TITLE: Novel insight into the patterns of in vitro short abortive RNA release by Escherichia coli RNA polymerase

AUTHOR'S NAME: Artun K. Kadaster

MAJOR: Biochemistry

DEPARTMENT: UW Department of Biochemistry, CALS

MENTOR: Dr. M. Thomas Record, Jr.

DEPARTMENT: UW Department of Biochemistry; UW Department of Chemistry

MENTOR 2: Claire E. Evensen

DEPARTMENT: UW Department of Biochemistry, L&S

YEAR: 2020

The author hereby grants to University of Wisconsin-Madison the permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.
Novel insight into the patterns of *in vitro* short abortive RNA release by *Escherichia coli* RNA polymerase

Artun K. Kadaster

Senior Honors Thesis

Record Laboratory
Department of Biochemistry
Department of Chemistry
University of Wisconsin – Madison
Madison, Wisconsin
Abstract

*E. coli* RNA polymerase (RNAP) catalyzes prokaryotic transcription with the σ^{70} specificity factor, which directs the binding of RNAP to promoter DNA. RNAP operates on the promoter to form an open complex. In transcription initiation, complementary ribonucleotide triphosphates (NTP) bind stepwise to the template DNA in the active site and RNAP catalyzes phosphodiester bond formation. After each NTP addition, the RNA-DNA hybrid translocates into the active site cleft while maintaining promoter contacts. In productive complexes, free energy buildup from translocation allows RNAP to break its specific contacts and escape from the promoter to transcribe full-length RNA. Nonproductive complexes stall before escape, release a short RNA, and re-initiate a cycle of abortive (short RNA) transcription. Dinucleotide RNA is the shortest possible abortive product. Experimental evidence and a thermodynamic rationale are presented indicating that most of the RNA synthesized by nonproductive complexes *in vitro* for the conditions investigated is 3-mer or longer.
**Background**

The transcription of RNA from DNA, catalyzed by the RNA polymerase (RNAP) enzyme, is a fundamental step in the process of gene expression. Transcription occurs in three phases: initiation, elongation, and termination. In prokaryotes, initiation begins when RNAP binds to the DNA promoter with the aid of the σ specificity factor (complexed RNAP core enzyme and σ are referred to as a holoenzyme). Seven σ–factors are present in *E. coli* which are responsible for regulating different types of genes\(^1\). Genes which are transcribed by σ\(^{70}\) holoenzyme are responsible for growth and maintenance of the cell.

Initial *E. coli* σ\(^{70}\) holoenzyme-promoter binding interactions generate a closed complex in which the holoenzyme is bound to linear DNA (Figure 1). RNAP then induces a series of conformational changes, bending and wrapping the duplex DNA around its active-site cleft. Next, RNAP unwinds the DNA and forms a single stranded transcription bubble, positioning the DNA transcription start site (TSS) adjacent to the Mg active site cofactor. After transitioning through several intermediates, a transcriptionally active RNA-DNA open complex (OC) is formed. The strength of these interactions is sequence dependent. *E. coli* RNAP OC with the bacteriophage λPr promoter has a lifetime (1/K\(_d\)) of 11 hours at 37 °C\(^1\).

When the OC is in the presence of ribonucleotide triphosphates (NTP), the nucleotide addition cycle begins. In this process, NTPs diffuse into the RNAP active site and complementary NTPs
bind to the template-strand DNA bases at positions +1 (the transcription start site) and +2 (one base downstream of the start site in the direction of transcription). RNAP then catalyzes formation of a phosphodiester bond with the preceding NTP, cleaving a pyrophosphate (PP_{i}) molecule as a byproduct. NTP binding, the reaction extending the DNA-RNA hybrid, and generation of PP_{i} are thermodynamically favorable and drive the energetics of the process. RNAP then translocates, positioning the subsequent DNA position adjacent to the active site to begin the cycle again. During transcription initiation, the initial transcribing complex (ITC) translocates by unwinding downstream DNA and pulling it into the RNAP cleft. Meanwhile, σ^{70}'s upstream promoter contacts are maintained, resulting in a DNA “scrunching” phenomenon.

Translocation stress is the product of the resultant scrunching free energy buildup and clashes between the σ^{70} region 3.2 and the 5’ end of the RNA transcript. This strain can allow RNAP to break the upstream contacts sequentially, release the σ^{70} factor, and escape the promoter to transcribe the full-length (productive) RNA product (Figure 2).

