Title
The mechanism of recognition and processing of DNA damage and modifications by RNA polymerase II

Permalink
https://escholarship.org/uc/item/9mn2j734

Author
Shin, JI

Publication Date
2016

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

The mechanism of recognition and processing of DNA damage and modifications by RNA polymerase II

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Ji Hyun Shin

Committee in Charge:

Professor Dong Wang, Chair
Professor James Kadonaga, Co-Chair
Professor Shannon Lauberth

2016
The Thesis of Ji Hyun Shin is approved and it is acceptable in quality and form for publication and on microfilm and electronically:

Chair

University of California, San Diego

2016
TABLE OF CONTENTS

Signature Page ................................................................. iii
Table of Contents ............................................................. iv
List of Figures ................................................................. v
List of Tables ................................................................. vii
Acknowledgements ........................................................... viii
Vita .................................................................................. ix
Abstract of the Thesis ......................................................... x
Introduction ........................................................................ 1
Project Design and Rationale ............................................... 8

Chapter 1: The recognition of unnatural nucleoside triphosphates
by RNA Pol II................................................................. 10

Chapter 2: RNA Pol II transcriptional arrest and bypass of
regioisomeric ethylated thymidine lesions ................................. 30

Chapter 3: RNA Pol II transcriptional arrest by non-covalent
minor groove binders: Py-Im polyamides ................................. 40

Summary & Future Implications ............................................. 49

References .......................................................................... 53
LIST OF FIGURES

**Figure i.** Structure of RNA polymerase II elongation complex.......................2

**Figure ii.** RNA pol II transcription elongation ...........................................3

**Figure iii.** Structural overlay of RNA pol II elongation complexes that accommodates CPD or CydA lesion.................................................................4

**Figure iv.** The structure of RNA pol II elongation complex with 5caC..............6

**Figure 1.1:** Unnatural base pair.................................................................12

**Figure 1.2:** Single nucleotide incorporation into dTPT3 template ..................16

**Figure 1.3:** Quantitative analysis of single nucleotide incorporation into dTPT3 template...........................................................................................................18

**Figure 1.4:** Quantitative analysis of single nucleotide incorporation into dNaM template ........................................................................................................18

**Figure 1.5:** Relative K values for each enzyme shows template dependent nucleotide selectivity.................................................................19

**Figure 1.6:** Single nucleotide incorporation of rNaM opposite different templates........................................................................................................20

**Figure 1.7:** Single nucleotide incorporation of rTPT3 opposite different templates........................................................................................................21

**Figure 1.8** NTP chase experiments.................................................................23

**Figure 1.9:** Elongation of the UBP by pol II..................................................25

**Figure 1.10:** Elongation of the UBP by E. coli RNAP......................................27

**Figure 1.11:** Elongation of the UBP by T7 RNAP.........................................28

**Figure 2.1:** RNA pol II transcription elongation of the damaged templates....33

**Figure 2.2:** Single nucleotide incorporation opposite the damage sites.........36

**Figure 2.3:** Extension from the damaged site is lesion specific....................37

**Figure 3.1:** Py-lm polyamide structure.........................................................42
Figure 3.2: Py-Im polyamides block RNA pol II transcription elongation……… 44

Figure 3.3: Molecular modeling of RNA pol II transcriptional pausing and arrest by Py-Im polyamide………………………………………………………46
LIST OF TABLES

Table 1.1: Single nucleotide incorporation efficiency by each enzyme ............ 17

Table 1.2: Single nucleotide incorporation efficiency opposite different templates ......................................................... 22
ACKNOWLEDGEMENTS

I would like to acknowledge Professor Dong Wang for his support as the chair of my committee. I would also like to acknowledge Dr. Liang Xu for his direct research guidance.

The Introduction, in part, contains material that has been submitted for publication of the material as it may appear in Cell & Bioscience, 2016, Shin, Ji H.; Xu, Liang; Wang, Dong, Mechanism of Transcription-Coupled DNA Modification Recognition. The thesis author was the primary author of this paper.

