3-27-2018

R1 Retrotransposons in *Drosophila melanogaster* are Transcribed by RNA Polymerase I Upon Heat Shock

Himanshu Shekhar Raje

*Louisiana State University and Agricultural and Mechanical College, hraje1@lsu.edu*

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool_dissertations](https://digitalcommons.lsu.edu/gradschool_dissertations)

Part of the Biology Commons, and the Genomics Commons

**Recommended Citation**


[https://digitalcommons.lsu.edu/gradschool_dissertations/4523](https://digitalcommons.lsu.edu/gradschool_dissertations/4523)

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).
R1 RETROTRANSPOSONS IN DROSOPHILA MELANOGASTER ARE TRANSCRIBED BY RNA POLYMERASE I UPON HEAT SHOCK

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College In partial fulfillment of the Requirements for the degree of Doctor of Philosophy

in

The Department of Biological Sciences

by

Himanshu Shekhar Raje
B.S., University of Mumbai, 2003
M.S., University of Mumbai, 2009
May 2018
# TABLE OF CONTENTS

ABREVIATIONS...........................................................................................................iv

PRIMER SEQUENCES USED IN THIS STUDY...............................................................vi

ABSTRACT...................................................................................................................vii

CHAPTER 1: INTRODUCTION
1.1 Transposable Elements.........................................................................................1
1.2 Eukaryotic RNA Polymerases.............................................................................2
1.3 R1 and R2 Retrotransposons Within the Ribosomal DNA of  
   *Drosophila melanogaster*....................................................................................3
1.4 References..........................................................................................................7

CHAPTER 2: EFFECT OF HEAT SHOCK ON THE EXPRESSION OF R1  
RETROTRANSPOSONS
2.1 Introduction.........................................................................................................9
2.2 Materials and Methods.......................................................................................11
2.3 Results and Discussion.....................................................................................12
2.4 References........................................................................................................14

CHAPTER 3: EFFECT OF OXIDATIVE AND NUCLEOLAR STRESS ON  
R1 EXPRESSION
3.1 Introduction.......................................................................................................16
3.2 Materials and Methods.....................................................................................19
3.3 Results and Discussion.....................................................................................19
3.4 References........................................................................................................22

CHAPTER 4: EFFECTS OF BLOCKING POL I AND POL II  
TRANSCRIPTION ON R1 EXPRESSION DURING HEAT  
SHOCK
4.1 Introduction......................................................................................................25
4.2 Materials and Methods.....................................................................................33
4.3 Results and Discussion.....................................................................................35
4.4 References........................................................................................................39

CHAPTER 5: RNA POL I READS THROUGH TO TRANSCRIBE R1  
UPON HEAT SHOCK
5.1 Introduction......................................................................................................42
5.2 Materials and Methods.....................................................................................44
5.3 Results and Discussion.....................................................................................47
5.4 References........................................................................................................54
CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS
6.1 Conclusions ................................................................................................................56
6.2 Future Work ..................................................................................................................57
6.3 References ....................................................................................................................60

VITA ........................................................................................................................................61
## ABBREVIATIONS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complimentary DNA</td>
</tr>
<tr>
<td>rDNA</td>
<td>Ribosomal DNA</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>lnc-RNA</td>
<td>Long non-coding RNA</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>bp</td>
<td>Basepairs</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobases</td>
</tr>
<tr>
<td>CTD</td>
<td>C-terminal domain</td>
</tr>
<tr>
<td>DPE</td>
<td>Downstream promoter element</td>
</tr>
<tr>
<td>IGS</td>
<td>Intergenic sequence</td>
</tr>
<tr>
<td>ETS</td>
<td>External transcribed spacer</td>
</tr>
<tr>
<td>FC</td>
<td>Fibrillar center</td>
</tr>
<tr>
<td>DFC</td>
<td>Dense fibrillar component</td>
</tr>
<tr>
<td>GC (nucleolus)</td>
<td>Granular component</td>
</tr>
<tr>
<td>GAF</td>
<td>GAGA factor</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescence protein</td>
</tr>
<tr>
<td>HSF</td>
<td>Heat shock factor</td>
</tr>
<tr>
<td>HSP</td>
<td>Heat shock protein</td>
</tr>
<tr>
<td>ITS</td>
<td>Internal transcribed spacer</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>Nopp140</td>
<td>Nucleolar phosphoprotein of 140 kDa</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>LTR</td>
<td>Long terminal repeat</td>
</tr>
<tr>
<td>LINE</td>
<td>Long interspersed nuclear element</td>
</tr>
<tr>
<td>SINE</td>
<td>Short interspersed nuclear element</td>
</tr>
<tr>
<td>TIF1A</td>
<td>Transcription initiation factor 1A</td>
</tr>
<tr>
<td>Pol I</td>
<td>RNA Polymerase I</td>
</tr>
<tr>
<td>Pol II</td>
<td>RNA Polymerase II</td>
</tr>
<tr>
<td>Pol III</td>
<td>RNA Polymerase III</td>
</tr>
<tr>
<td>PRO-seq</td>
<td>Precision run-on sequencing</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>JNK</td>
<td>c-Jun N terminal kinase</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleoprotein</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>S2</td>
<td><em>Drosophila</em> Schneider 2 cells</td>
</tr>
<tr>
<td>MMuLV</td>
<td>Moloney murine leukemia virus</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>NTP</td>
<td>Nucleotide triphosphate</td>
</tr>
<tr>
<td>TDF</td>
<td>Tiled data file</td>
</tr>
<tr>
<td>BAM</td>
<td>Binary version of SAM</td>
</tr>
<tr>
<td>IGV</td>
<td>Integrative genomics viewer</td>
</tr>
<tr>
<td>Gene</td>
<td>Primer sequences</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Actin 5C</strong></td>
<td>Forward – 5' CTC ACC TAT AGA AGA CGA AGA AGT TGC TGC TCT 3'&lt;br&gt;Reverse – 5' CTA ACT GTT GAA TCC TCG TAG GAC TTC TCC AAC G 3'&lt;br&gt;Intronic forward – 5' CAG CGC AGT CCA AGG AAA CCA CGC 3'</td>
</tr>
<tr>
<td><strong>ITS2</strong></td>
<td>Forward – 5' TGG AGT ACT ATG GTT GAG GGT TG 3'&lt;br&gt;Reverse – 5' CGA ACC AAC GAA GAA TAA TAA CAT AAC C 3'</td>
</tr>
<tr>
<td><strong>R1</strong></td>
<td>Forward – 5' CGC TAA GGA TTG TGT CTT GGG ACA G 3'&lt;br&gt;RT-Reverse (+685 bp downstream) – 5' CAG CGA TTT TAG CAG CAG TGG AAA C 3'&lt;br&gt;Reverse (+272 bp downstream) – 5' CTG TCC CAA GAC ACA ATC CTT AGC G 3'&lt;br&gt;Reverse (+116 bp downstream) – 5' CGC GAA AAT TTG CGC ACC ACT TCC ACG G 3'</td>
</tr>
<tr>
<td><strong>R2</strong></td>
<td>Forward – 5' ATG ATG TGC GGA AGG GGA ATT TTA C 3'&lt;br&gt;Reverse – 5' TTT GCT GTG AGC TCA ACC TCC TTT C 3'</td>
</tr>
<tr>
<td><strong>Copia</strong></td>
<td>Forward – 5' TAT GGG CCC AGT CCA TGC CTA ATA AAC 3'&lt;br&gt;Reverse – 5' CGA CGC CAA ACT TTT TCG TTC ATA AAC 3'</td>
</tr>
<tr>
<td><strong>Hsp26</strong></td>
<td>Forward – 5' CCC CAT CTA CGA GCT TGG ACT G 3'&lt;br&gt;Reverse – 5' TGT AGC CAT CGG GAA CCT TGT AGC 3'</td>
</tr>
<tr>
<td><strong>Trx2</strong></td>
<td>Forward - 5' CAT TTT CAT TTG CAG GCC GAT CTC GAT GG 3'&lt;br&gt;Reverse – 5' GAA CTC TTC GAC CTT GAC GGC GTT 3'</td>
</tr>
<tr>
<td><strong>28S</strong></td>
<td>Forward – 5' TCT AAT TAG TGA CGC GCA TGA ATG 3'&lt;br&gt;Forward (-69 bp upstream of R1 insertion) – 5' CAA ATG CCT CGT CAT CTA ATT AGT GAC GC 3'&lt;br&gt;Reverse – 5' TGG CTA GGA AAT GAT ACA CGT TCC 3'</td>
</tr>
<tr>
<td><strong>18S</strong></td>
<td>Forward – 5' ACA GAT TGA TAG CTC TTT CTC GAA TC 3'&lt;br&gt;Reverse – 5' CCA TTT AAG AAG CTA GTG TCC TTA TAA TGG G 3'</td>
</tr>
</tbody>
</table>
ABSTRACT

The ribosomal RNA genes of *Drosophila melanogaster* reside within centromere-proximal nucleolar organizers on both the X and Y chromosomes. Each locus contains between 200-300 tandem repeat *rDNA* units that encode 18S, 5.8S, and 28S ribosomal RNAs (*rRNAs*) for ribosome biogenesis. In arthropods like *Drosophila*, about 60% of *rDNA* genes are inserted with *R1* and/or *R2* retrotransposons at specific sites within the 28S regions; these units likely fail to produce functional 28S rRNA. We showed previously that *R2* expression increases upon nucleolar stress caused by the loss of a ribosome assembly factor, the Nucleolar Phosphoprotein of 140 kDa (Nopp140). Here we show that *R1* expression is selectively induced by heat shock. Actinomycin D, but not α-amanitin, blocked *R1* expression in S2 cells upon heat shock, indicating that *R1* is transcribed by Pol I. RT-PCR analysis confirmed read-through transcription by Pol I from the 28S gene region into *R1*. Using a genome wide precision run-on sequencing (PRO-seq) data set available at NCBI-GEO, we showed that Pol I activity on *R1* elements is negligible under the normal non-heat shock condition but increases dramatically upon heat shock. We propose that prior to heat shock, Pol I pauses within ~350 bp of the 5' end of *R1* wherein we find ‘pause button’ like sequence motifs, and that heat shock releases Pol I for read-through transcription into *R1*. 
CHAPTER 1: INTRODUCTION

1.1 Transposable Elements

Transposable elements are interspersed repeats that have a unique ability to move in the genome. These mobile genetic elements were discovered by Barbara McClintock during the 1940s. There are two types of transposable elements: the DNA transposons and retrotransposons. The DNA transposons are enzymatically cleaved from their location in the genome and insert their DNA elsewhere in the genome. The DNA transposons are primarily found in prokaryotes and do not require their RNA intermediate for their insertion into the new genomic location. Whereas, retrotransposons are first transcribed to their RNA intermediate, which is then reverse transcribed and their cDNA acts as a substrate for insertion of these retrotransposons into their genomic target sites. Although there are some DNA transposons in all eukaryotes, most of the transposable elements in eukaryotes are retrotransposons. There are two types of retrotransposons: the LTR type and the non-LTR type. The LTR type retrotransposons have a central protein coding region flanked by element specific LTR (long terminal repeat) sequences. The upstream LTR acts as a promoter sequence that directs the host cell RNA polymerase for the transcription of LTR retrotransposons (Lodish 2013). The non-LTR type retrotransposons lack the LTR repeats. Non-LTR retrotransposons appear in two forms: LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) (Figure 1.1). LINEs are about 6 kb long, whereas SINEs are just ~300 bp long. Some LINEs possess a single open reading frame (ORF) and the other LINEs have two ORFs (Lodish 2013). Most of the LINEs have their own internal promoter in their 5' UTR.
(untranslated region), and they are transcribed by RNA Polymerase II (Pol II) (Heras et al. 2006, Mizrokhi et al. 1988, Ostertag and Kazazian 2001).