Figure 2: Scrunching during initial transcription for the λP_r promoter
Depicted is the process of DNA scrunching as downstream ssDNA is pulled into the enzyme during translocation of the ITC. RNAP (gray), the transcription bubble consisting of both template (blue) and non-template (black) strands, the active site (yellow), the nascent RNA (green squares), and the +1 and +2 DNA positions (red and pink squares) are represented schematically. The leftmost diagram depicts the OC in an unscrunched position. Scrunching proceed as the ITC translocates to each successive position. Below each illustration is the free energy accumulation that occurs at each depicted step along the path to promoter escape. For the λP_r promoter, an accumulation of 9.0 kcal/mol is sufficient for RNAP to break all promoter contacts and transition to elongation, coinciding with the formation of the 11-mer.

The RNA length at which promoter escape occurs depends on many factors, one of which is the strength of the RNAP-promoter interactions. In λP_r, promoter escape occurs after the 10-nucleotide (10-mer) RNA has been formed. Alternatively, the ITC may stall before this requisite free energy buildup is achieved. In this case, the abortive, short RNA product is ejected and the ITC cycles back to the TSS.
non-productive ITC will continue to cyclically reinitiate, stall, and release abortive products. The productive fraction will transcribe their full-length RNA products within 10 seconds\(^3\), whereas the abortive cycling will occur indefinitely. The total population of OCs spontaneously branches into these two distinct fractions of ITCs upon NTP addition\(^4\). Critically, the productive fraction of ITCs will not release short RNA products and the non-productive fraction of ITCs can never produce or release a full-length RNA. While the molecular properties cause this branching phenomenon are unknown, it has been shown that these are not intrinsic to the enzyme as the isolated productive or non-productive fraction of ITCs will spontaneously re-branch when reinitiated\(^4\).

*In vitro* studies elucidating patterns of short abortive RNA formation and release have significant implications on the overall process of transcription initiation and its regulation *in vivo*. Previous work has shown that, during the stationary phase of *E. coli* growth, short RNA oligonucleotides ranging in length from 2-mer to 4-mer can be used as primers to initiate transcription instead of NTPs\(^5\). Under these conditions, the TSS is shifted so that transcription begins at what would normally be the -1 base of the DNA template strand. During log phase growth, these short RNAs cannot be used as a primer. Thus, short abortive RNAs can provide *E. coli* with a mechanism for regulating the transcription start site and subsequent gene expression based on the nutrient availability of the cell.

The dinucleotide product pppApU (2-mer) is the shortest length of abortive RNA product that can be released. Its synthesis is an exception to the nucleotide addition cycle. RNAP does not translocate to add the UTP in the second position. Thus, no translocation stress buildup contributes towards promoter escape occurs during this step. In previous transcription assays, both transient 2-mer on the way to full length RNA and abortive 2-mer buildup have been challenging to observe and quantify using polyacrylamide gel electrophoresis. This is because the 2-mer band is usually not separated from smaller MW bands that represent background noise and free moving \(\alpha\)-\(^{32}\)P radiolabeled NTP\(^3\). Thus, nearly all previous kinetic models have had to simulate the kinetics of 2-mer buildup as a transient intermediate on the way to full-length RNA synthesis. Meanwhile, abortive 2-mer kinetic and thermodynamic parameters
are unknown. Here, it is shown that, in vitro, abortive 2-mer release does not occur under the NTP conditions investigated. The thermodynamic and molecular basis for this claim is provided.

**Methods**

To measure abortive RNA synthesis, transcription assays were performed on the benchtop on a medium-long time scale from 10 to 480 seconds. OC was incubated for 1 hour at 37 °C. Quench solution contained 8 M urea, 0.15 M EDTA, 0.05% (wt/vol) xylene cyanol, and 0.05% (wt/vol) bromophenol blue suspended in Tris/borate/EDTA buffer. Transcription buffer (TB) contained 40 mM Tris base (pH 8.0), 5 mM MgCl₂, 1 mM DTT, 0.1 mg/mL BSA, and 60 mM KCl. The holoenzyme mixture contained 200 nM λPr DNA and 50 nM holoenzyme suspended in TB. Reaction initiation solution contained 0.25 mg/mL heparin, 17.5 nM α-³²P UTP, 10 μM UTP, 200 μM ATP, and 200 μM GTP (not present in NTP subset experiments) suspended in TB. The ratio of α-³²P labeled NTPs to unlabeled NTPs for the limiting nucleotide was 0.00175:1 for quantitation purposes. RNA products were run half-way down 23% polyacrylamide gels, exposed for 18 h on a phosphorimaging cassette, and phosphorimaged using a GE Typhoon FLA 9000 Phosphorimager. Quantification was performed using ImageQuant software.

