Examining the Role of RNA Secondary Structure on Human RNA Polymerase II Pausing

Meagan N. Paul

University of Colorado, Boulder, mepa0216@colorado.edu

Follow this and additional works at: https://scholar.colorado.edu/honr_theses

Part of the Biochemistry Commons, Biophysics Commons, and the Structural Biology Commons

Recommended Citation


https://scholar.colorado.edu/honr_theses/1078

This Thesis is brought to you for free and open access by Honors Program at CU Scholar. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
“Examining the Role of RNA Secondary Structure on Human RNA Polymerase II Pausing”

Meagan Nicole Paul

Defense Date: March 31st, 2016

Thesis Advisor:
Dr. Dylan Taatjes, Department of Chemistry and Biochemistry

Committee Members:
Dr. Joseph Falke, Department of Chemistry and Biochemistry
Dr. Robert Kuchta, Department of Chemistry and Biochemistry
Dr. Alysia Marino, Department of Physics
# Table of Contents

Abstract.........................................................................................................................................................3

Introduction.....................................................................................................................................................4
  The Central Dogma of Biology ..........................................................................................................................4
  Assembling the Machinery for Transcription .....................................................................................................5
  The Stages of Transcription .............................................................................................................................6

Background ..........................................................................................................................................................7
  The Biological Importance of Pausing...............................................................................................................7
  The Native *HSP70* Promoter ..........................................................................................................................9
  Pausing Regulation by the Nascent Transcript in Bacteria ................................................................................9
  Pausing Regulation by Protein Factors ............................................................................................................11
  RNA Secondary Structure ..............................................................................................................................14

Methodology ......................................................................................................................................................15
  Approach I: The *In Vitro* Transcription Assay ............................................................................................17
  Approach II: SHAPE Chemical Probing ..........................................................................................................20

Detailed Methods ...............................................................................................................................................22
  *In Vitro* Transcription Assay ..........................................................................................................................22
  Preparing RNA for SHAPE ..............................................................................................................................24
  SHAPE Chemical Probing ...............................................................................................................................26
  SHAPE Data Analysis .......................................................................................................................................27

Hypotheses .........................................................................................................................................................27

Results ...............................................................................................................................................................28
  *In Vitro* Transcription Assay ..........................................................................................................................28
  SHAPE Chemical Probing ................................................................................................................................30

Conclusion.........................................................................................................................................................32

Bibliography .......................................................................................................................................................33
**Abstract**

The process of transcription, in which DNA is converted to RNA, is an integral process in the expression of our genetic information. Regulation of transcription is critical for healthy, functioning cells and defects in transcriptional regulation can cause a plethora of ill health effects including cancer and improper cardiac or neuronal development. Human RNA Polymerase II (Pol II) is the enzyme that transcribes all protein-coding and most non-coding RNA genes in the human genome. Within the past 10 years, it has become evident that Pol II function is regulated after it initiates transcription; specifically, the Pol II enzyme pauses after transcribing 20-100 bases of DNA into RNA. This process is highly regulated by numerous protein factors, but no groups have yet addressed the nascent RNA present during transcription as a potential regulatory factor. This project examines the nascent RNA secondary structure and its potential for regulating Pol II pausing at the native *HSP70* human promoter. A combination of approaches was used to dissect the importance of RNA secondary structure in Pol II pausing. First, *in vitro* transcription assays were used to test the role of RNA secondary structure in pausing by incorporating structure disrupting nucleotide triphosphate analogues into the nascent RNA. Second, SHAPE chemical probing and structural prediction algorithms were used to predict the secondary structure of the *HSP70* nascent RNA. By utilizing both *in vitro* chemical methods and directed structural prediction, this project has begun to dissect the role of RNA secondary structure in human Pol II pausing.
Introduction

The Central Dogma of Biology

Francis Crick first summarized the central dogma of biology in 1958, leading to an expanding field of research in the biochemical fundamentals of genetics (Voet, 2011). Transcription, the first step in the central dogma, explains how organisms express their genetic material by converting double-stranded DNA into single-stranded RNA. This RNA can then be used in the process of translation to be converted to proteins that are expressed in the body. Recent research and scientific curiosity have explored the relationship between the regulation of these pathways and human diseases such as cancer, inflammation, and aging. It is the larger goal of this project to study the first step of the central dogma, transcription, and strive to deduce molecular regulation of this important pathway. The research of

Figure 1: The central dogma of biology includes the processes of replication, transcription and translation. Highlighted in yellow is the topic of my project, transcription, which converts DNA to RNA.

Adapted from: Voet, 2011
transcription and its regulation is critical for understanding some of the largest medical questions of our generation including gene expression, development, aging, disease and cancer.

**Assembling the Machinery for Transcription**

The process of converting DNA into RNA requires complex protein machinery to assemble before a gene can be transcribed. The key enzyme that converts double-stranded DNA into single-stranded RNA is called RNA Polymerase II (Pol II). Pol II uses a specific DNA sequence, the promoter, to serve as a scaffold onto which the protein machinery is assembled. These general transcription factors—TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, Mediator, and RNA polymerase II

![Figure 2A](image_url)

*Figure 2A:* A diagram detailing the steps of assembly of each protein factor required for the Pre-Initiation Complex. These protein factors assemble sequentially to allow Pol II to transcribe a gene and allow for high levels of regulation of transcription.