Figure 1.1 Types of transposable elements. DNA transposons are enzymatically cut from their original genomic location and insert into other parts of the genome. DNA transposons do not need their RNA intermediate for insertion into genome. Whereas, LTR and non-LTR retrotransposons require RNA intermediate for their insertion into genome. Their RNA intermediate is reverse transcribed to cDNA during insertion into their target sites. LINEs and SINEs are two types of non-LTR retrotransposons. LINEs encode their own proteins such as reverse transcriptase and endonuclease, whereas SINEs do not encode their own proteins and depend on cellular enzymes.

1.2 Eukaryotic RNA Polymerases

Most of the eukaryotes have three known RNA polymerases. RNA Polymerase I (Pol I) synthesizes ribosomal RNA (rRNA) which constitutes about 80% of total cellular RNA. The activity of Pol I is nucleolar. RNA Polymerase II (Pol II) transcribes mRNAs (messenger RNAs) from all protein coding genes. Pol II also transcribes some regulatory RNAs and non-coding RNAs. The activity of Pol II is exclusively nuclear. RNA Polymerase
III (Pol III) transcribes only short length RNAs such as tRNAs, 7SL RNA and 5S rRNA. All eukaryotic RNA polymerases share subunits (Lodish 2013). However, their largest subunits are unique to each polymerase. Pol II is structurally distinct from the other two polymerases because its largest subunit has a long CTD (C terminal domain) that is phosphorylated when Pol II is transcriptionally active (Chapter 4). Almost all of the transposable elements are transcribed by Pol II. SINE elements are transcribed by Pol III. \textit{R1} and \textit{R2} retrotransposons reside in the ribosomal DNA of all arthropods including \textit{Drosophila melanogaster}. \textit{R2} is transcribed by Pol I but it is not yet clear which polymerase transcribes \textit{R1} (Ye and Eickbush 2006).

\subsection*{1.3 \textit{R1} and \textit{R2} Retrotransposons Within the Ribosomal DNA of \textit{Drosophila melanogaster}}

In \textit{Drosophila melanogaster}, the ribosomal DNA (\textit{rDNA}) is present as tandem repeats in the centromere proximal regions of the X and Y chromosomes. There are about 200-300 \textit{rDNA} unit repeats at each locus. Each unit is separated by an intergenic sequence (IGS) that contains the \textit{rDNA} core promoter (Figure 1.2 A). The coding region of each \textit{rDNA} unit consists of an external transcribed spacer (ETS) region, the 18S region, internal transcribed spacer 1 (ITS1), the 5.8S region, ITS2, and the 28S region. The 18S ribosomal RNA (rRNA) assembles into the small ribosomal subunit, whereas the 5.8S and 28S rRNA contribute to the formation of large ribosomal subunit (Nikolaev et al. 1979). The ITS1 and ITS2 regions are transcribed as a part of the full length pre-ribosomal RNA (pre-rRNA), however they are quickly processed out of the pre-rRNA, and degraded. Therefore, ITS1 and ITS2 can be used as markers for ribosomal DNA transcription (Allmang et al. 2000).
In all insect species including *Drosophila*, the 28S region of many *rDNA* units is interrupted by type I and/or type II insertion sequences (Long and Dawid 1979) which are now referred to as *R1* and *R2* retrotransposons, respectively (Perez-Gonzalez and Eickbush 2002) (Figure 1.2C and inset). *R1* and *R2* elements insert into the 28S regions in a sequence dependent manner (Figure 1.3). While ~40% of *rDNA* units do not have either of these insertions, ~44% of the units have just *R1* insertions, and ~11% of the units have just *R2* insertions in the 28S region. Only a small fraction of the *rDNA* units (~5%) have both *R1* and *R2* insertions (Ye and Eickbush 2006). There is a 74 bp sequence of 28S DNA separating *R1* and *R2* insertion sites, with *R2* located upstream of *R1*.

In *Drosophila* embryonic cell chromatin, the *rDNA* units with or without *R1* and/or *R2* insertions are similar with respect to their nuclease sensitivity, psoralen crosslinking, and histone H3 and H4 modifications. Furthermore, once activated for transcription, the rates of transcription initiation are similar for inserted and uninserted *rDNA* units (Ye and Eickbush 2006).
Figure 1.2 Organizational structure of Drosophila rDNA. (A) Intergenic sequence (IGS) region that contains the rDNA promoter. Regions I, II and III represent enhancer elements, upstream promoter element and the core promoter element respectively (B) Multiple rDNA units present in tandem array on the X and Y chromosomes of Drosophila melanogaster. Some of these units are inserted by R1 and/or R2 retrotransposons in their 28S region. (C) A single rDNA unit inserted by both R1 and R2 (inset). The insertion sites of R1 and R2 are about 7 kb away from the rDNA transcription start site (+1).

Figure 1.3 R1 and R2 insertion sites (nucleotides indicated in bold) within the 28S region of the Drosophila melanogaster rDNA. Sequences are taken from Genebank, accession no. M21017.1.

Both R1 and R2 belong to the LINE class of the non-LTR retrotransposons that do not have long terminal repeat (LTR) sequences at their ends (Eickbush 2002). Full length R1 element is about 5.3 kb. R1 has two open reading frames (ORFs): ORF1 codes for a
protein with unknown function, and ORF2 encodes a protein with an apurinic/apyrimidinic endonuclease domain and a reverse transcriptase domain (Eickbush 2002). Nuclear run-on experiments showed significantly lower abundance of $R1$ transcripts beyond ~1 kb from the $R1$ insertion site leading to the hypothesis that RNA polymerase I (Pol I) complexes dissociate from the template within the 5’ end of $R1$ (Ye and Eickbush 2006). However, how $R1$ is transcribed has remained uncertain.

Full length $R2$ element is about 3.2 kb in length. Unlike $R1$, $R2$ has one ORF that encodes a protein with a DNA binding domain, an endonuclease domain, and a reverse transcriptase domain (Eickbush 2002). $R2$ is co-transcribed with $rDNA$ by Pol I (Ye and Eickbush 2006). The 5’ end of the $R2$ transcript encodes a self-cleaving hammerhead type ribozyme that splices itself out of the pre-rRNA (Eickbush and Eickbush 2010). Nuclear run-on experiments detected almost equal abundance of 5’ and 3’ $R2$ transcripts (Ye and Eickbush 2006). This observation rules out the possibility that Pol I transcription complexes dissociate from the template within $R2$ during transcription.

Recent reports show that $R2$ transcription rates increase upon nucleolar stress caused by the loss of Nopp140, a nucleolar ribosome assembly factor (He et al. 2015), and that both $R1$ and $R2$ along with several other transposable elements increase expression upon Lamin knockout in Drosophila (Chen et al. 2016). We know that environmental stress conditions influence the transcription of several retrotransposons (Cavrak et al. 2014, Ikeda et al. 2001, Lavie et al. 2004). For example, the $L1$ retrotransposon in mammals and ONSEN in plants are known to be upregulated upon heat shock (Cavrak et al. 2014, Lavie et al. 2004). $L1$ is also a non-LTR type retrotransposon like $R1$ and $R2$. However, unlike $R2$ which has one ORF, $L1$ has two
ORFs. \textit{L1} is therefore structurally similar to \textit{R1}; however, \textit{L1} has its own internal Pol II promoter that responds to heat shock and oxidative stress (Giorgi et al. 2011, Lavie et al. 2004, Speek 2001). It is not known how \textit{R1} and \textit{R2} respond to environmental stress. We observed changes in \textit{R1} transcript levels in response to heat shock (Chapter 2) and pursued further investigation of how \textit{R1} and \textit{R2} elements respond to cellular stress (Chapter 3).

1.4 References


Mizrokhi LJ, Georgieva SG, Ilyin YV. 1988. Jockey, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell 54:685-691.


CHAPTER 2: EFFECT OF HEAT SHOCK ON THE EXPRESSION OF R1 RETROTRANSPOSONS

2.1 Introduction

2.1.1 Heat shock response

Most higher organisms adapt to their own optimum temperature. Any increase above the highest temperature point in the optimum temperature range is considered a heat shock. Heat shock is the predominant environmental stress that all organisms are regularly exposed to (Richter et al. 2010). The optimum temperature range for *Drosophila melanogaster* is about 18°C-28°C, so a temperature above 28°C is considered a heat shock, and a temperature below 18°C is considered a cold shock. Upon heat shock, organisms exhibit intracellular changes that include aggregation of misfolded proteins, cytoskeletal damage, fragmentation of the Golgi complex and the endoplasmic reticulum, loss of mitochondria, and a subsequent decrease in the amount of ATP (Welch and Suhan 1985). Messenger RNA (mRNA) splicing is impaired after heat shock (Vogel et al. 1995) indicating that the damage is not limited to the cytoplasm but extends to the nucleus as well.