<table>
<thead>
<tr>
<th>Table 1: Sequences Used for PCR and Transcription Assay ITR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shown are the primer sequences used to generate the DNA promoter with polymerase chain reactions. The initial transcribed region of the DNA promoter is provided with the transcription start site (bold) and CTP stops (underline) annotated.</td>
</tr>
<tr>
<td>λP, WT Forward (-71 to -12) 5’ – CCACGAATTCAATGATAATCTACACCCGT GCGTGTTGAATATTTTTACCTCCTGCGGTG – 3’</td>
</tr>
<tr>
<td>λP, WT Reverse (-24 to +31) 5’ – ACAAAACCTCATGACGACCTCTACAT GCAACCATTATACCCCGCCAGGT – 3’</td>
</tr>
<tr>
<td>HTOP Primer 5’ – CCAGCATTCCCTCCAGAATTC – 3’</td>
</tr>
<tr>
<td>HBOT TXN Primer 5’ – CACCTGCACCGACAAAAACCTTT – 3’</td>
</tr>
<tr>
<td>Initial Transcribed Region 5’ – ATGTAAGGAGGAGTTC – 3’</td>
</tr>
</tbody>
</table>

The λPr promoter was generated with a polymerase chain reaction (PCR) using custom forward and reverse primers and Phusion DNA polymerase (Table 1). Agarose gel purification was performed to isolate the desired DNA products. The promoter construct was 124 bp, spanning from the -82 to the +42
position. The sequence contained a guanosine at positions +17 and +32, promoting product buildup at positions +16 and +31 in the absence of CTP from the reaction mixture.

Results

Since at least 2017, the Record lab’s studies of abortive and productive initiation, used to determine the thermodynamics and mechanism of transcription, have been unable to resolve a quantifiable 2-mer band on polyacrylamide gels. This resulted from an inability to separate this band from the background noise and free α-32P radiolabeled NTPs during polyacrylamide gel electrophoresis (PAGE) using a 20% polyacrylamide gel, as seen in an unpublished gel image courtesy of Lindsey Felth (Figure 3). Thus, the kinetics of 2-mer synthesis were inferred using experimental data of product buildup at other lengths. Previous work in the Record lab was done to find if migrating the products different distances down the gel would separate the 2-mer band more optimally from the free NTP band. In order to be sure that 2-mer was being observed, the reaction conditions were modified to prematurely stall the ITC after the abortive 2-mer had been synthesized. This was done by providing a subset of only UTP and ATP with α-32P radiolabeled UTP for visualization. The dominant product under these reaction conditions is 2-mer, because GTP is absent and misincorporation is therefore required to synthesize 3-mer. Upon a comparison of the phosphorimaged gel with the stopping points of each lane, it was determined that allowing the RNA products to migrate roughly halfway down the gel provided the greatest separation between the 2-mer band and the free α-

Figure 3: Typical 20% acrylamide gel image
A representative gel image is shown displaying product buildup from a transcription assay. Reaction conditions are 10 μM UTP, 200 μM GTP, and 200 μM ATP at 37 °C. The lowest observable band is 3-mer, while 2-mer is lost in the background noise and blends in with the dark band of free radiolabeled NTPs. (Unpublished image by Felth, L.C.)
32P NTP band. Thus, an assay in which the goal was to resolve 2-mer buildup could theoretically be performed provided special considerations were given to the migration distance of the products during electrophoresis.

Although 2-mer buildup could be observed in assays where only ATP and UTP were added, 2-mer synthesis by nonproductive complexes was not detected when GTP was present, even with a shorter gel migration. Abortive 2-mer synthesis had eluded all previous attempts to quantify its buildup. This necessitated the use of a different approach to further separate the abortive 2-mer band and quantify its buildup.

2-mer can be resolved with quantifiable separation via high percentage PAGE

An NTP subset transcription assay was performed using 10 µM UTP and 200 µM ATP to stall synthesis after production of the 2-mer. The products of this experiment were run on a 23% polyacrylamide gel which was phosphorimaged (Figure 4). The results showed very good separation between the dinucleotide band and the free NTP band.