![Figure 2B](image_url)

*Figure 2B:* A diagram depicting the pre-initiation complex formation on a human gene with relevant transcription factors and machinery shown, indicating the complexity and tight regulation of human transcription.

*Adapted from: Israel Journal of Chemistry, 2013*
(Pol II)—assemble on the promoter to form a Pre-Initiation Complex (PIC), as seen in Figure 2A. Together these proteins allow for Pol II to initiate transcription and progress along the gene transcribing DNA into RNA. The binding of these transcription factors allows for a high level of transcriptional regulation to occur via protein-protein and protein-DNA binding. Transcription factors can act as activators or repressors to increase or decrease, respectively, the expression of a single gene (Figure 2B). These binding events act as important regulatory mechanisms for the cell to carefully control gene expression.

The Stages of Transcription

![RNA Pol II](image)

**Figure 3:** A brief schematic showing RNA Polymerase II pausing, then continuing with transcription, forming a nascent RNA strand.

*Adapted from: Core, Waterfall and Lis, 2008*

The procession of Pol II along a gene occurs in three stages—initiation, elongation, and termination. During initiation, Pol II binds to DNA at the promoter as part of the PIC and prepares the DNA for conversion to RNA. In elongation, the Pol II enzyme traverses down the DNA strand reading each DNA base of the template DNA strand and matching it with a complementary RNA nucleotide to synthesize a new strand of RNA. This RNA strand contains important molecular instructions encoding for amino acids, the building blocks of proteins. After Pol II has synthesized
the proper RNA, it dissociates from the DNA in the final step called termination. It has been recently discovered that between the initiation and elongation stages, Pol II undergoes a “pausing” phase (shown in Figure 3) in which Pol II suspends RNA synthesis shortly after initiation. This critical pausing effect of Pol II has been confirmed over the last decade by several scientific groups and has been shown to occur between 20 and 70 bases downstream of the transcription start site (O’Brien and Lis, 1991; Nechaev, et. al., 2010; Core and Lis, 2008). Several groups have also found that Pol II pausing plays a critical regulatory role in 30%-90% of expressed genes (Core and Lis, 2008; Adelman and Lis, 2012). The biological details and mechanism of this pausing phase are currently poorly understood. However, better understanding of this paused phase of Pol II transcription is key to understanding transcription regulation as a whole. This project investigates the mechanism of Pol II pausing as a critical step in regulating transcription of human genes.

Background

While there is still much to learn about how transcription is regulated in the cell, there has been a significant amount of research done to substantiate the importance of Pol II pausing in transcription.

The Biological Importance of Pausing

A key area of research in my project first relies on understanding how Pol II operates and is regulated. Rougvie and Lis first published work introducing Pol II pausing in 1988. Their research discovered that after Pol II binds to DNA, it begins transcribing the DNA but then stops, or pauses,
shortly after beginning (Rougvie and Lis, 1998). They hypothesized that this could be an integral step for cell regulation, allowing cells to synchronously activate genes for transcription, for example in developing or cancerous cells (Henriques, 2013). Zeitlinger and Young substantiated the biological importance and prevalence of pausing in their 2007 paper, which used Chromatin Immunoprecipitation (ChIP) assays in Drosophila to identify which genes contained paused Pol II. They found that genes that display paused Pol II are highly enriched in functions for development, including neurogenesis and muscle differentiation (Zeitlinger and Young, 2007). Zeitlinger and Young also categorized non-paused Drosophila genes and discovered that these genes were typically ubiquitously expressed and played a role in general cell functions such as metabolism and cell proliferation. Using ChIP-chip assays then confirming their results by permanganate footprinting, this group was further able to calculate that, on average, Pol II pauses 50bp downstream of the promoter in Drosophila pause-regulated genes (Zeitlinger and Young, 2007).

The Adelman group further identified and confirmed several critical functions of this pausing effect: using similar ChIP-chip assays and permanganate footprinting, Adelman and Muse identified that about 1/3 of paused genes had functions related to development and that these genes were overrepresented in their ability to respond directly to stimuli (Muse and Adelman, 2007). Adelman hypothesized functionality for Pol II pausing, including the ability for genes to quickly and synchronously be turned on in the case of an environmental signal, i.e. during development or in cancerous cells, and the ability for transcription to be controlled at multiple levels, allowing for greater molecular regulation (Adelman and Lis, 2012). Adelman and Muse propose that pausing allows Pol II to be poised for activation after PIC assembly and transcription initiation, allowing for a more rapid response to stimuli (Muse and Adelman, 2007). Since the
discovery of Pol II pausing, groups have substantiated the idea that Pol II pausing is critical for proper development and the control of disease through gene expression.

The Native *HSP70* Promoter

For *in vitro* study of Pol II pausing, the native *HSP70* promoter was chosen as the gene of choice for several reasons. Most importantly, the *HSP70* gene has been confirmed by numerous groups to be regulated by Pol II pausing, so we would expect to see both paused and elongated Pol II along the promoter (Core and Lis, 2008; Adelman and Lis, 2012; Muse and Adelman, 2007). Second, *HSP70* is an endogenous human gene and is highly conserved from *Drosophila* to humans. Therefore, it has direct biological relevance over an engineered, non-biological promoter and allows for more direct application of these results to *in vivo* studies in the future. Finally, *HSP70* is a widely studied promoter in the field of Pol II pausing but most of the work using *HSP70* has been done *in vivo*. By choosing the *HSP70* promoter, we can compliment the previous cell-based work done by others in the field, while also adding a new approach to the *HSP70* studies using our *in vitro* approach.