Cells respond to heat shock primarily by increasing transcription and translation of several gene classes. The most important and well studied genes are those encoding Heat Shock Proteins (HSPs) or ‘molecular chaperons’ that help with proper refolding of misfolded proteins (Ashburner and Bonner 1979). Apart from HSPs, several other categories of proteins such as DNA/RNA modifying enzymes and DNA repair proteins are also upregulated as a part of heat shock response (Richter et al. 2010).

Another category of Pol II transcripts that respond to heat shock are transposable elements. Transposons constitute a significant part of the mammalian genome, yet their
functions and impacts are not well understood. They are mobile genetic elements that cause genetic alterations and recombination wherever they insert, which leads to overall genome instability. There are two known types of transposons: 1) DNA transposons that are enzymatically cleaved from their location and insert elsewhere in the genome, and 2) Retrotransposons like \( R1 \) and \( R2 \) that are first transcribed into RNA molecules that are reverse transcribed into double stranded DNA that then insert into the genome.

All HSP genes, genes coding for mRNAs, and genes encoding some regulatory non-coding RNAs depend on RNA polymerase II (Pol II) for their transcription. This heat shock induced expression of HSPs by Pol II is largely mediated by Heat shock factor (HSF) and GAGA factor (GAF), both of which are well studied Pol II transcription factors (Wilkins and Lis 1997). However, the nuclear effect of heat shock is not limited to Pol II transcription. In mouse lymphosarcoma cells, Pol I mediated transcription is selectively downregulated by heat shock (Ghoshal and Jacob 1996). In addition, heat shock downregulates \( rDNA \) transcription in mammalian cells due to epigenetic changes in the \( rDNA \) promoter, as well as to the downregulation of the basal \( rDNA \) transcription factor, TIF-1A (Transcription initiation factor 1A discussed in Chapter 3) (Zhao et al. 2016). The same report claimed a simultaneous rise in transcription of a long-noncoding RNA (Inc-RNA) that is transcribed in the antisense orientation and is associated with transcriptional repression of human \( rDNA \) (Zhao et al. 2016).

2.1.2 The Nucleolus

The nucleolus is a nuclear subdomain that contains the \( rDNA \) and the known site for ribosome assembly. Recently emerging evidence suggests that the nucleolus plays a vital role as a cellular stress sensor (Mayer et al. 2005). \( rDNA \) transcription and
processing occur within the nucleolus, and both processes determine nucleolar integrity. However, the effect of heat shock on the integrity of nucleolus is not well understood. I wanted to test the effect of heat shock on the expression of $R1$ and $R2$ retrotransposons.

2.2 Materials and Methods

The Wild type (WT) stock was $w^{1118}$ Drosophila melanogaster (obtained from the Bloomington Drosophila stock center, stock #3605) grown at room temperature (22-24°C) on standard medium. For heat shock treatment, ~20 well fed third instar larvae were placed into an empty vial which was submerged in a water bath set at 37°C for 1 hr. For the non-heat shocked control, total RNA was isolated from ~20 well fed non-heat shocked third instar WT larvae kept in a similar vial at room temperature for 1 hr. Total RNA from heat shock and control larvae was isolated using TRIzol (Invitrogen) using the manufacturer’s instructions. RNA samples were treated with DNase I (Promega) for 1 hr at 37°C and then ethanol precipitated. RNA concentrations were measured using a spectrophotometer. Equal masses of total RNA were subjected to first strand cDNA synthesis using MMuLV reverse transcriptase (New England BioLabs). Neither $R1$ nor $R2$ are known to have poly(A) tails, therefore specific primers (Table 1) were used to synthesize their first strand cDNAs. Equal volumes of cDNA were then used to perform semi-quantitative RT-PCR with 25 cycles of amplification, using gene specific primers for all of the genes tested (Table 1).

PCR products were resolved on 1% agarose TAE gels and imaged using a Bio-Rad gel imager. Equal volumes of cDNAs were also used for quantitative RT-PCR (qRT-PCR) using a New England BioLabs Luna universal qPCR master mix and a Quantstudio 6 qPCR instrument from Applied Biosystems.
2.3 Results and Discussion

Transcription of all HSP genes including HSP26 is induced by heat shock (Ashburner et al. 1979). Therefore, to check the influence of heat shock on R1 and R2 expression, we used HSP26 expression as a positive control. (Ashburner and Bonner 1979) We also measured the copia retrotransposon RNA in this experiment. Copia does not insert within ribosomal DNA and there are mixed reports in the literature describing the influence of heat shock on its transcription (Gilmour and Lis 1985, Strand and McDonald 1985). Since Actin mRNA is abundant and stable, Actin 5C was used as a general control for RT-PCR. Upon 1 hr of heat shock, the transcription of R1 was significantly induced, whereas Actin 5C, R2, and copia transcript levels remained unchanged according to the semi-quantitative RT-PCR (Figure 2.1A). HSP26 transcript was upregulated as expected (Figure 2.1A). In order to determine if R1 expression resulted from an overall increase in rDNA transcription, we measured the expression of 18S, ITS2 and 28S regions by using region specific primer pairs. Approximate positions of primers in each region are indicated in Figure 2.1B by arrows, and their sequences are shown in Table 1. There was no observable change in rDNA transcription upon heat shock (Figure 2.1B). In particular, the ITS2 expression; which is indicative of the rDNA transcription, was unchanged upon heat shock.
Figure 2.1 Effect of heat shock on $R1$, $R2$, and $rDNA$ transcription. (A) Semi-quantitative RT-PCR gel pictures compare expressions between non-heat shocked and heat shocked third instar larvae. (B) Effect of heat shock on $rDNA$ transcription. Approximate positions of primer pairs for the indicated regions of $rDNA$ are shown by arrows. Primer sequences are listed in Table 1. All of the PCR reactions were carried out for 25 cycles with same denaturation and extension times.

To quantitate the expression of $R1$ upon heat shock, I used qRT-PCR to measure the Ct values. The raw Ct values for each gene were normalized to those for the $Actin\ 5C$ control, and the relative expression of each gene under non-heat shock conditions was normalized to one. qRT-PCR data analysis was performed as described below:

1. Each Ct value for every gene for a particular experimental treatment was subtracted from the mean loading control Ct value ($\Delta$Ct).
2. Consider one of these subtractions as ‘X’. To proceed further, $2^X$ value was calculated for each subtraction.
3. Mean $2^X$ value was calculated and each $2^X$ value was divided by mean $2^X$ value ($\Delta\Delta$Ct).
4. The average of the division described in step 3 was plotted as relative fold expression.

Relative expression of R1 was induced by about 12 fold upon heat shock, whereas copia and R2 were downregulated. HSP26 was induced by about 2 fold (Figure 2.2). These results show that R1 responds to heat shock. Although R1 and R2 belong to the same non-LTR retrotransposon family, they seem to behave differently in their regulated expression under heat shock conditions. However, it is not clear which polymerase transcribes R1 upon heat shock.

Figure 2.2 Bar graph depicting qRT-PCR analysis to show the effect of heat shock on R1 and R2 expression. R1 expression was induced about 12 fold whereas, copia and R2 were downregulated upon heat shock. Ct values of all genes were normalized to those of Actin 5C.

2.4 References


CHAPTER 3: EFFECT OF OXIDATIVE AND NUCLEOLAR STRESS ON R1 EXPRESSION

3.1 Introduction

3.1.1 Oxidative stress

Oxidative stress is defined as a disturbance in the balance between the formation of intracellular reactive oxygen species and cells’ antioxidant defense. It is essentially a result of the production of free radicals that are not neutralized by antioxidants (Betteridge 2000). If left unchecked, oxidative stress could lead to tissue damage. Lipid peroxidation and oxidation of DNA and proteins are some of the predominant effects of oxidative stress (Barrera 2012, Barzilai and Yamamoto 2004). At the cellular level, prolonged oxidative stress can cause cell death by both apoptosis and necrosis (Ryter et al. 2007). Oxidative stress has also been established as a consequence of underlying diseases such as diabetes, cancer, Alzheimers, Parkinsons and cardiovascular diseases (Betteridge 2000). One study looked at gene expression patterns in *Drosophila melanogaster* by microarray analysis, upon different stress conditions, and observed that gene expression patterns overlap between oxidative stress and aging. On the other hand, minimal gene expression similarities were seen between oxidative stress/aging and heat shock (Landis et al. 2012). This would suggest that the transcriptional response to heat shock is different from transcriptional changes observed upon oxidative stress/aging.

Exposure to certain chemicals can induce oxidative stress. The most common chemical that can cause oxidative stress is hydrogen peroxide (H$_2$O$_2$) (Wijeratne et al. 2005). H$_2$O$_2$ is a product of several cellular biochemical pathways that involve oxidase enzymes. H$_2$O$_2$ is lipid soluble and hence can diffuse across cell membranes if exposed to cells (Horwitz et al. 1996). Upon cellular exposure to H$_2$O$_2$, there is a predictable
increase in reactive oxygen species. Prolonged exposure to \( \text{H}_2\text{O}_2 \) causes DNA damage (Betteridge 2000). Several genes are transcribed upon oxidative stress. They include but are not limited to antioxidant enzymes like \textit{Thioredoxin-2} (Svensson and Larsson 2007), cytokines and immune response genes (Elmarakby and Sullivan 2012), tumor suppressors (Jenkins et al. 2011) and some transposable elements. In fact, one such DNA-transposon \textit{Bari-Jheh} is associated with expression of antioxidant defense genes \textit{Jheh1} and \textit{Jheh2} in \textit{Drosophila melanogaster} (Guio et al. 2014). \textit{Tf1}, which is an LTR type retrotransposon in fission yeast that integrates into the promoters of oxidative stress response genes, is reported to act as an enhancer sequence to assist the expression of \textit{Tf1} inserted stress response genes (Feng et al. 2013). The \textit{L1} retrotransposon in mammals, \textit{ONSEN} in \textit{Arabidopsis}, and \textit{MAGGY} in fungi are also upregulated by oxidative stress (Cavrak et al. 2014, Ikeda et al. 2001, Lavie et al. 2004).

The nucleolus also responds to oxidative stress. In yeast, a homologue of TIF1A (Transcription initiation factor 1A, discussed in Chapter 4) called Rrn3 translocates from the nucleolus to the cytoplasm upon oxidative stress. This redistribution of Rrn3 results in the downregulation of \textit{rDNA} transcription in yeast (Lewinska et al. 2010). This chapter describes the effects of oxidative stress on \textit{R1} and \textit{R2} expression.