Chapter 3, in part, is a reprint of the material as it appears in Proceedings of the National Academy of Sciences of the United States of America, 2016, Xu, Liang; Wang, Wei; Gotte, Deanna; Yang, Fei; Hare, Alissa A.; Welch, Timothy R.; Li, Benjamin C.; Shin, Ji H.; Chong, Jenny; Strathern, Jeffrey; Dervan, Peter B.; Wang, Dong, RNA polymerase II senses obstruction in the DNA minor groove via a conserved sensor motif. The thesis author was a co-author of this paper.
VITA

2015 Bachelor of Science, University of California, San Diego

2015-2016 Teaching Assistant, Department of Biological Sciences, University of California, San Diego

2016 Master of Science, University of California, San Diego

PUBLICATIONS


ABSTRACT OF THE THESIS

The mechanism of recognition and processing of DNA damage and modifications by RNA polymerase II

by

Ji Hyun Shin

Master of Science in Biology

University of California, San Diego, 2016

Professor Dong Wang, Chair
Professor James Kadonaga, Co-Chair

RNA polymerase II (pol II) recognizes many obstacles during transcription elongation, including DNA damage lesions and modifications, via specific interactions and leads to distinct transcriptional outcomes. We investigate three specific types of modifications/lesions in DNA and how they affect the pol II transcription process:
1) unnatural synthetic nucleotides (dNaM and dTPT3), 2) regioisomeric alkylated thymidine lesions (O²-, N³-, O⁴-EtdT), and 3) non-covalent minor groove DNA binders pyrrole-imidazole (Py-Im) polyamides. In Chapter 1, we investigate pol II transcription and elongation in the presence of synthetic nucleotides (dNaM and dTPT3), and the ability of pol II to distinguish between natural NTPs and the unnatural triphosphates. Selective incorporation of rNaM by pol II only occurs when dTPT3 is in the template strand, and loses its selectivity when dNaM is in the template. In Chapter 2, we discovered distinct patterns of pol II transcriptional bypass for each of the alkylated thymidine lesions. We found that pol II bypass of O²-EtdT is essentially error free, bypass of O⁴-EtdT is efficient and highly error prone, and bypass of N³-EtdT is extremely slow. In Chapter 3, we found that Py-Im polyamides bound to the minor groove at sequence specific sites causes prolonged pol II arrest upstream of the binding site, due to two specific residues in the pol II Switch 1 region that contribute to the early detection of the obstruction in the minor groove. Taken together, these studies highlight the importance of pol II recognition of DNA damage and modifications in the maintenance of transcriptional fidelity.
INTRODUCTION

RNA polymerase II (pol II) is the enzyme responsible for the transcription and synthesis of pre-messenger RNA and noncoding RNA transcripts (1). During the process of transcription, pol II reads along the template strand of genomic DNA and incorporates the matched nucleotide substrate with high fidelity to ensure accurate genetic transfer and minimize transcriptional errors. Transcriptional fidelity during elongation is maintained via at least three fidelity checkpoint steps: the nucleotide insertion step, RNA transcript extension step, and proofreading step (1). Unavoidably, pol II may encounter various DNA modifications or lesions during its long transcriptional ‘journey’ moving along the DNA template. In such situations, pol II utilizes several important motifs to ‘sense’ these DNA modifications. The distinct interactions between pol II conserved motifs and these DNA modifications also induce appropriate transcription-coupled responses, which may lead to transcriptional mutagenesis, transcription-coupled repair pathway, or apoptosis (2-4).

There are several important conserved structural components of pol II involved in DNA template base recognition and fidelity control, including the trigger loop and bridge helix of the Rbp1 subunit (Figure i). The trigger loop (TL) is a highly conserved domain in various multisubunit RNA polymerases that is responsible for the rapid catalysis of phosphodiester bond formation and maintaining substrate specificity (1, 5, 6). In the presence of a matched NTP substrate, complementary to the DNA template in the active site, the TL undergoes a conformational change from open, inactive states to a closed, active state and positions the substrate for catalysis. The bridge helix is a long alpha helix domain that bridges over the two halves of pol II and separates the pol II catalytic
site from the downstream main channel and the secondary channel (5, 7, 8). All of these components are important for pol II enzymatic activity, but they also contribute to the ability of pol II to sense DNA modifications and damage during transcription elongation.

**Figure i. Structure of RNA polymerase II elongation complex.** The incoming NTP enters the pol II active site through the secondary channel of pol II (dashed circle). The bridge helix (BH) is shown in green, while the RNA, template DNA (TS), and non-template DNA (NTS) are shown in red, blue, and cyan, respectively.