Pausing Regulation by the Nascent Transcript in Bacteria

Most of the research studying pausing by transcriptional polymerases has been done in bacterial studies using *E. Coli*. These studies, while investigating a much simpler organism, are very important in substantiating the reason for investigating the role of secondary structure in pausing in human systems. Researchers have found that the secondary structure of the nascent RNA can regulate pausing in bacteria. Wang, et al. has found that in *E. Coli* transcription, an 11-nucleotide
hairpin structure forms in the nascent RNA that is able to regulate pausing (Wang, et. al, 1997). The Wang group has published a model for interaction between the nascent RNA hairpin structure and the RNA Polymerase to stabilize the paused state, shown in Figure 4. This group has found that in in vitro studies of E. Coli transcription, pausing is seen even with no other protein factors present besides the polymerase. They found that when the nascent RNA is digested, by introducing RNase, pausing decreases dramatically. They have also found significant interaction between the RNA hairpin and the polymerase using cross-linking studies (Wang, et al, 1997). This model proves that the nascent RNA secondary structure is critical for pausing in E. Coli systems. While many of the regulatory mechanisms of pausing are not conserved between bacteria and human transcription, the hairpin-stabilized pausing found in bacteria lends justification to study the possibility of this model being a regulatory mechanism in human transcription as well.

**Figure 4:** A proposed mechanism of hairpin-stabilized RNA polymerase pausing in E. Coli. The 11-nt hairpin structure formed by the nascent RNA can independently stabilize pausing in bacteria without the need for other interacting protein factors.

*Adapted from: Wang, et. al. 1997*
Pausing Regulation by Protein Factors

While many groups have hypothesized about the importance and reasoning behind this pausing phenomenon, very little is known about how Pol II pausing is affected by other factors. Some research has been done to study the effects of proteins that bind to Pol II and alter the duration or frequency of pausing. For example, the Gilchrist group has found that a protein complex called Negative Elongation Factor (NELF) successfully induced pausing of Pol II; the Hargreaves group found another protein complex, P-TEFb, that has the opposite effect and reduces pausing of Pol II near the promoter (Gilchrist, 2010; Hargreaves, 2009). The Taatjes and Espinosa labs have found that another protein complex, the CDK8 module, releases Pol II from pausing in conjunction with P-TEFb and that the overexpression of the CDK8 protein can be oncogenic (cancer-causing) (Donner, 2010).

In addition to studying the role of protein factors in pausing, some groups investigated the potential for specific DNA sequences to regulate Pol II Pausing. In 2008, work by the Adelman lab used ChIP-chip assays in Drosophila to identify the role of NELF and the GAGA promoter element in pausing. The GAGA promoter element is a repeat of GA bases in the DNA template sequence that binds a specific transcription factor. The Adelman

<table>
<thead>
<tr>
<th>Motif</th>
<th>LOGO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverted GAGA</td>
<td><img src="https://example.com/inverted_gaga_logo.png" alt="Inverted GAGA Logo" /></td>
</tr>
<tr>
<td>GAGA</td>
<td><img src="https://example.com/gaga_logo.png" alt="GAGA Logo" /></td>
</tr>
<tr>
<td>Inr</td>
<td><img src="https://example.com/inr_logo.png" alt="Inr Logo" /></td>
</tr>
<tr>
<td>Pause button</td>
<td><img src="https://example.com/pause_button_logo.png" alt="Pause button Logo" /></td>
</tr>
</tbody>
</table>

Figure 5: The combination of promoter core elements the Hendrix and Levine group identified as a way to identify and predict paused genes.

Adapted from: Hendrix and Levine, 2008.
lab found that 39% of paused genes contain both the GAGA element and bound NELF on the gene. The authors used this occurrence to hypothesize that NELF and GAGA can facilitate pausing through some collaborative mechanism, but that this pausing may be a transient but necessary transcriptional checkpoint (Lee and Gilmour, 2008). While the GAGA element seemed to be predictive of some paused genes, no group has identified a specific sequence that acts directly as a pause button motif at which Pol II recognizes the sequence and pauses directly. Other than regulation by protein factors, some groups have studied the ability of certain sequence-specific elements to regulate Pol II pausing. Hendrix and Levine used ChIP-chip assays of Drosophila to identify a combination of promoter elements – the GAGA factor, Inr, and “Pause Button” sequence elements—that are highly enriched in paused genes, shown in Figure 5 (Hendrix and Levine, 2008). They found that this combination of promoter core elements are enriched 26.6 fold in paused genes and that less than 2% of non-paused promoters contain this same combination of elements (Hendrix and Levine, 2008). While this combination of promoter elements is shown to be overrepresented in paused genes, the authors are not suggesting that Pol II actually pauses at any of these sites. The mechanism of pausing and being able to predict where along the gene Pol II pauses is not addressed by this pause button motif. Despite finding some predictive sequences that tend to indicate paused Pol II, no DNA sequence motif has been found that Pol II recognizes and pauses at specifically. Therefore, studying the RNA secondary structure as a potential structural motif that induces pausing, as opposed to a DNA sequence, is a particularly interesting and relevant project. Furthermore, Hendrix and Levine found that the promoters of paused genes, including human HSP70, are especially GC-rich. The authors hypothesized that the increased GC content causes the template DNA to have a higher melting point, which may act as a “speed-bump” for the polymerase. Another possible function of the GC content could be to stabilize nucleosome
architecture to act as another regulatory step in transcription (Hendrix and Levine, 2008). While
GC content does not suggest a direct sequence that affects pausing, this interesting tendency for
paused genes to be GC rich may be suggesting some conserved RNA structural elements, as
opposed to conserved sequences that could be affecting Pol II pausing. While these authors
purport to have found a combination of sequence-related motifs that are overrepresented in
paused genes, no group has found a DNA sequence motif where Pol II pauses directly. Therefore,
understanding where and how Pol II pauses may be a question addressed by studying the nascent
RNA secondary structure instead.