3.1.2 Nucleolar stress

The nucleolus is a subnuclear membrane-less region within the nucleus that contains tandem arrays of ribosomal DNA genes. Critical cellular processes such as \textit{rDNA} transcription, \textit{rRNA} processing, ribosome assembly, and the maturation of ribosomal subunits occur within a nucleolus (Pederson 1998). The nucleolus is divided into three ultrastructural components: the fibrillar center (FC), the dense fibrillar
component (DFC) and the granular component (GC). Furthermore, the nucleolus is now regarded as 'plurifunctional' because it takes part in several other cellular processes such as tRNA maturation, U6 RNP assembly, telomerase RNP assembly, genome maintenance and telomere replication (Pederson 1998). The nucleolus is also widely considered as a cellular stress sensor. One way in which the nucleolus responds to stress conditions is by downregulating rDNA transcription (Mayer et al. 2005). This (stress induced) downregulation of the rDNA transcription is mediated by JNK (c-Jun N terminal kinase) which adds an inhibitory phosphate group to TIF1A. The subsequent translocation of TIF1A from the nucleolus to nucleoplasm decreases Pol I activity within the nucleolus (Mayer et al. 2005).

Failure in ribosome biogenesis or function that ultimately leads to disruption in cellular homeostasis is now known as nucleolar stress (Boulon et al. 2010). Nucleolar stress could be induced by perturbing nucleolar homeostasis in one of several ways such as inhibiting Pol I transcription initiation or elongation, inhibiting pre-rRNA processing, disturbing ribosome biogenesis or inhibiting export of ribosomal subunits from the GC (Boulon et al. 2010, Wang and DiMario 2017). Our lab has established that a gene knockout of an essential nucleolar phosphoprotein (Nopp140) induces nucleolar stress in Drosophila (He et al. 2015). Drosophila has two isoforms of Nopp140 named as Nopp140-True and Nopp140-RGG (Waggener and DiMario 2002). In Drosophila both of these isoforms localize to nucleoli in addition to Cajal bodies. Currently we know that Nopp140-True is essential for chemical modifications of pre-rRNA (He et al. 2015). Furthermore, transcription of R2 retrotransposon is induced upon Nopp140 gene knockout (He et al. 2015). While Nopp140 is shown to interact with Pol I in mammalian cells (Chen et al.
1999), its role in rDNA transcription is still under investigation. Since \( R1 \) and \( R2 \) insert within the majority of rDNA units, I wanted to test the effect of nucleolar stress on \( R1 \) expression.

**3.2 Materials and Methods**

3.2.1 Induction of oxidative stress in *Drosophila* third instar larvae

To induce oxidative stress, well fed third instar larvae were cultured for 1 day on 0.5% agar plates with or without 3% \( \text{H}_2\text{O}_2 \). All plates were supplemented with granular yeast lightly sprinkled on their surface to avoid starvation. Total RNA was isolated using TRizol. Equal concentrations of DNase treated RNA samples were subjected to first strand cDNA synthesis using MMuLV reverse transcriptase (New England BioLabs) and subsequently to PCR/qPCR using gene specific primers (Table 1).

3.2.2 Nucleolar stress treatment

For nucleolar stress treatment (He et al. 2015), total RNA was isolated from WT larvae, homozygous WH\(^+\) larvae (an internal control with a piggyBac element inserted into the \( P5CDh1 \) gene which resides immediately downstream of the \( Nopp140 \) gene), and homozygous \( Nopp140 \) gene deletion (KO121) larvae (He et al. 2015). Equal concentrations of RNA were subjected to first strand cDNA synthesis using MMuLV reverse transcriptase (New England BioLabs) and subsequently to PCR/qPCR using specific primers (Table 1).

**3.3 Results and Discussion**

Upon oxidative stress, expression of \( R1 \), \( R2 \) and \textit{copia} were not induced (Figure 3.1A). In fact, qRT-PCR data suggests that the expression of \( R1 \) and \textit{copia} were downregulated upon oxidative stress (Figure 3.1 B). \( R2 \) expression was slightly elevated
by qPCR, but the significance was minimal (Figure 3.1B). Thioredoxin-2 (Trx-2) has been used as a marker gene for oxidative stress (Svensson and Larsson 2007). Its expression was induced by about 2.5 fold following oxidative stress as seen by qRT-PCR (Figure 3.1B). Actin 5C was used as a general positive control for RT-PCR as its transcripts are abundant and stable.

Upon nucleolar stress caused by Nopp140 gene knockout, expression of R2 was significantly induced by about 5 fold as expected, but expression of R1 and copia remained unchanged as measured by semi-quantitative RT-PCR (Figure 3.2, 3.3). Copia was in fact downregulated in homozygous Nopp140 knockout larvae as seen by qPCR (Figure 3.3).

![Figure 3.1 Effect of oxidative stress on the expression of R1, R2, and copia](image)

A) Semi-quantitative RT-PCRs were performed with 25 cycles for each gene and PCR products were resolved on agarose gels. Gel pictures depict expression of each gene under control (without oxidative stress) and oxidative stress conditions. B) Bar graph showing qRT-PCR data analysis. Ct values for all the genes were normalized to Actin 5C expression. Trx2 normally responds to oxidative stress, and was used as a positive control.
Figure 3.2 Effect of nucleolar stress on the expression of \(R1\), \(R2\), and \(copia\) by semi-quantitative RT-PCR performed with 25 cycles for each gene. WH is an internal control with a piggyBac element inserted downstream of \(Nopp140\) gene.

Figure 3.3 Effect of nucleolar stress on \(R1\), \(R2\), and \(copia\) expression measured by qRT-PCR. Ct values for all of the genes are normalized to \(Actin\ 5C\) expression. WT is wild type and WH is an internal control with a piggyBac element inserted downstream of \(Nopp140\) gene.

The data shown in Chapters 2 and 3 together suggest that \(R1\) and \(R2\) behave differently in their regulated expression. \(R1\) does not behave like a general stress induced retrotransposon. Instead, its expression responds specifically to heat shock. On the other hand, \(R2\) expression is specifically induced by nucleolar stress. As mentioned in Chapter
1, R2 is transcribed by RNA Pol I, but nuclear run-on experiments showed very low transcription of R1 beyond 1 kb of its 5’ end leading to the hypothesis that Pol I leaves the template within 5’ end of R1 (Ye and Eickbush 2006). Upon heat shock, however, R1 transcription is induced, but it is still unclear which polymerase transcribes R1 under heat shock conditions. All heat shock induced genes identified so far are transcribed by RNA Pol II. Since, there is no report suggesting any activity of RNA Pol II within nucleolus, we wanted to selectively inhibit RNA Pol I and Pol II transcription separately to test if either of these treatments have an effect on R1 expression (Chapter 4).

3.4 References


CHAPTER 4: EFFECTS OF BLOCKING POL I AND POL II TRANSCRIPTION ON R1 EXPRESSION DURING HEAT SHOCK

4.1 Introduction

RNA Pol I transcribes the ribosomal DNA, and therefore its activity is exclusively nucleolar. Pol I activity essentially dominates cellular transcription, as ribosomal RNA (rRNA) constitutes about 90% of the total RNA within the cell. Pol I synthesizes rRNA on the border between the fibrillar center (FC) and the dense fibrillar component (DFC) of the nucleolus. The rRNA is further processed as it assembles with early ribosomal proteins in the DFC, and then nearly completes assembly into the ribosomal subunits in the granular component before export to the cytoplasm (Paule and White 2000).

4.1.1 Ribosomal DNA promoter

Pol I initiates rDNA transcription from the promoter located in the highly repetitive IGS region. Although the sequence of various rDNA promoters show minimal conservation between species, the overall structure of the promoters is highly conserved. The rDNA core promoter lies about 50 bp upstream of the primary transcription start site (+1) which marks the beginning of the ETS (Figure 4.1). The core promoter is sufficient for basal transcription, and contains an AT-rich sequence element called initiator (Inr) located upstream of the transcription start site (Paule and White 2000, Perna et al. 1992). An upstream promoter element (UPE) is located about 150-200 bp upstream of the transcription start site. UPE is the binding site for UBF in vertebrates. UPE enhances transcription initiation from the initiator sequence (Reeder 1984). IGS region has multiple repeat sequence elements that contain minor transcription initiation sites from which wandering Pol I molecules could possibly begin transcription, essentially directing Pol I to the core promoter. The proximal terminator (PT) is a sequence element upstream of the
UPE that primarily helps in terminating Pol I transcription that has initiated from any of the minor initiation sites other than +1 (Henderson et al. 1989). PT has other architectural functions that are reported only in certain species, such as helping chromatin remodeling on the rDNA promoter (Langst et al. 1998). The rest of the IGS upstream of the PT contains several enhancer elements that are present as repeats (Figure 4.1). The mechanism by which these elements enhance rDNA transcription is not yet known.

![Figure 4.1](image-url)

Figure 4.1 The organizational structure of the IGS region that contains the rDNA core promoter.

4.1.2 RNA polymerase I transcription

Pol I is a large enzyme complex made up of 14 subunits. It shares five of its subunits with Pol II and Pol III. Two other subunits are shared just with Pol III, and the rest are unique to Pol I. This high degree of homology between cellular RNA polymerases suggests that they have a similar general structure and mechanism. The largest and second largest subunits are unique to Pol I, and they make up the catalytic portion of the Pol I holo-enzyme (Goodfellow and Zomerdiijk 2012). In Drosophila, the largest Pol I subunit is RPI 185, and the second largest subunit is RPI 135. Unlike Pol II, Pol I requires fewer factors for its pre-initiation complex formation. Apart from the core subunits, a few transcription factors are known to take part in Pol I pre-initiation complex (PIC) formation. In vertebrates, UBF (Upstream Binding Factor, UAF in Yeast) binds the UPE as a dimer
and creates a docking site for selectivity factor 1 (SL1) (Bazett-Jones et al. 1994). SL1 is a complex of TBP (TATA Binding Protein) and TAF1 (TBP Associated Factor) that binds the core promoter. Upon receiving appropriate transcription activation signals such as a growth stimulus, SL1 recruits Pol I to the rDNA promoter (Comai et al. 1992). TIF1A (Rrn3 in Yeast) is phosphorylated upon growth signals, and it binds Pol I. The TIF1A – Pol I complex then binds SL1 resulting in the formation of pre-initiation complex on the rDNA promoter (Goodfellow and Zomerdijk 2012) (Figure 4.2). Drosophila has TIF1A, and the SL1 homologues have been recently identified and named as TAF1B and TAF1C like factors (Zhang et al. 2014). However, there is no known UBF in Drosophila, thus the mechanism of Pol I recruitment and pre-initiation complex formation remains uncertain in Drosophila.