Genomic DNA is under constant attack, including endogenous reactive oxygen species and free radicals, and external factors like UV irradiation. As a result, these attacks cause many DNA lesions, including base modifications, strand breaks, crosslinks, and bulky, DNA-distorting lesions. Pol II may encounter these lesions or modifications during RNA transcript synthesis (Figure ii). A critical question in the field is how Pol II recognizes and senses these DNA modifications or lesions. Recent studies provided new insights into understanding this critical question.
Figure ii. RNA pol II transcription elongation. (a) Elongation of RNA polymerase II may encounter different types of DNA modifications. (b) These include bulky, DNA-distorting lesions (e.g. UV damage CPD, oxidative damage CydA), small but mutagenic DNA damage (e.g. 8-oxo-guanine), and enzyme-catalyzed endogenous DNA modifications (e.g. 5caC).

Bulky DNA-distorting lesions serve as a strong road block for pol II elongation (9). UV-induced cyclobutane pyrimidine dimer (CPD) lesions form 1,2-intrastrand cross-links that significantly distort the DNA template structure. These lesions strongly inhibit pol II transcription by reducing the rate and fidelity of substrate incorporation and extension (10, 11). Intriguingly, a structurally unrelated bulky DNA lesion, cyclopurines (CydA), which arise form oxidative damage, also strongly inhibit pol II transcription elongation in the similar manner (12, 13). In both cases of transcriptional stalling, pol II utilizes the A rule, a phenomenon in which AMP is preferentially incorporated in a non-template dependent manner, rather than the correct substrate, opposite a damaged DNA base in a non-template manner (13), indicating that pol II may recognize these...
structurally different DNA lesions in a similar manner. Intriguingly, further structural analysis indeed revealed that both lesions are accommodated above the bridge helix (Figure iii) and arrested in a similar position in which the damaged base is stuck at the half-way position of template translocation between the i+1 and the i+2 position (11, 13). Interestingly, such DNA damage induced translocation-arrested states were very similar to the transient translocation intermediate states of normal pol II translocation of a non-damaged DNA template observed by molecular dynamic simulation (14). These translocation intermediate states were proposed to be rate-limiting steps during normal translocation, as they require significant conformational changes for the DNA template base to crossover the bridge helix to progress through the active site (14).

Figure iii. Structural overlay of RNA pol II elongation complexes that accommodates CPD or CydA lesion at the “above-bridge-helix” conformation (dashed circle) and causes transcriptional arrest. The bridge helix is shown in green, and RNA and DNA are shown in red and blue, respectively.
Therefore, the presence of bulky DNA lesions introduces a great steric barrier to the crossover of the bridge helix and causes pol II arrest at this ‘half-way’ translocation state. These common lesion arrest mechanisms indicate that the rate-limiting bridge helix crossover step acts as a critical checkpoint for pol II to examine the DNA template and recognize bulky DNA lesions that greatly compromise DNA backbone flexibility and integrity.

Some small DNA lesions do not affect the DNA backbone significantly and therefore do not block transcription elongation. Rather, some of these DNA lesions cause error-prone transcriptional lesion bypass. For example, 8-Oxo-2’-deoxyguanosine (8-oxo-dG), a common endogenous oxidative damage, is one such mutagenic DNA lesion (15). Pol II can either insert a matched cytosine or a mismatched adenine when it encounters 8-oxo-dG during transcription (16, 17). However, the presence of the 8-carbonyl group of 8-oxo-dG destabilizes the canonical anti conformation of template base, making ATP misinsertion and extension much more energy favorable (17). Consequently, the presence of 8-oxoG at the DNA template causes a specific C→A mutation in the RNA transcript, termed transcriptional mutagenesis (18). Emerging evidence suggests that transcriptional mutagenesis could contribute to cancer, aging, and a variety of neurodegenerative diseases.

The third class of DNA modifications are generated by endogenous enzymes. For example, the methylation of cytosine to 5-methylcytosine (5mC) by DNA methyltransferases (DNMTs) is the most common epigenetic DNA modification, often enriched at enhancer and promoter regions. 5mC functions as an epigenetic mark and plays an important role in regulating gene transcription and chromatin structure (19).
Figure iv. The structure of RNA pol II elongation complex with 5caC, in which 5caC adopts the similar “above-bridge-helix” conformation. 5caC can form a specific hydrogen bond with key residue Q531 of the Rpb2 subunit. The bridge helix is shown in green, and RNA and DNA are shown in red and blue, respectively.