While very little research has been done on the role of nascent RNA secondary structure on Pol II
pausing, one group attempted to address the issue using a structural prediction algorithm. As a
further investigation into NELF as a regulatory protein in Pol II pausing, the Adelman group used a
secondary structure prediction web-based program called Mfold to predict significant RNA
secondary structures that may be a regulator in Pol II pausing. However, by using structural
prediction without chemical probing, they found no substantial conserved sequence motifs or
secondary structural motifs common to Drosophila paused genes (Lee and Gilmour, 2008).
Without further investigation and chemical probing methods, this secondary structure prediction
using Mfold is unable to determine whether a conserved secondary structure could be key to
regulating Pol II pausing in paused genes. Many groups have investigated the role of protein
factors and sequence elements in facilitating pausing, however no research has yet thoroughly
investigated the secondary structure of the nascent RNA as a potential regulatory element in Pol II
pausing in human systems, and this gap in knowledge will be the focus of my project.
RNA Secondary Structure

Since RNA is single-stranded, as opposed to double-stranded DNA, it cannot form the typical double-helix structure we see in DNA. Rather, RNA can form its own complex secondary structures by folding and chemically binding in unique ways to itself along the RNA strand. RNA can form base pairs just like DNA does to find its lowest free energy conformation. In RNA, adenosine will base pair with uracil, and guanine will base pair with cytosine in interactions stabilized by hydrogen bonding. This extensive base pairing can lead to a variety of RNA secondary structures. Some typical RNA secondary structures are shown in Figure 6. Some of the most stable and common structures are hairpins, where the RNA shares a long complementary sequence and is able to base pair with itself to form a hairpin as shown above. The nascent RNA formed by Pol II pausing, like other RNA molecules, can form complicated secondary structures. These RNA secondary structures can bind to proteins with very high affinity and specificity. Examples of this phenomenon are used in creating pharmaceutical aptamers and in SELEX to create RNAs that will bind to specific proteins or other targets. Because of the ability of

![Figure 6: Some simple secondary structures formed by single-stranded mRNA. These structures can be predicted using the SHAPE-prediction algorithm. Adapted from: Lehninger Principles of Biochemistry, 2008](image-url)
RNA to bind to proteins, the nascent RNA extending from paused Pol II may be able to act as a scaffold to bind to proteins and regulate pausing.

A complete understanding of the factors which regulate Pol II pausing will allow researchers to determine how general transcription factors and other proteins are able to regulate transcription as a whole. And while the mechanisms of pausing remain largely unknown, I seek to look specifically at the role of the structure of the nascent RNA in this process.

**Methodology**

I will investigate the role of RNA secondary structure in RNA Pol II Pausing using a dual approach to provide both biological relevance and structural prediction (shown in Figure 7). The dual approach to this project includes two complementary techniques: an *in vitro* transcription assay and SHAPE prediction chemistry. The *in vitro* transcription assay will assess the biological relevance of Pol II pausing by adding in nucleotide analogues that disrupt secondary structure and then viewing pausing versus full length RNA produced. SHAPE prediction chemistry will then allow me to determine the secondary structure of the nascent RNA using chemical probing followed by computer algorithm analysis.
*Figure 7:* A diagram outlining the dual-approach strategy of my project to address both biological relevance of pausing and to predict the secondary structure of the nascent HSP70 RNA using an *in vitro* transcription assay and SHAPE chemical prediction, respectively.
**Approach I: The *In Vitro* Transcription Assay**

We first looked at the biological relevance of the secondary structure of the nascent RNA in Pol II pausing by using an *in vitro* transcription assay. The Taatjes Lab has developed a strategy to probe Pol II pausing in a cell-free *in vitro* system. The transcription assay is a biochemical tool that allows us to conduct transcription in a controlled manner using DNA templates and transcriptional proteins of our choosing. After allowing transcription to proceed, we can infer the fraction of reactions with a paused Pol II as compared to the fraction of reactions with a fully elongating Pol II by comparing the sizes of RNA produced. The DNA template used was the native *HSP70* promoter shown in **Figure 8**. The fully elongated Pol II will produce +216 nucleotide RNA transcript, whereas the paused Pol II will produce some shorter RNA, most likely between 20-100 nucleotides in length. By comparing these two transcription products, we can analyze the fraction of Pol II that is paused versus elongated.