Once initiated from the rDNA promoter, Pol I is not known to pause or stop transcribing rDNA. However, a report describes abortive ETS transcripts upon Actinomycin D treatment (Shcherbik et al. 2010). Also, in Drosophila, Pol I is thought to terminate within the 5' end of the R1 retrotransposon (Ye and Eickbush 2006). In general, after promoter escape, Pol I elongation rates are found to be extremely high, on the order of 95 nucleotides per second with about 100 Pol I complexes transcribing a single rDNA unit (Dundr et al. 2002). Pol I terminates transcription at a terminator sequence located within 1 kb downstream of the 28S region. This region is a site for termination factor binding, and the orientation of their binding seems to be critical for proper termination (Goodfellow and Zomerdijk 2012).

rDNA transcription is tightly coupled with the growth status of the cell. Actively dividing cells including cells that have lost control over their division cycle (cancer cells)
have high demands for Pol I activity (Russell and Zomerdijk 2005). A few small molecules such as Actinomycin D and CX5461 are widely used chemicals to inhibit Pol I transcription, and are being tested as chemotherapy drugs against cancer cells. Actinomycin D has a DNA binding ability that can selectively block Pol I progression at relatively low concentrations. Actinomycin D is known to inhibit Pol I transcription in several different cell lines, but its action is dose-dependent (Bensaude 2011). CX5461 is a competitive inhibitor of SL1 (Bensaude 2011, Whitten et al. 2008). CX5461 has been exclusively used on cancer cell lines, and it is currently in clinical trials as a potential chemotherapy drug to treat cancer patients.
1. UBF binds rDNA promoter

2. Upon growth factor stimulation TIF1A is phosphorylated allowing it to bind Pol I. UBF dimer recruits SL1.

3. Binding of SL1 to UBF recruits TIF1A-Pol I complex to complete formation of the pre-initiation complex.

Figure 4.2 Steps involved with the formation of Pol I pre-initiation complex in vertebrate cells.

4.1.3 Inhibition of RNA polymerase I transcription by Actinomycin D

Actinomycin D is isolated from bacteria that belong to the *Streptomyces* genus. It has been widely used as an anti-cancer chemotherapy drug. It is a DNA intercalator that binds GC-rich regions of DNA to hinder transcription elongation (Figure 4.3). All three eukaryotic RNA polymerases are sensitive to Actinomycin D. However, the sensitivity is dose-dependent (Bensaude 2011). RNA polymerase I transcription is the most sensitive
to Actinomycin D treatment at about 0.05 μg/mL. RNA polymerase II transcription is affected at about 0.5 μg/mL. A higher dose of 5 μg/mL inhibits RNA Pol III transcription (Bensaude 2011). Since its low dose selectively affects rDNA transcription, it has been extensively used in studies involving Pol I transcription.

![Figure 4.3](image)

**Figure 4.3** Mechanism of the Actinomycin D mediated inhibition of the RNA polymerase I transcription. Actinomycin D binds to the GC-rich regions in the DNA to inhibit the elongation of polymerase molecules. The Pol I transcription is the most sensitive to Actinomycin D mediated transcription inhibition.

4.1.4 RNA Polymerase II transcription

RNA Polymerase II transcribes all messenger RNAs, several small nuclear RNAs, and many micro or non-coding RNAs in eukaryotic cells. Pol II is a large enzyme complex typically made up of 12 subunits that constitute the Pol II holoenzyme. Pol II promoters are bound by several general transcription factors and specific regulatory proteins making the Pol II transcription machinery the most complex within the cell with about 60 polypeptides bound together (Nikolov and Burley 1997). The largest subunit of Pol II is
called RPB1. The C terminal domain (CTD) of RPB1 is extremely important for Pol II transcription; it contains about 52 heptapeptide repeats Tyr-Ser-Pro-Thr-Ser-Pro-Ser (YSPTSPS) in vertebrates. Other organisms also have multiple repeats within the RPB1 subunit with similar amino acid sequences. For example, Pol II in wild type *Saccharomyces cervisiae* has 26 repeats of this sequence. Any number of repeats less than ten is found to be lethal for *Saccharomyces cervisiae* (Lodish 2013). These repeats are unique to RNA Pol II, and there is no evidence of similar repeats in either Pol I or III. The transcription initiation of Pol II depends heavily on the phosphorylation status of serine and tyrosine residues within these repeats.

The formation of the Pol II pre-initiation complex and the anatomy of Pol II promoters is well studied. Commonly found elements in Pol II promoters include a TATA box located about 25 bp upstream of the transcription start site, a pyrimidine-rich initiator sequence located at the transcription start site, CpG islands for TATA-less promoters, a downstream promoter element (DPE), and cis/trans-acting enhancer sequences (Lodish 2013). The general Pol II transcription factors include TFIID, B, F, E and H. TFIID contains the TBP and a TAF complex that loads Pol II onto the promoter. TFIID is highly conserved among species (Lodish 2013). Although the TBP subunit of TFIID binds a TATA box, TBP binding is also detected at TATA-less promoters (Pugh and Tjian 1991). TFIIB binding occurs only after TBP is bound to the TATA box. A preformed complex of TFIIF and Pol II joins the assembly of proteins at the promoter, and TFIIF correctly positions Pol II. TFIIE and TFIIH binding completes the pre-initiation complex formation. TFIIB and the helicase subunit of TFIIH assist in the melting of DNA strands to create a transcription bubble and to unwind the DNA at the transcription start site. TFIIB dissociates as Pol II starts
transcribing. A subunit of TFIIH phosphorylates the CTD of Pol II and puts Pol II into the elongation phase (Lodish 2013). After initiation, Pol II is often found in a promoter-proximal paused state. P-TEFb (positive transcription elongation factor b) which is a cyclin-dependent kinase is known to play a role in productive elongation of Pol II partly by regulating the pause complex of DSIF (DRB sensitivity inducing factor) and NELF (Negative elongation factor) (Levine 2011). When Pol II reaches the end of a gene it gradually slows down over the terminator, and the poly(A) signal sequence appears in the nascent transcript. The 3’ end cleavage and polyadenylation complex binds to the poly(A) signal sequence and cleaves mRNA from chromatin. Pol II continues to transcribe the template DNA even after the mRNA release; however, these transcripts are short-lived and are quickly degraded by exonucleases (Lodish 2013). Intercalation of chemicals into the coding DNA or DNA damage induces Pol II transcription arrest that then attracts ubiquitin ligases for the degradation of arrested complexes.

4.1.5 Inhibition of RNA Pol II transcription

Several chemicals such as Streptolygidin, Streptovaricin, Adriamycin and Rifamycin inhibit Pol II transcription by interfering with the formation of the pre-initiation complex (Bensaude 2011). Actinomycin D blocks Pol II transcription at higher doses by binding GC-rich regions in the coding sequences. α-amanitin is a widely used specific inhibitor of Pol II transcription. It is a cyclic peptide of 8 amino acids and possibly the deadliest amatoxin isolated from the mushrooms of Amanita genus (Bensaude 2011). It is primarily known to inhibit Pol II transcription at 10-20 μg/mL concentration with Pol III being sensitive at higher doses. Pol I is completely insensitive to α-amanitin. One way by which α-amanitin blocks Pol II transcription is by its interaction with the bridge-helix of Pol
II and interfering with Pol II elongation (Bushnell et al. 2002, Gong et al. 2004). The other way is by stimulating the ubiquitination of RPB1, the largest subunit of Pol II (Szeberenyi 2006). These combined actions make Pol II sensitive to α-amanitin at 10-20 μg/mL concentration (Nguyen et al. 1996). A major limitation of α-amanitin is that it is a slow acting polymerase inhibitor. Reports indicate that it takes about 24-36 hrs for it to effectively inhibit Pol II transcription (Bensaude 2011).

We wanted to test the effects of Pol I and/or Pol II transcription inhibition on R1 expression. These results would determine the type of RNA polymerase that transcribes R1 under heat shock conditions.

4.2 Materials and Methods

4.2.1 The inhibition of RNA Pol I transcription using Actinomycin D

Actinomycin D (Sigma, cat. no. A9415) at a very low concentration (0.08 μg/mL) was used to inhibit RNA Pol I transcription in Drosophila Schneider 2 (S2) cells grown in culture. Precise control of Actinomycin D dosage prevented the use of Drosophila larvae for these experiments. S2 cells were grown in four 25 cm² culture flasks using 5 mL of the Schneiders’ cell medium (Gibco) with 10% FCS (Fetal calf serum). Each of the S2 cell populations in these 25 cm² flasks received a different treatment. Cells in one flask were treated with 0.08 μg/mL of Actinomycin D for 6 hrs. Cells in the second flask were heat shocked for 1 hr by suspending the culture flask in a 37°C water bath. This flask did not receive Actinomycin D treatment. Cells in the third flask were treated with 0.08 μg/mL Actinomycin D for 6 hrs and then suspended in a 37°C water bath for a 1 hr heat shock. While suspended in the water bath, the S2 cells in this flask were still exposed to Actinomycin D. Cells in the fourth flask received neither the heat shock nor the
Actinomycin D treatment. This flask was kept at room temperature until the cells in other flasks were ready for RNA isolation. Total RNA was isolated using TRIzol (Invitrogen) from S2 cells in these four cell culture flasks. Total RNA samples were treated with DNase I (Promega), and the RNA was ethanol precipitated. First strand cDNA synthesis was performed with these RNA samples using specific primers (Table 1) and MMuLV reverse transcriptase (New England BioLabs). These cDNA samples were used for semi-quantitative RT-PCR and quantitative RT-PCR using specific primers (Table 1).