![Figure 4: 2-mer band resolved on 23% acrylamide gel](image)

A gel image is shown displaying product buildup from a transcription assay. Product buildup is displayed at 10, 30, 90, 150, 240, and 480 s. Reaction conditions are 10 µM UTP and 200 µM ATP at 37 °C. The 2-mer band (red arrow) is well resolved and clearly separated from the free radiolabeled NTP and background noise. This band is not visible in the negative control lane.
**2-mer is not observed given 3-NTP conditions**

Four transcription assays, two with an NTP subset of 10 µM UTP and 200 µM ATP and two with a full complement of 10 µM UTP, 200 µM ATP, and 200 µM GTP, were performed and resolved on the same 23% polyacrylamide gel (Figure 5). Unlike previous experiments, this allowed for a direct, side by side comparison and straightforward identification of where the 2-mer band should occur. Both assays where only ATP and UTP were present showed clear 2-mer kinetic buildup across the 480 second time course, with no correlated band in the negative control lane. Extrapolation of the steady-state region to the y-intercept yielded 0.01 (± 0.01) initial 2-mer RNA per OC. The abortive synthesis rate was 0.7 (± 0.0) 2-mer RNA per OC per 1000 s (Figure 6). The trials where GTP was also present showed no bands at the same length.

![Figure 5: NTP subset and 3-NTP comparison experiment on a 23% acrylamide gel](image)

A gel image is shown displaying product buildup from several benchtop transcription assays. All assays display product buildup at 10, 30, 90, 150, 240, and 480 s. The two left most trials are the NTP subset experiment, whereas the two rightmost trials are the 3-NTP assay. Reaction conditions for the NTP subset trials are 10 µM UTP and 200 µM ATP at 37 °C. Reaction conditions for the 3-NTP assays are 10 µM UTP, 200 µM GTP, and 200 µM ATP at 37 °C. Kinetic buildup of 2-mer is observed in assays with only ATP and UTP (red arrow), while the corresponding position when GTP was also present shows no significant 2-mer band (blue arrow). No bands are observed in negative control lanes.
Discussion

The issue of resolving 2-mer on polyacrylamide gels is one previously met by many researchers studying transcription initiation. Previous publications have marked the 2-mer band on gel images when any sort of 2-mer band in the images is not separable from the free NTP band in the manner reported here\(^8\). This suggests that certain studies have assumed the 2-mer band to be present and simply hidden from view under similar 3-NTP conditions to those investigated in this work. It is therefore important to acknowledge, both through energetic analysis and predicted product distributions of abortive ITCs, why abortive 2-mer release is not observed under the conditions investigated.

The lack of 2-mer at the first time point (10 s) in the 3-NTP transcription assays means that it is not a detectable transient intermediate in full length RNA synthesis by productive complexes\(^3\). This is due to the speed in which transient 2-mer buildup and conversion to 3-mer occurs. With a simulated second order rate constant of $4.4 \pm 0.6 \times 10^5 \text{ M}^{-1} \text{s}^{-1}$, 2-mer buildup peaks and decays before 3 seconds have elapsed since transcription initiation\(^3\). This can be compared to the $1.5 \pm 0.7 \times 10^5 \text{ M}^{-1} \text{s}^{-1}$ and $1.2 \pm 0.1$...
x $10^5$ M$^{-1}$ s$^{-1}$ rate constants of 3-mer and 4-mer, respectively. Thus, 2-mer synthesis is nearly 3 times faster than the subsequent NTP incorporations. These rate constants apply to transient species in the productive pathway and, while they probably are a reasonable estimate of the rate of synthesis of the first 2-mer made by a nonproductive complex, they are only an upper bound on the rate of abortive synthesis of 2-mer because abortive synthesis rates in general are determined by the rate of release of the abortive product and restart.

Previous work by Kate Henderson et al. has examined RNA product length distributions released by various promoters in both the fast (< 30 s) and abortive cycling (> 30 s) phases of transcription$^6$. Extrapolating the steady-state region of transcription to its intercept shows that approximately one RNA per OC is produced in the fast, initial phase of transcription. Performing the same analysis by individual transcript length for the λPr promoter shows that RNAP transcribes 0.25 (±0.01) 3-mer, 0.04 (±0.02) 4-mer, 0.01 (±0.01) 5-mer, and 0.43 (±0.02) full-length initial RNAs per OC in the fast phase. While initial 2-mer RNA buildup could not be directly measured, an estimate can be given by subtracting the measured, initial short and full-length RNA amounts from one, the total amount of RNA produced per OC in the fast phase. This estimate yields an amount of roughly 0.2 2-mer initial RNA per OC. A comparison with product buildup in the slower phase of transcription shows that most of the abortive cycling is done in this longer timeframe. RNAP transcribes 4.2 (±1.4) 3-mer, 1.6 (±0.31) 4-mer, and 0.59 (±0.16) 5-mer RNA per OC per 1000s in the slower phase$^6$. Given the similar magnitude of initial 2-mer (estimated) and 3-mer RNAs produced per OC, as well as the very faint 3-mer band at the 10 s timepoint of a transcription assay, it is understandable that little 2-mer is seen at 10 s. The gel results reported here demonstrate that, under conditions with high GTP, the abortive rate of 2-mer synthesis must be lower than the abortive rates of 3-mer and 4-mer synthesis as comparatively little-to-no 2-mer transcript buildup occurs in the slower phase. The quantifications of the assays in which GTP is withheld yield an initial fast-phase buildup of 0.01 (± 0.01) 2-mer RNA per OC and an abortive synthesis rate of 0.7 (± 0.0) 2-mer RNA per OC per 1000 s. Error values are notably high due to a lack of replication. This data serves as an
upper limit for an abortive 2-mer synthesis rate because it was generated under conditions in which GTP was withheld. This supports that the abortive 2-mer rate is far lower than the abortive 3-mer synthesis rate under conditions in which 200 μM GTP is provided.