**Figure 8:** A diagram of the *HSP70* promoter, indicating the TATA binding region at -24bp and two regulatory binding regions (MBS and HSE, shown). The transcribed gene is the *HSPA1B* gene and runoff transcription is 216 nucleotides long.
To assess the role(s) of the nascent RNA extending from a paused Pol II, we decided to use nucleotide triphosphate (NTP) analogues that have been demonstrated or are likely to disrupt native RNA secondary structures. Literature research of possible analogues uncovered two excellent candidates—tCTP synthesized by the Kuchta Lab at CU Boulder, and Thio-G synthesized by the Tor Lab at UCSD. The structures of each candidate analogue are shown in Figures 9 and 10 and a comparison of both analogues and their published properties can be found in Figure 11 below. Thio-G is created using a (3,4)thienopyrimidine nucleoside core, while tCTP’s 1,3-diaza-2-oxophenothiazine presents a much more significant modification. Both NTP analogues have been shown to be incorporated readily by bacterial polymerase T7 into RNA. While both analogues are successfully utilized by T7 polymerase in the RNA polymerization reaction, structural differences between the human and bacterial polymerases prevent any guarantee from human Pol II being able synthesize RNA using these analogues. Additionally, both analogues have been shown in the literature to disrupt native secondary structure. Thio-G has also been shown to disrupt higher level structures within the RNA such as tertiary and quaternary structure where elements such as hairpins can interact with other RNA elements to form higher-level structures.
However, for our purposes, disrupting secondary structure will also disrupt higher-level structures and will allow us to accurately probe the shape and structure of the nascent RNA. With each analogue, we can replace all or some of the typical GTP or CTP with the fluorescent $^{\text{th}}$GTP (thio-GTP) or tCTP molecules. If secondary structure of the nascent RNA is important for Pol II pausing, we’d expect to see a shift in the ratio of paused polymerase to elongating polymerase. The $^{\text{th}}$GTP has been shown to incorporate successfully in bacteria at a 95% incorporation rate (McCoy 2014) and a native RNA of over 800 bases was created using only tCTP in place of CTP (Stengel, G, Urban, M, Purse, BW & Kuchta, RD, 2010). By assessing two main concerns—the incorporation of the NTP analogues by human Pol II and their ability to disrupt secondary structure—I was able to

<table>
<thead>
<tr>
<th>NTP Analogue Considerations</th>
<th>tCTP (Stengel, Urban, Purse, &amp; Kuchta, 2010)</th>
<th>Thio-G (McCoy 2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Can the analogue be used as a substrate for transcription?</td>
<td>T7 Polymerase</td>
<td>Human Pol II</td>
</tr>
<tr>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
</tr>
<tr>
<td>2. Will the analogue disrupt secondary structure?</td>
<td>Disrupts $2^0$ structure with high hydrophobicity</td>
<td>Disrupts $2^0$ and higher level structure in hammerhead ribozymes</td>
</tr>
<tr>
<td>Evidence: Difference in polyacrylamide gel migration</td>
<td>Evidence: Attenuated ribozyme catalytic activity</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 11:** A comparison chart of both NTP analogues chosen to be used in this procedure. By satisfying both conditions for an NTP analogue as described in the literature, tCTP and Thio-G were chosen for the *in vitro* assay experiments.
choose these two candidate analogues for the *in vitro* transcription assay. The *in vitro* assay can then examine the difference between paused and elongated Pol II after disrupting the nascent RNA using incorporated NTP analogues.

**Approach II: SHAPE Chemical Probing**

**SHAPE: Selective 2’-Hydroxyl Acylation analyzed by Primer Extension**

![SHAPE Process Diagram]

*Figure 12:* A diagram showing the overview of the SHAPE process. Chemical probing is used to generate a graph of nucleotide reactivity for each nucleotide in the nascent RNA using QuSHAPE software. A secondary structure is then predicted with high accuracy using the RNAStructure algorithm.
The second approach to this project will examine the structure of the nascent RNA using SHAPE (Selective 2’-Hydroxyl Acylation analyzed by Primer Extension) prediction chemistry (Weeks, et. al. 2013). The overview of the SHAPE process is shown in Figure 12 and a complete diagram detailing each step of the SHAPE process is shown in Figure 13. The SHAPE method uses a specific SHAPE reagent that chemically reacts with RNA that is structurally flexible. After probing for flexible nucleotides, the RNA is reverse transcribed. The reverse transcriptase enzyme cannot proceed past the bulky adduct added in SHAPE probing and will create a series of short DNA segments with information about the location of the added SHAPE adducts. These short DNA fragments are then analyzed using Sanger sequencing to obtain a plot of fluorescent intensity versus elution time. This graphical result is then directly analyzed using the software program QuSHAPE (Weeks, et. al. 2013). This software program aligns the sequenced short DNA fragments with the known RNA sequence to create a SHAPE reactivity plot. The next step is then to predict the secondary structure of the nascent RNA.
using the software program RNAStructure, which takes the SHAPE reactivity plot as its input (Reuter, J. & Mathews, D, 2010). RNAStructure utilizes thermodynamic parameters to predict RNA secondary structure based on the lowest free energy conformation. RNAStructure uses set thermodynamic parameters for the $\Delta G$ values of RNA folding interactions. A favorable base-stacking or hydrogen bonding interaction will yield a negative $\Delta G$ contribution, while the decrease in entropy due to restriction of possible states will contribute a positive $\Delta G$ contribution because this interaction in non-favorable.