4.2.2 Inhibition of RNA Pol II transcription using α-amanitin

α-amanitin (Sigma, cat. No. A2263) at 15 μg/mL was used to inhibit Pol II transcription in S2 cells. S2 cells were grown in four 25 cm² culture flasks using Schneiders’ cell medium (Gibco) with 10% FCS (Fetal calf serum). Each of the S2 cell cultures received a different treatment. Cells in one flask were treated with 15 μg/mL α-amanitin overnight. Cells in the second flask were heat shocked for 1 hr by suspending the culture flask in a 37°C water bath. This flask did not receive α-amanitin treatment. Cells in the third flask were treated with 15 μg/mL α-amanitin overnight and then suspended in a 37°C water bath for a 1 hr heat shock. The S2 cells in this flask were still in the presence of α-amanitin during heat shock treatment. Cells in the fourth flask did not receive either the heat shock treatment or the α-amanitin treatment. This flask was kept at room temperature until the cells in the three other flasks were ready for RNA isolation. Total RNA was isolated using TRIzol (Invitrogen). These total RNA samples were treated with DNase I (Promega), and the RNA was ethanol precipitated. First strand cDNA synthesis was performed with these RNA samples using specific primers (Table 1) and
MMuLV reverse transcriptase (New England BioLabs). These cDNA samples were used for semi-quantitative RT-PCR and quantitative RT-PCR using specific primers (Table 1).

4.3 Results and Discussion

While $R_2$ is transcribed by RNA Pol I, nuclear run-on experiments by Eickbush’s group showed that there are hardly any $R_1$ transcripts beyond 1 kb of the 5’ end of $R_1$, leading to a hypothesis that Pol I dissociates from the template somewhere within the first 1 kb of $R_1$ (Ye and Eickbush 2006). However, experiments described in Chapter 1 show that $R_1$ transcription is induced by heat shock. Since Pol I does not normally transcribe beyond the 5’ end of $R_1$, it is not clear which RNA polymerase is responsible for transcribing $R_1$ under heat shock conditions.

A low dose (0.08 μg/mL) of Actinomycin D was used to specifically inhibit Pol I transcription. ITS2 transcription was used as a marker for Pol I transcription. Semi-quantitative RT-PCR, as well as qRT-PCR data indicated that the ITS2 expression was reduced by about half upon Actinomycin D treatment (Figures 4.4, 4.5). While Pol I mediated transcription was inhibited by Actinomycin D, Pol II transcription was unaffected, as indicated by unchanged expression of an Actin 5C intron-exon junction (Figure 4.4). To amplify this intron-exon junction of Actin 5C gene we designed a forward primer within the second intron and a reverse primer (used for cDNA synthesis) in the third exon of the gene (Table 1). Since introns are expected to degrade quickly, pre-mRNAs that contain intron-exon junctions of any housekeeping gene such as Actin 5C are expected to have a very short half-life, and thus serve as indicators of active Pol II transcription. $R_1$ expression was slightly downregulated upon Actinomycin D treatment in the absence of heat shock but significantly elevated upon heat shock alone (Figures 4.4, 4.5). When heat
shock was preceded by Actinomycin D treatment, \( R1 \) expression remained similar to the untreated control (Figure 4.4). These results indicated an active role of Pol I in transcribing \( R1 \) under heat shock conditions because heat shock treatment failed to induce \( R1 \) expression in the presence of Actinomycin D.

**Figure 4.4** The effect of Actinomycin D inhibition of RNA Pol I transcription on \( R1 \) expression. Semi-quantitative RT-PCR gel pictures show that the inhibition of Pol I transcription prevents induction of \( R1 \) expression upon heat shock.

**Figure 4.5** The effect of Actinomycin D inhibition of Pol I transcription on \( R1 \) expression as measured by qRT-PCR. Ct values for all of the genes were normalized to \( Actin 5C \) expression.
Assuming all heat shock induced transcription is mediated by Pol II we asked if the transcriptionally active Pol II localizes to nucleoli upon heat shock. We stained non-heat shocked and heat shocked third instar larvae with anti-\textit{Drosophila} Pol II antibody (Abcam, cat. No. ab5408) that is specific for the phosphorylated carboxy terminal domain of RPB1. This antibody specifically detects transcriptionally active Pol II complexes. Although there were strong signals within the nuclei of the non-heat shocked as well as heat shocked tissues and distinct foci in the nuclei of heat shocked cells, there was no nucleolar staining in the cells with either of these treatments (Figure 4.6). While this is largely a negative result, it shows that there was no detectable translocation of Pol II to the nucleolus upon heat shock.

![Staining with anti-pol II pCTD](image)

Figure 4.6 The immunofluorescence images showing the localization of transcriptionally active Pol II in the nuclei of heat shocked and non-heat shocked third instar larvae.

In order to directly rule out a possibility of the active involvement of Pol II in \textit{R1} expression, α-amanitin at 15 μg/mL was used to selectively inhibit Pol II transcription in
S2 cells. The expression of *Actin 5C* intron-exon junction served as an indicator for Pol II transcription. Under these conditions, downregulation of the *Actin 5C* intron-exon junction expression indicated efficient Pol II transcription inhibition (Figures 4.7, 4.8), whereas, no change in the ITS2 expression showed that Pol I transcription was unaffected by α-amanitin treatment. *R1* expression remained high even after α-amanitin treatment and heat shock indicating there was no active involvement of Pol II in expressing *R1* (Figures 4.7, 4.8).

![Figure 4.7](image1.png)

Figure 4.7 The effect of α-amanitin on RNA Pol II transcription inhibition on *R1* expression. Semi-quantitative RT-PCR gel images show that the inhibition of Pol II transcription fails to suppress *R1* expression upon heat shock.

![Figure 4.8](image2.png)

Figure 4.8 The effect of Pol II transcription inhibition on *R1* expression as measured by qRT-PCR. Ct values for all of the genes were normalized to ITS2 expression.
The data presented so far shows that \( R1 \) is transcribed by Pol I upon heat shock. As far as we know, this is the first functional association between the Pol I transcription machinery and heat shock. However, since there is hardly any transcription beyond 1 kb of \( R1 \) under normal non-heat shock conditions, two possibilities could explain the \( R1 \) transcription upon heat shock. Either Pol I could elongate from a potential pause site somewhere within the first 1 kb of \( R1 \), or Pol I could be recruited onto \( R1 \) upon heat shock. Although in vitro experiments have shown that U-tracts (stretches of 7 or more uridines in rRNA) could pause Pol I (David Schneider, personal communication), there is no in vivo data in the literature that would support either of these two possibilities. Pol I is not known to have a regulatory pause site anywhere within the \( rDNA \), and Pol I is not known to be recruited to any genomic region other than the \( rDNA \) promoter located within the IGS region. The \( rDNA \) promoter is located several kilobases away from \( R1 \). If heat shock was to influence Pol I activity or recruitment, or if normally heterochromatic \( R1 \) inserted \( rDNA \) units were to become euchromatic upon heat shock, we would have seen upregulation of the \( rDNA \) transcription as well. However, \( rDNA \) transcription remained unchanged between heat shock and non-heat shock treatments (Chapter 1). Therefore, it remains important to test if \( R1 \) is co-transcribed along with the 28S rRNA, or if \( R1 \) is selectively transcribed by Pol I upon heat shock. These results would help us understand whether heat shock permits Pol I to read-through into \( R1 \) or recruits Pol I for \( R1 \) expression. Experiments described in Chapter 5 test these hypotheses.

4.4 References


CHAPTER 5: RNA POL I READS THROUGH TO TRANSCRIBE R1 UPON HEAT SHOCK

5.1 Introduction

5.1.1 RNA Pol II elongation from a pause site is a rate-limiting step in transcription

Until recently, RNA polymerase recruitment to the promoter was regarded as the rate-limiting step during transcription. However, recent evidence suggests that in the case of many genes, the release of the promoter-associated Pol II could act as a rate-limiting step as well. The first direct evidence of transcriptional regulation by Pol II release came from a set of developmentally regulated segmentation genes such as slp1, engrailed, and wingless in the Drosophila embryo (Wang et al. 2007). Many Hox genes or genes encoding tissue determinants, or components of cell signaling pathways also contain Pol II associated with their promoters prior to their activation (Zeitlinger et al. 2007). Several studies using chromatin immunoprecipitation (ChIP) indicated the presence of ‘stalled’ Pol II on the promoters of certain genes (Levine 2011). However, more than 20 years ago, several studies independently confirmed the presence of promoter-paused Pol II on Drosophila heat shock genes, as well as on c-myc and HIV early genes (Bentley and Groudine 1986, Gilmour and Lis 1986, Kao et al. 1987). This ‘paused’ Pol II was regarded as a subset of ‘stalled’ Pol II that binds the template DNA, undergoes promoter clearance and after production of short abortive transcripts of about 20-50 bp, stably pauses downstream of the transcription start site (+1) to wait for an ‘activation signal’.

Heat shock acts as one of the activating signals that triggers release of paused Pol II to provide rapid and robust expression of the heat shock genes (Gilchrist et al. 2010). The transcription of genes in which Pol II pauses downstream of the transcription start site is controlled by Pol II elongation rather than initiation. Paused Pol II is phosphorylated
on serine-5 residue within the heptapeptide C terminal repeat domain mentioned in Chapter 4. Alternatively, release of paused Pol II from the promoter is associated with phosphorylation of serine-2 in the C terminal repeats and is mediated by P-TEFb that phosphorylates DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor) and thus the CTD of Pol II (Boehm et al. 2003). A consensus definition of ‘paused’ Pol II is the activated Pol II positioned between 20-50 bp downstream of the transcription start site (+1) and associated with a 5’ cap modified nascent transcript (Levine 2011). Genome-wide studies speculated that as many as one-third of the genes in a typical genome have ‘paused’ Pol II complexes at some point in the organisms’ life cycle (Levine 2011). However, there is an inverse relationship between the half-life of paused Pol II and fresh Pol II initiation from a gene’s promoter; that is, stably bound paused Pol II prevents the formation of the new pre-initiation complex at that gene’s promoter (Shao and Zeitlinger 2017).

Efforts have been made to look for the prevalence of sequence motifs where Pol II tends to pause. Data revealed a set of sequence elements near Pol II pause sites. The pause button motif, a GAGA motif, a DPE, an Inr or an inverse GAGA element were all detected at or near the 3’ end of the nascent transcript attached to paused Pol II complexes (Hendrix et al. 2008). Heat shock factor (HSF) and GAGA factor (GAF) have been implicated in heat shock mediated pause release of Pol II (Duarte et al. 2016).

Both GAF and HSF play major roles in Pol II mediated induction of heat shock genes in *Drosophila*, and they recruit chromatin remodelers (Wilkins and Lis 1997). GAF is present on the promoters of all heat shock genes along with a paused Pol II. However, upon heat shock, HSF binds the promoter, GAF association persists, but GAF now binds
throughout the gene body rather than being confined just at the promoter, and Pol II escapes into productive elongation (Wilkins and Lis 1997). Also, mutations in GAGA elements (GAF binding sites) in the heat shock genes are reported to inhibit the formation of paused Pol II complex near the promoters of those heat shock genes (Croston et al. 1991). HSF also has been found to restructure nucleosomes on the HSP26 gene promoter (Sandaltzopoulos et al. 1995).

