


Gwizdek, C., Iglesias, N., Rodriguez, M.S., Ossareh-Nazari, B., Hobeika, M., Divita, G., Stutz, F.,


Johnson, A.W. (1997). Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. Mol. Cell. Biol. 17, 6122–6130.


Villanyi, Z., Ribaud, V., Kassem, S., Panasenko, O.O., Pahi, Z., Gupta, I., Steinmetz, L.M., Boros,