This algorithm has been proven to be greater than 95% accurate for RNA strands of this size by correctly predicting the conformations of folded RNA with known structures (Low and Weeks, 2010). SHAPE chemistry will very accurately allow me to determine the secondary structure of the nascent HSP70 RNA. This well-tested and reliable technique could also be used in future experiments to determine the secondary structure of other genes known to be regulated by pausing in hopes of discovering a conserved pausing motif.

**Detailed Methods**

*In Vitro* Transcription Assay

*HSP70* Template Mix is created by combining 1.6µL of 1M HEPES (pH 7.6), 1.6µL 0.05M DTT, 6.4µL 100mM MgCl₂, 0.5µL Rnase OUT, 65.3µL filtered HPLC water, and 4.5µL of 100nM *HSP70* Template. Reactions are then assembled by combining 8µL *HSP70* Template Mix with 7.8µL DB100 buffering system, and 3µL of a 5:3:2 ratio of the combined P-Cell fractions P1M:P0.5M:P0.3M, respectively. A series of Low C Analogue NTP Mixes were created for each
unique RNA NTP analogue and substituted the native NTP with the analogue so that the total concentration of the NTP and its NTP analogue together was always 400µM final concentration. NTP Analogue mixes for tCTP and Thio-G both tested a range of final concentrations of NTP Analogue (tCTP or Thio-G) of 0µM, 25µM, 50µM, 100µM, 200µM, 300µM, and 400µM. These reactions were set up at 4ºC and then incubated at 30ºC for 30 minutes. After incubation, 1µL of “Low C Analogue NTP Mix”, and 1µL 32P-labeled CTP were combined and added to each reaction. These reactions were then incubated for a 30 minutes at 30ºC. To stop the transcription reactions, 150µL of Stop Solution containing 148µL Stop Buffer (2mL 0.5M EDTA, 2mL 5M NaCl and 2.5mL 20% SDS), 1µL of Glycogen, and 1µL Proteinase-K. RNA was then extracted using a phenol:chloroform extraction and then ethanol precipitated overnight at -20ºC. The samples were centrifuged at 13,000 RPM for 30 minutes, the supernatant removed, and the RNA pellets were resuspended in Formamide Loading Buffer (FLB). These reactions were then ran on a 20cm X 40cm sequencing gel and the RNA was visualized by exposing the gel on a Phosphor Image screen for 18 hours to 3 days and imaging on the Typhoon imager.
Preparing RNA for SHAPE

To prepare RNA for chemical probing via SHAPE, an *HSP70* native promoter was chosen and PCR was used to amplify the chosen DNA template. This amplified DNA product was then transcribed using T7 Polymerase, a bacterial RNA polymerase, in the presence of NTPs to convert the DNA to the nascent *HSP70* RNA. Because the nascent RNA is predicted in its equilibrium structure apart from the polymerase using SHAPE, it was not necessary that we use human Pol II for the transcription step. T7 polymerase was chosen for its ability to produce a very high yield of RNA product that would be visible by UV shadowing and further purification steps that could decrease

![Diagram](image)

**Figure 14:** A diagram showing the necessary steps to convert DNA to RNA and isolate purified RNA for SHAPE chemical probing showing the DNA and RNA I isolated and used in this experiment.
the amount of RNA available for SHAPE. The transcribed RNA was then purified using UV shadowing and gel extraction and imaged using Polyacrylamide gel electrophoresis and Sybr Gold staining. The overall process is shown in **Figure 14** with images of the gels showing the DNA and RNA I isolated and used in this SHAPE experimentation.

After generating a DNA template using PCR, RNA transcripts were generated using a T7 Transcription protocol adapted from the Sauer Lab. For each RNA generated, 10µL of Transcription buffer (250µL 1M Tris-HCl pH 7.5, 75µL 1M MgCl₂, 25µL 1M DTT, and 10µL 1M Spermidine), 10µL 20mM each NTPs, 10µL DNA Template generated via PCR, 2µL T7 Polymerase, and 68µL dI Water. Each reaction was then incubated overnight at 37°C and stopped by adding 5µL 500mM EDTA. The RNA reactions were then phenol:chloroform extracted and ethanol precipitated overnight at -20°C. The RNA reactions were then resuspended in Formamide Loading Buffer and ran on a 20cm X 40 cm sequencing gel. The RNA was imaged using UV-shadowing and the proper length bands were excised and gel-extracted using an RNA gel elution buffer (20mM Tris-HCl pH 7.5, 0.25M Sodium acetate, 1mM EDTA pH 8.0 and 0.25% SDS), following a protocol adapted from Cold Spring Harbor. After gel extraction, the generated RNA transcripts were purified from contaminating DNA using a phenol:chloroform extraction followed by a chloroform extraction and overnight ethanol precipitation at -20°C. The extracted RNA pellets were then resuspended in a well-tested human system folding buffer (50mM KCl, 1mM MgCl₂, and 50mM HEPES buffer). The RNA concentration was then quantified using A₂₆₀ data from a NanoDrop 200 UV-Vis Spectrophotometer. The calculated RNA concentration was 1046.7 ng/µL, which was of sufficiently high concentration to dilute to 1µM for RNA SHAPE reactions.
SHAPE Chemical Probing