Data described in Chapter 4 showed that \( R1 \) is transcribed by Pol I upon heat shock. We further wanted to test whether \( R1 \) upregulation is a result of Pol I recruitment or elongation. We also wanted to address whether \( R1 \) transcription begins within \( R1 \) or within the 28S rDNA immediately upstream of \( R1 \).

### 5.2 Materials and Methods

#### 5.2.1 PRO-seq data to assess Pol I densities

To test the hypotheses mentioned above, we used the PRO-seq (Precision run-on sequencing) data set submitted to NCBI-GEO (gene expression omnibus) (accession # GSE77607) by Duarte et al. (2016). These data contain whole genome raw sequence reads from non-heat shocked and 20 min heat shocked Drosophila S2 cells. Some of the other treatments in this data set include individual RNAi mediated depletions of HSF and GAF with or without heat shock. While PRO-seq experiments and dataset submission was performed by Lis lab (Duarte et al. 2016), we performed additional computational analysis for \( R1 \) and rDNA.

PRO-seq is a technique wherein the density of transcriptionally active polymerases on any genomic region is measured. It takes advantage of the fact that the number of nascent transcript reads of any genomic region are directly proportional to the density of
active polymerases on that region (Mahat et al. 2016). In this technique, nuclei are isolated from cells. These nuclei are incubated with Biotinylated NTPs (nucleotide triphosphates) so that the active polymerase molecules on any genomic region would incorporate a few of these labeled NTPs. RNA isolation, RNA fragmentation, and enrichment for Biotinylated RNA molecules is performed immediately following the run-on transcription. PRO-seq technique provides a resolution of 2-3 basepairs. 5’ ends of the enriched RNA fragments are de-capped and the adapter oligonucleotide molecules of known sequence are ligated to these fragments on both 3’ and 5’ ends after 5’ end phosphorylation. These adapter molecules provide a template to anneal primers for reverse transcription and PCR amplification. The last experimental step in this technique is the next generation sequencing (NGS) and data analysis (Mahat et al. 2016).

We downloaded the *Drosophila* whole genome PRO-seq data set generated by Duarte et al (2016) from the NCBI-GEO database (accession GSE77607). For further analysis, the raw reads were uploaded to the Galaxy server (Afgan et al. 2016) by FTP (file transfer protocol) (along with indexed custom FASTA (fast adaptive shrinkage thresholding algorithm) files of both R1 and one entire *Drosophila* rDNA unit containing R1 and R2. We used Bowtie2 with default parameters to generate binary versions of Sequence Alignment Map (SAM) files. These binary files are commonly known as a Binary version of SAM (BAM) files. However, to visualize the BAM files using a genome browser like Integrative Genomics Viewer version 2 (IGV2) (Robinson et al. 2011), UGENE browser, or Savant, we converted the BAM files to the Tiled Data File (TDF) format using IGV-tools. TDF files were analyzed after scaling their Y axes equally.
5.2.2 Visualization of GFP tagged HSF within the nuclei of heat shocked and non-heat shocked cells

To test whether HSF localizes to nucleoli upon heat shock, we used GFP tagged HSF fly line (P[w+, UAS–HSF–EGFP]) which was a gift from John Lis. Virgin females from this line were crossed with daughterless:GAL4 males to overexpress HSF. A set of progeny larvae were heat shocked for 1 hr at 37°C, and another set of similarly aged larvae were kept at room temperature for 1 hr. GFP signal was visualized by a Zeiss Axioskop® equipped with SPOT Pursuit® camera and software.

5.2.3 RT-PCRs to detect read-through transcription of R1

To verify the possibility of read-through transcription of Pol I into the R1 elements, we performed a series of RT-PCRs with reverse primers spanning the 5' end of R1 and a common forward primer placed 69 bp upstream of the R1 insertion site in the 28S region. One of the R1 reverse primers was placed +685 bp within R1, the second was at +272, and the third was at +116 bp (Table 1). First strand cDNA synthesis was performed on total RNA isolated from non-heat shocked and heat shocked third instar larvae as described in Chapter 2. However, this time the cDNA synthesis reactions were separate using individual sequence specific reverse primers within R1 (+685 bp, +272 bp and +116 bp) (Table 1). This experiment was designed to test if R1 is selectively transcribed upon heat shock or if it is co-transcribed with the 28S region. If R1 is indeed transcribed by read-through transcription from the 28S region, then PCRs with a forward primer placed within the 28S will generate defined products.
5.3 Results and Discussion

5.3.1 Analysis of the PRO-seq data set with and without heat shock

Upon visualizing the TDF files showing active Pol I density along full length R1 elements for non-heat shocked and heat shocked S2 cells, we observed that R1 was induced all along its length of ~5.3 kb upon heat shock. This observation was consistent with results showing heat shock induced R1 expression described in Chapter 2. However, active Pol I density was heterogeneous along the length of R1 (Figure 5.1) suggesting pause sites.

![Figure 5.1](image)

**Figure 5.1** Densities of “active” RNA Pol I on R1 elements in non-heat shock and heat shock conditions measured by PRO-seq reads per million (RPM). Graphs are generated using IGV2 by scaling equally on their Y axes.

While assessing active Pol I densities over the rDNA, we could not separate the uninserted rDNA units from those that are exclusively inserted with R1, and hence we could not assess 28S-R1 junctions. We observed similar Pol I densities on R2 elements in either heat shocked or non-heat shocked cells (Figure 5.2). Also, while semi-quantitative RT-PCR showed no difference for ITS2 transcript levels with or without heat shock (Chapter 2), we observed significantly lower Pol I densities over 28S region upon heat shock with the PRO-seq data analysis (Figure 5.2). This loss of rDNA transcription during heat shock is supported by previous observations (Ghoshal and Jacob 1996, Zhao et al. 2016).
Figure 5.2 Composit comparison of “active” Pol I densities on 28S region of an *rDNA* unit containing *R1* and *R2*. NHS represents non-heat shock treatment and HS represents heat shock treatment. Graphs are generated using UGENE browser and scaled equally on their Y axes.

Although there is no known interaction between HSF or GAF either with the *rDNA* or with Pol I, we looked at *R1* expression with or without heat shock in the PRO-seq data sets wherein HSF and GAF were selectively depleted using RNAi. Results of these comparisons showed that while the heat shock treatment resulted in induced *R1* expression, in cells lacking HSF or GAF still this heat shock induction was not as robust as with just heat shock treatment with HSF and GAF present (Figure 5.3).

Figure 5.3 IGV2 generated bar graphs that depict the comparison of Pol I densities on *R1* between heat shock and non-heat shock treatments with and without HSF and GAF depletion. NHS stands for non-heat shock treatment. Whereas, HS stands for heat shock treatment.
These results argue that HSF and GAF play active roles in heat shock mediated $R1$ expression. Although physical interaction of HSF or GAF with Pol I has not yet been reported, the data described above suggest that transcriptional response to heat shock might be conserved in eukaryotic cells regardless of the type of polymerase involved.

While comparing active Pol I densities within the treatments described in Figure 5.3, we noticed that while all treatments involving heat shock showed higher Pol I densities on $R1$ compared to non-heat shock treatments, active Pol I densities remained consistently low within the first $\sim$300 bp of $R1$. However, heat shock treatment showed robust accumulation of transcripts after position +300 within $R1$. Therefore, analyzed the TDF file with the heat shock treatment with HSF and GAF present, showing active Pol I densities around the +300 region of $R1$ alongside the actual corresponding $R1$ sequence. A closer look at the Pol I densities upon heat shock around +300 within $R1$ revealed that $R1$ transcripts were more abundant after nucleotide position +349 (T) (Figure 5.3A, B). Interestingly, it turned out that this nucleotide is part of an inverted repeat sequence, similar to the consensus HSF binding site (GAATnTTC) (Trinklein et al. 2004). Multiple HSF binding sites are observed in a typical heat shock gene promoter (Trinklein et al. 2004). However, this inverted repeat sequence could potentially act as a cryptic heat shock element.

5.3.2 Visualization of GFP tagged HSF within cells treated with and without heat shock

To test whether HSF localizes to nucleoli upon heat shock we used GFP tagged HSF fly line (P[w+, UAS–HSF–EGFP]) which was a gift from John Lis (Materials and Methods). Although the overexpressed GFP tagged HSF failed to localize to nucleoli, it aggregated around nucleoli under heat shock conditions (Figure 5.5). At this time we do
not know why HSF would aggregate around nucleoli upon heat shock. Upon actinomycin D mediated inhibition of rDNA transcription, the Pol I transcription complexes are found at the periphery of nucleoli (Craig Pikard, personal communication). The heat shock induced HSF aggregation around nucleoli might indicate an ectopic site for Pol I mediated R1 transcription.

![Graph showing Pol I density upon heat shock](image)

**Figure 5.4** (A) Location of an inverted repeat sequence (GACtCGTC) near +349 within R1. While comparing the treatments shown in panel (B), we observed higher Pol I densities starting at nucleotide 349 within this inverted repeat sequence.

5.3.3 RT-PCRs to show read-through transcription of Pol I into R1 elements

Under normal growth conditions, RNA Pol I has been proposed to dissociate from the template within first 1 kb of the 5’ end of R1 (Ye and Eickbush 2006). To test for 28S-R1 co-transcripts in third instar larvae upon heat shock, we used RT-PCR with a forward primer placed within the 28S region at -69 bp upstream of the R1 insertion site and a series of RT reverse primers that should anneal to R1 transcripts (Table 1).
The first RT primer was placed at a position +685 bp within the $R1$ element. There was a clear upregulation of this 28S-$R1$ co-transcript upon heat shock (Figure 5.6). A second RT reverse primer positioned at +272 in the $R1$ transcript reproducibly produced a fainter product upon heat shock. Finally, a third RT primer now at +116 in the $R1$ transcript consistently failed to show upregulation of a 28S-$R1$ co-transcript upon heat shock due perhaps to a Pol I pause site at or just downstream of +116 under normal non-heat shock conditions (Figure 5.6).