On the other hand, 5mC can also undergo active demethylation, a process catalyzed by ten eleven translocation (Tet) proteins to generate the oxidized mC (oxi-mC) intermediates, 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), before being removed by thymine DNA glycosylase (TDG) to regenerate the unmodified cytosine (20). Recent evidence suggests that 5fC and 5caC are not merely reaction intermediates, but also play novel functional roles in gene regulation, as they are able to recruit various transcription factors and DNA repair protein complexes, as well as to induce transient pausing of pol II in vitro (21, 22). Recently, structural studies revealed that pol II interacts with 5caC via specific interactions between pol II and the 5caC. These specific interactions drag the majority of 5caC to be accommodated above the bridge helix (Figure iv). Further structural analysis revealed that a conserved ‘epi-DNA recognition loop’, located in the fork region of the Rpb2 subunit of pol II, is responsible...
for the recognition of 5caC in the major groove of the template strand (Figure iv) (23). Notably, the presence of 5caC can still support Watson-Crick base pair with incoming GTP substrate. However, the specific hydrogen bonds between the epi-DNA recognition loop and 5caC disrupts proper alignment of the substrate and 3’-RNA terminus, and results in a partially open conformation of the trigger loop (23). Without full closure of the trigger loop, GTP addition efficiency is significantly reduced. The Q531A mutant abolishes the ability of epi-DNA recognition loop to form the hydrogen bond with 5caC and consequently gained a significant increase in GTP incorporation specificity. Conclusively, the evidence showed that the specific hydrogen bonding between Q531 of pol II and the carboxyl group of 5caC causes a positional shift of the incoming GTP and compromises nucleotide addition, resulting in the significant reduction of pol II elongation.

Taken together, evidence suggests that RNA pol II is able to sense and detect the noncanonical DNA structures during transcription via specific interactions with the modification or lesion, based on its structural or chemical properties, and dictates the resulting transcriptional response: transcriptional arrest, pausing, and error-free or error-prone transcriptional bypass

The Introduction, in part, contains material that has been submitted for publication of the material as it may appear in Cell & Bioscience, 2016, Shin, Ji H.; Xu, Liang; Wang, Dong, Mechanism of Transcription-Coupled DNA Modification Recognition. The thesis author was the primary author of this paper.
PROJECT DESIGN & RATIONALE

Previous studies mentioned above support the idea that pol II is a specific sensor that detects DNA modifications during transcription. The specific interactions between DNA lesions/modifications and pol II govern the specific transcriptional outcomes: transcriptional arrest, pausing, and error-prone or error-free transcriptional lesion bypass. To help answer the critical question of how pol II recognizes and processes various DNA damage, lesions, and noncanonical structures, here we have selected three specific types of modifications/lesions/structural variations in DNA to investigate how they affect the pol II transcription process.

First, we focus on the recognition mechanism of pol II for non-natural nucleotides. The majority of pol II transcription studies involve pol II recognition of DNA damage and modifications that occur within the natural genome. However, the question arises as to how pol II may recognize and process unnatural, synthetic nucleotides. The development of synthetic nucleotides is a major point of interest in the growing field of synthetic biology, as they have a wide variety of applications in medicine and are useful in the study of DNA structures. We aim to investigate pol II transcription and elongation in the presence of unnatural, synthetic nucleotides (dNaM and dTPT3) (24), and the ability of pol II to distinguish between natural NTPs and the unnatural triphosphates.

Next, we look at DNA alkylation, a major type of DNA damage (25), and how pol II recognizes and processes this type of lesion, by specifically focusing on three different regional positions of alkylation on thymine (O\(^2\), N\(^3\), and O\(^4\)-EtdT). Alkylated DNA lesions generally do not cause significant pol II arrest, but depending on the position of
the alkylation, pol II demonstrates different methods of transcriptional bypass. By examining the pol II recognition and bypass mechanism, the transcriptional effects and downstream biochemical consequences of these three major damage lesions can be identified.