On the path to promoter escape, the ITC passes through several transient energetic barriers which promote free energy buildup and sequential detachment from the promoter. In the first of these, the synthesis of a 5-mer from a 2-mer, translocation stress causes the discriminator region of the promoter to dissociate from σ70 region 4.2 and the core recognition element. This series of steps generates roughly 33% of the free energy buildup required to escape the promoter, none of which occurs during the binding or catalysis steps which form ITC2 (ITC with 2-mer RNA) because the first translocation step at the λPr promoter occurs after 2-mer synthesis. The lack of abortive 2-mer release under the observed conditions suggests that an abortive pausing event may be caused by a defect in the preceding translocation step. The lack of a preceding translocation step could alternatively affect the kinetics of abortive release by altering the rate of pyrophosphorolysis. Pyrophosphorolysis involves the reverse reaction of NTP catalysis in which a PPi molecule is assimilated to break the phosphodiester bond and release an NTP. This reverse rate would be much higher in the 2-mer rather than later NTP addition steps, as 2-mer would not need to first proceed through a translocation step to release a nucleotide. In this scenario, it may be much faster for an ITC2 faced with a translocation defect in the forward direction to perform pyrophosphorolysis and eject the UTP.

The final explanation of this phenomenon lies in the NTP concentrations present in the reaction mixture. Manipulating the concentration present of the subsequent nucleotide at any step can lead to pausing and RNA release. This is what allows for product buildup at the DNA positions 1 NTP upstream of a CTP-stop in the experimentally used λPr sequence. Likewise, the withholding of GTP in the NTP subset experiments creates a sufficient energetic barrier after ITC2 formation to stall all complexes. A 200 μM GTP concentration is evidently high enough to not present a sufficient barrier for the incorporation of a GTP to ITC2.
The findings in this work have given new insights into patterns of short abortive RNA synthesis. Future experiments will be conducted to determine the exact bounds by which the reaction conditions can be manipulated to increase abortive 2-mer synthesis rates. This will involve testing a concentration gradient of PPi in solution. Preliminary experimentation has shown that increased PPi concentrations in the reaction mixture may induce a translocation barrier. Under these circumstances, 1 mM and 2 mM concentrations of PPi may lead to significant pausing at the 2-mer step and abortive product release. This can be tested by performing transcription assays with 3 NTPs present and a concentration range of 0-2 mM PPi. More replicates of the data reported in this work will be conducted to reduce error for quantified values. Lastly, the 23% acrylamide gel separation technique will be used to re-test previously published reaction conditions in an attempt to elucidate further trends of abortive 2-mer synthesis.

Acknowledgements

The author would like to thank Claire Evensen, Ben Palatnik, Dylan Plaskon, Dr. Kate Henderson, Lindsey Felth, and Dr. M. Thomas Record, Jr. for their contributions to this work.
References

   Biomolecules, 5. 1035-1062. doi:10.3390/biom5021035

   transcription initiation intermediate of Escherichia coli RNA polymerase. PNAS, 113(43). E6562-E6571. 
   doi:10.1073/pnas.1605038113

   58(18). 2339-2352. doi: 10.1021/acs.biochem.9b00049

   RNA polymerase. JMB, 256(3). 449-457. doi.org/10.1006/jmbi.1996.0100


   Mechanism of transcription initiation and promoter escape by E. coli RNA polymerase. PNAS, 114(15). 
   E3032-E3040. doi:10.1073/pnas.1618675114

   productive and non-productive pathways during initial transcription in bacteria. Nature Communications, 
   9(1). doi:10.1038/s41467-018-03902-9

   Escherichia coli RNA polymerase. 2. Formation and characterization of two distinct classes of initial 
   transcribing complexes. Biochemistry, 42(13). 3787-3797. doi:10.1021/bi0269613