Figure 5.5 Localization of overexpressed HSF-EGFP in non-heat shocked (control) and heat shocked third instar larval gut cells. Panels on the left show DAPI stained nuclei under non-heat shock and heat shock conditions whereas panels on the right show GFP tagged HSF expressed by Gal4 promoter. The lower right panel shows that HSF seems to be aggregating around nucleoli upon heat shock.
Figure 5.6 Semi-quantitative RT-PCRs showing co-transcription of R1 upon heat shock. The cDNAs synthesized with the reverse primer placed at +685 generated an intense band upon heat shock. cDNAs with +272 reverse primer showed a fainter PCR product. Whereas, cDNAs with +116 reverse primer consistently failed to show induction upon heat shock. NHS: non-heat shock, HS: heat shock.

Overall, these data are consistent with Pol I pausing at or within the 5’ end of R1 upon normal growth conditions, but the data indicate that upon heat shock Pol I reads through to transcribe R1 as marked by the modest accumulation of transcripts detected at +272, but abundant transcripts detected at +685. These results are also consistent with the heat shock induced abundance of R1 transcripts detected beyond +349(T) within R1 by PRO-seq data analysis (Figure 5.4).

‘Polymerase pausing’ is not yet characterized as a regulatory mechanism for Pol I transcription. This is the first report which suggests that Pol I transcribes R1 upon heat shock by its release from a ‘pause’ site within the first ~350 bp. As described in Chapter 5, polymerase pausing was first proposed as a regulatory mechanism for Pol II. Although Pol II does not transcribe R1 (Chapter 4), and since all of the eukaryotic polymerases share some subunits (Introduction, Chapter 4), it would be worthwhile to look for sequences within R1 that are known to be associated with a ‘paused’ Pol II. Therefore, we searched for the presence of a sequence motif within the 5’ end of R1, typically associated with Pol II pausing. One of these sequence motifs called a pause button; (T/GCGPuWCG) (Figure 5.7) immediately stood out at the very 5’ end of R1. Surprisingly,
$R1$ starts with a perfect ‘pause button’ sequence motif described earlier (Hendrix et al. 2008). The first six nucleotides of $R1$ along with a nucleotide (T) from the 28S region immediately upstream of $R1$ insertion site make a perfect ‘pause button’ (Figure 5.7 right panel).

Figure 5.7 $R1$ starts with a perfect pause button sequence. The left panel shows the consensus pause button motif sequence wherein K represents a keto (T or G), R is a purine (A or G), and W is a weak base (A or T) (Hendrix et al. 2008). The panel on the right shows two pause button like sequences identified within $R1$ and the inverted repeat sequence starting at +346. 5 nucleotides from the 28S region upstream of $R1$ insertion are shown with capitalized T indicating the nucleotide after which $R1$ insertions are found.

This sequence motif was first described in developmentally regulated genes of $Drosophila$ and was associated with ‘paused Pol II’ within promoter proximal sequences (Hendrix et al. 2008). Pol II was shown to pause at or about 20-50 nucleotides downstream of this motif (Levine 2011). We also observed an imperfect ‘pause button’ like sequence at +175 of $R1$ (Figure 5.7 right panel). These observations are consistent with the experimental data described in Figure 5.6, indicating that Pol I pauses within the first 350 bp of $R1$.

The results described in this chapter also suggest the involvement of HSF and GAF in the Pol I mediated $R1$ transcription, since RNAi depletions of HSF and GAF reduced the amount of active Pol I on $R1$ elements during heat shock. Also, the RT-PCR
experiments strengthen the possibility of the presence of a Pol I pause site around +116 within \( R1 \) and the co-transcription of \( R1 \) upon heat shock. Although it was interesting to note the presence of the cryptic heat shock element near +349 and two pause button like sequences within the 5' end of \( R1 \), further experiments are needed to precisely determine their functional significance and the exact site of Pol I pause.

5.4 References


CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Conclusions

Under normal conditions, $R_1$ and $R_2$ are transcribed at very low levels (Ye and Eickbush 2006). However, our data shows that upon heat shock $R_1$ is expressed at much higher levels (Chapter 2), and upon nucleolar stress, the $R_2$ expression is induced (He et al. 2015). These observations suggest that although $R_1$ and $R_2$ belong to the same class of retrotransposons, they differ in their regulated expression. Distinct changes in the $rDNA$ chromatin structure upon different stress conditions might contribute to the differential expression of $R_1$ and $R_2$ retrotransposons.

The molecular mechanism underlying the expression of all heat shock genes is well elucidated. All heat shock induced transcription of HSP genes in a eukaryotic cell is mediated by Pol II (Ashburner and Bonner 1979). Even the retrotransposons like $L1$ that respond to heat shock have their own internal Pol II promoters (Lavie et al. 2004). Our data shows the first functional association between RNA Pol I and heat shock (Chapter 4). The PRO-seq data analysis described in Chapter 5 serves as an independent verification of our experimental results showing heat shock-induced $R_1$ expression. PRO-seq data analysis with HSF and GAF depletions suggests active roles of HSF and GAF in the Pol I mediated $R_1$ expression. These observations also suggest that these two factors may be functionally conserved in response to heat shock, regardless of the type of RNA polymerase involved.

The series of RT-PCR experiments described in Chapter 5 support the pre-existing hypothesis that Pol I falls off within 1 kb of $R_1$. However, these results also indicate for the first time that instead of dissociating from the template, Pol I pauses within the first
350 bp of \( R1 \), and is released from the pause upon heat shock. PRO-seq data analysis hints at a region in R1 (around +349 T) from where we begin to see the accumulation of labeled \( R1 \) transcripts (active Pol I) upon heat shock. This observation is consistent with our RT-PCR results showing modest accumulation of \( R1 \) transcripts at +272, but a robust increase in transcripts at +685. Although we have not ruled out the possibility of \( R1 \) having its own promoter, our RT-PCR data with a forward primer placed within the 28S region immediately upstream \( R1 \) insertion site indicates read-through transcription from the 28S region into \( R1 \).

Based on these results, we propose a model (Figure 6.1) wherein Pol I ‘pauses’ around nucleotide position +116 within \( R1 \). However, heat shock releases Pol I from this pause site and Pol I continues to transcribe \( R1 \). Transition from initiation to elongation modes by RNA polymerases as a rate-limiting step in transcription is an emerging notion. All of the literature so far is consistent with this notion for Pol II. Our data indicate that even Pol I could transcribe by elongation control rather than initiation control under certain conditions such as heat shock.

**6.2 Future Work**

It would be interesting to test if the inverted repeat sequence depicted in Figure 5.4 acts as a cryptic heat shock element by cloning a full length IGS region along with the 28S region normally upstream of \( R1 \) and the 5’ end of \( R1 \) with or without the inverted repeat sequence followed by a reporter gene such as \( LacZ \) into a plasmid suitable for transfection into \textit{Drosophila} S2 cells. Semi-quantitative RT-PCR measurements of \( LacZ \) transcripts with and without heat shock would tell us whether the inverted repeat sequence acts as an active heat shock element.
Figure 6.1 Proposed model showing RNA Pol I stalled under non-heat shock conditions near the ‘pause button’ motif (yellow) located at the beginning of $R1$. Thin irregular lines attached to Pol I complexes denote rDNA transcripts. Heat shock releases Pol I allowing it to read into $R1$ resulting in accumulation of $28S$-$R1$ co-transcripts with $R1$ sequences represented as the thicker line. Inverted repeat sequence at +346 is shown in green.

Similar experiments could be performed to directly test if the ‘pause button’ motifs within $R1$ act synergistically to pause Pol I around +116. Cloning of the $R1$ 5’ end of with and without the ‘pause button’ motifs downstream of Pol I promoter and placing a reporter gene downstream of $R1$ in a plasmid would provide us with a construct to transfec into S2 cells. Measurement of reporter gene expression in the transfected cells would indicate the effectiveness of these pause button motifs.

Lastly, our lab is currently working on testing the effect of aging on the rDNA transcription. We recently observed induced $R1$ expression with aging in flies aged for 5, 30 and 50 days. Expression of Actin 5C and $R1$ was undetectable in flies aged for 5 days but Actin 5C expression remained constant in flies aged for 30 and 50 days while $R1$ expression was induced (Figure 6.2).
Figure 6.2 R1 expression in aged flies. R1 transcription seems to be induced in the flies aged for 50 days compared to that in the flies aged for 30 days.

This is a very preliminary observation that needs more biological replicates with proper loading controls. However, it would be interesting to compare R1 expression alongside rDNA expression in aged flies. Unpublished data from some laboratories claim increases in rDNA transcription upon aging. R1 expression could be a consequence of this increased Pol I activity. Conversely, R1 expression might be one of the contributing factors to rDNA instability observed upon aging in arthropods (Guarente 1997, Vijg and Suh 2013).

To summarize, we observed R1 transcription upon heat shock in Drosophila melanogaster. After assessing R1 and R2 expression under different stress conditions we showed that both of these LINE elements in Drosophila rDNA behave differently in their regulated expression. Actinomycin D and α-amanitin experiments described in Chapter 4 confirmed that the heat shock induced R1 expression is mediated by Pol I. As far as we know this is the first report showing induced Pol I transcription upon heat shock. PRO-seq data analysis independently verified our observations and suggested possibility of HSF and GAF involvement in R1 transcription. Our strand specific RT-PCR experiments described in Chapter 5 and the presence of a ‘pause button’ sequence motif
at the beginning of \textit{R1} showed that \textit{R1} is transcribed by elongation of paused Pol I. Further work is needed to confirm the functional association of the ‘pause button’ motifs within the 5’ end of \textit{R1}. Since HSF localized to nucleolar periphery upon heat shock and PRO-seq data suggested its active role in \textit{R1} transcription, it would be interesting to study the subcellular location of \textit{R1} transcription upon heat shock. Overall, the study of the biology of \textit{R1} and \textit{R2} retrotransposons would shed light on their function and explain their evolutionary persistence in the genome of all arthropods.

6.3 References


VITA

Himanshu Raje received his Bachelor of Science degree in Life sciences from University of Mumbai, India in 2003 and Master of Science degree by research from P. D. Hinduja National Hospital affiliated with University of Mumbai in 2009. Himanshu joined Louisiana State University as a graduate student in 2010. Himanshu changed his major from Veterinary medicine to Biological sciences in 2011 and started working in the lab of Dr. Patrick DiMario. Himanshu will graduate with a Doctor of Philosophy degree in Biological sciences in May 2018.