Purified RNA was diluted to 1μM and was snap-cooled to ensure one properly folded conformation population by heating the RNA for 3 minutes at 95°C and then cooling for 10 minutes on ice. An RNA Fold Mix was then created by mixing 4μL 0.5X TE buffer, 1μL 1μM RNA, and 3μL 3.33X Fold Mix (333mM HEPES buffer pH 8.0, 20mM MgCl₂, 333mM NaCl, and 17.7mL water) per reaction. The RNA was then NMIA-labeled by mixing 8μL of RNA Fold Mix with 1μL 0.5x TE buffer and 1μL DMSO for the sequencing and control reactions or NMIA for the positive NMIA-labeled reaction (Step 1 of the below SHAPE figure). These reactions were then incubated for 5 half-lives of NMIA at 25°C. After incubation, 3μL of fluorescently labeled DNA primer oligomers were added to each reaction. The RNA was then melted at 65°C for 5 minutes and the oligos were then annealed at 35°C for 10 minutes. An Enzyme Mix was created by mixing 3.3μL water, 2μL 3x Enzyme Mix (500mM Tris buffer pH 8.3, 750mM KCl, 5mL water), 1μL 10mM dNTPs, 0.1μL 1M DTT, 0.6μL SuperScript III Reverse Transcriptase Enzyme per reaction. For the sequencing reactions, ddCTP was substituted for the water in the Enzyme Mix. To each reaction, 7μL of the Enzyme Mix was added to the 13μL NMIA-labeled RNA reactions and Reverse Transcription Extension was carried out at 62°C for 12 minutes. The reaction was then de-activated by adding 1μL of 4M NaOH and heat-deactivating at 95°C for 10 minutes. The reactions were then ethanol-precipitated overnight and washed twice with 75% Ethanol before isolating and drying the pellet and resuspending in HiDi Formamide. 5μL of the Sequencing reaction was combined with each of the DMSO control and NMIA positive reactions and these 10μL reactions were submitted to GeneWiz for Fragment Analysis.
SHAPE Data Analysis

To analyze the fluorescence data obtained from Sanger Sequencing via GeneWiz, two programs were used to generate a predicted RNA secondary structure (Steps 2 and 3 of Figure 12). The first program, QuSHAPE, was downloaded from the Weeks Laboratory at University of North Carolina. The QuSHAPE software took an input of the fluorescence versus elution time data obtained from GeneWiz and generated a SHAPE reactivity plot using a curve integrator method. The second program, RNAstructure, created by the Mathews lab at the University of Rochester Medical Center, is used to generate a predicted secondary structure from the SHAPE structural data calculated using QuSHAPE. For both programs, the default settings were applied and the known RNA sequence was lined up by hand with the SHAPE reactivity curves to generate reactivity data by nucleotide.

Hypotheses

1. *If the nascent RNA secondary structure is disrupted, there will be a change in pausing and/or pause release when tested using an in vitro transcription approach.*

2. *If RNA Pol II Pausing is regulated by RNA Secondary Structure, I will be able to predict a conserved structural pausing motif using SHAPE chemical probing.*
Results

**In Vitro Transcription Assay**

RNA transcribed *in vitro* with an increasing concentration of tCTP analogue was tested using the described *in vitro* assay approach and radioactive imaging. As explained by my hypothesis, when tCTP analogue is titrated in place of CTP, I would expect to see a change in pausing versus runoff transcript when visualized on the Poly-acrylamide gel (Figure 15). However, the current approach using this analogue in the *in vitro* assay failed to reconstitute both paused (short-length product) and runoff product (216nt product). Interestingly, when tCTP concentration increased, the runoff RNA transcript product decreased. Importantly, however, when only tCTP was used with no native CTP, no runoff product was formed. Therefore, we can conclude that human RNA Pol II is not able to effectively use tCTP as a substitute for CTP in transcribing RNA since without CTP, no runoff RNA transcripts are formed. While I was not able to reconstitute a pause-length RNA product, the

![Figure 15: A 6% Polyacrylamide gel imaged using 32P-labeled UTP to visualize transcribed RNA using an increasing amount of tCTP in exchange for CTP. No runoff product in the 0μM CTP lanes indicates that tCTP cannot be effectively utilized by human Pol II.](image-url)
disappearance of RNA runoff product at 0μM allows us to conclude that RNA Pol II cannot effectively use tCTP as a catalytic substrate. Therefore, a new analogue needed to be chosen to examine its ability to be taken up by RNA Pol II and effectively disrupt RNA secondary structure to compare pausing versus runoff product.

The next analogue examined was Thio-G, created by the Tor Lab at UCSD (McCoy, Shin, and Tor 2014). Since Thio-G is modified with a (3,4)thienopyrimidine core, this more subtle modification was hopeful that Thio-G would be more likely to be used effectively by human Pol II. As my hypothesis above states, if secondary structure is disrupted by Thio-G, I would expect to see a change in RNA paused product versus runoff product.

Again, RNA was transcribed using the *in vitro* transcription assay approach, and increased Thio-G analogue was used in place of native GTP. Radioactive CTP was used to visualize RNA transcripts generated and compare runoff versus paused RNA products. Again, runoff product seemed to decrease generally with the addition of a greater concentration of Thio-G analogue. However, the current approach *in vitro* assay was unable to produce any paused products to compare to the

![Figure 16: A 6% Polyacrylamide gel imaged using ^32^P-labeled CTP to visualize transcribed RNA using an increasing amount of thio-G in exchange for GTP. Again, the appearance of no runoff product in the 0μM CTP lanes indicate that Thio-G cannot be incorporated by human Pol II.](image)
decreased runoff product. Even with the new Thio-G analogue, when only Thio-G is used in place of GTP, no runoff product is formed (Figure 16). Because no RNA transcripts can be formed without native GTP, it is conclusive that human RNA Pol II also cannot effectively utilize Thio-G in RNA transcription.

While I was not able to visualize a paused RNA product, I was able to conclude from the disappearance of runoff RNA transcript in the presence of only tCTP or Thio-G that human RNA Pol II cannot effectively incorporate these NTP analogues in place of CTP or GTP, respectively. This conclusion alone is valuable information for future groups and use of these analogues since, while they have shown to be effective for transcription in bacterial systems using T7 polymerase, they cannot be used to study human systems in which human Pol II is necessary for transcription. To continue to study the role of RNA secondary structure on human Pol II pausing, new NTP analogues will need to be chosen and investigated for their ability to be used effectively by human Pol II when synthesizing RNA.

**SHAPE Structural Prediction**

After examining the effect of RNA analogues on RNA Pol II Pausing, the next goal was to predict the secondary structure of the nascent *HSP70* RNA using SHAPE chemical probing and computer prediction algorithms. However, after two rounds of SHAPE experimentation, I was unable to obtain conclusive data from the SHAPE analysis. Due to two critical errors from a key collaborator, the reverse transcription reaction after NMIA-labeling was unsuccessful. First, the RNA was not designed with a specific Reverse Transcriptase (RT) primer binding site. During reverse transcription, the SuperScript III Reverse Transcriptase enzyme needs a DNA primer (a short
single-stranded DNA molecule) that binds to the RNA to provide a starting point for the reverse transcriptase to begin copying the RNA into the reverse-complement DNA. Since the RNA was not engineered with a specific sequence for this Reverse Transcriptase primer to bind, there was no primer to begin reverse transcription, and the reaction could not occur. Therefore, no short DNA fragments were produced from the reverse transcription reaction and no structural information was able to be obtained from the SHAPE computer analysis. Second, after incorporating a specific RT primer binding site into the HSP70 RNA, I was told that I was given the incorrect RT primer binding site sequence by the collaborator. Because the primer is a specifically designed sequence and requires its reverse complement to be present to bind with high affinity, I was not able to get any primer bound to the RNA for reverse transcription. Again, no reverse transcription reaction was able to occur without effective binding of the RT primer and no structural information could be obtained from the SHAPE chemical probing analysis. Given more time, this process needs to be repeated with correct incorporation of the proper 5’-3’ RT primer binding sequence into the nascent RNA so that effective RT primer binding will occur and reverse transcription will yield accurate structural information about the folded nascent HSP70 RNA. After troubleshooting this process, I was able to identify the issue that was preventing meaningful data from being obtained and by generating new nascent RNA with the proper RT primer binding site, SHAPE experiments could be repeated with likelihood of a successful prediction of the secondary structure of the nascent HSP70 RNA.
Conclusion

The success of this project was largely limited by the time it took to optimize the biochemical techniques used. While the human transcription assay is a powerful biochemical tool, it is a difficult technique to master and yields small amounts of RNA. The buffering conditions and ratios of the P-Cell fractions dramatically changed the yield and quality of the transcription products visualized. However, even with months of optimization, a paused RNA product could not be visualized using this approach. While the change in the runoff product still produces meaningful and interesting results, a comparison of the ratio of paused to runoff transcript is necessary to validate the potential role of secondary structure in RNA Pol II pausing and obtain quantitative data. Overall, two nucleotide analogues were tested and while both were shown in the literature to be taken up by the bacterial T7 Polymerase, neither analogue was successfully able to be taken up by human Pol II. This finding confirms important structural differences between the human and bacterial polymerase active sites since even minor changes in the NTP structure can prevent incorporation by the human Pol II enzyme. In the future, new NTP analogues, such as inosine, will need to be chosen carefully for their potential to be taken up by human Pol II to allow testing of the role of RNA secondary structure on Pol II pausing using an in vitro assay approach. Secondly, by optimizing the RNA construct for SHAPE chemical probing to allow a proper Reverse Transcriptase primer to bind the RNA template, the structure of the nascent HSP70 RNA can be predicted using the SHAPE technique described. This technique will also be useful for future structural prediction of other paused genes to possibly identify a secondary structure pausing motif conserved across paused genes.
Bibliography