Lastly, we investigate the effects of a sequence-specific, small molecule (pyrrole-imidazole, or Py-Im, polyamide) that binds noncovalently to the minor groove of DNA. This class of DNA binders acts a major obstruction to pol II during transcription elongation (26, 27) as pol II cannot bypass this obstruction. The ability of Py-Im polyamides to cause strong pol II transcription inhibition makes them a potentially useful tool in cancer therapeutics. By studying the mechanism of pol II arrest in the presence of Py-Im polyamides, we can gain further insight into the pol II recognition mechanism of noncovalent minor groove binders, as well as future applications in rational drug design.

For all three of these studies, we use a purified, defined, in vitro RNA pol II transcription system. Wild-type RNA pol II used for transcription was purified from Saccharomyces cerevisiae as described in previous studies (5, 28). We investigate the pol II elongation complex in the presence of various templates and nucleotide substrates for each of the three studies. By employing this defined system, we can unambiguously gain important mechanistic insights into how RNA pol II handles these three distinct types of DNA modifications or structural variations and how transcriptional fidelity is affected during processing of these modifications.
CHAPTER 1: The recognition of unnatural nucleoside triphosphates by RNA Pol II

1.1 Introduction

During transcription, pol II is responsible for reading along the DNA template strand and subsequently incorporating the matched nucleotide with high fidelity to ensure accurate genetic transfer and to prevent transcriptional errors. Although the impacts of various DNA lesions on pol II transcription has been well-documented, little is known about how pol II may recognize and process an unnatural, synthetic nucleotide in the DNA template strand, namely, if pol II is able to distinguish the novel nucleotide, and whether pol II will be able to selectively incorporate the complementary synthetic nucleotide to form an unnatural base pair in the newly synthesized transcript. Understanding the mechanisms by which pol II recognizes and processes unnatural synthetic nucleotides would have many implications in synthetic biology and provide a framework for the rational design of unnatural synthetic nucleotides that can be effectively utilized by natural systems.

The field of synthetic biology has rapidly progressed over the past few decades, with the major goal of developing synthetic nucleotides that could be efficiently inserted into DNA and stably propagated by living cells. One of the first pioneers in the field, Benner and his team began their work by modifying natural nucleotides, specifically by rearranging the hydrogen bonding scheme between base pairs, to synthesize iso-C and iso-G, which could by incorporated by DNA and RNA polymerases (29). However, the instability of the chemical structures of iso-C and iso-G caused frequent hydrolysis of
iso-C, as well as frequent mispairing of T with iso-G (29). Benner’s team later reported another unnatural base pair (UBP), termed P-Z, that was efficiently replicated with 99.8% retention during PCR (30). The Kool group developed a new base called F, or difluorotoluene, which mimicked the structure of T but replaced oxygen atoms with fluorine (31). F and other shape-mimicking bases efficiently paired through hydrophobic forces, and suggested that hydrogen bonding may not be required for stable duplex DNA formation (31). Furthermore, Romesberg and his team focused on the synthesis of unnatural base pairs that would pair based on predominantly hydrophobic interactions. After screening from a pool of 3,600 candidate base pairs, the UBP formed by dSICS and dMMO2 were identified to be the most efficiently and accurately replicated (32). In order to optimize the insertion of dMMO2 opposite d5SICS, dNaM, a derivative of dMMO2 was synthesized to form the dSICS-dNaM base pair, which was replicated and PCR-amplified with high fidelity (33, 34). However, the overall rate of replication for dNaM-d5SICS was still slower than that of natural base pairs (24). Intriguingly, Romesberg and his team showed that in the presence of exogenous nucleotide triphosphate transporters, dNaM and dSICS could be efficiently imported into E. coli, and could be accurately replicated by the endogenous replication machinery (35), resulting in one of the first semi-synthetic organisms with an expanded genetic alphabet. Further optimization of the d5SICS nucleobase led to the development of the dTPT3 nucleobase, which forms the specific dNaM-dTPT3 UBP, and remains the most efficiently replicated UBP identified to date (36) (Figure 1.1a).
We collaborated with the Romesberg lab to test whether these unnatural nucleoside triphosphates could be efficiently incorporated during transcription and we compared how RNA pol II, E. coli RNA polymerase (E. coli RNAP), and T7 RNA polymerase (T7 RNAP) recognize the dNaM-dTPT3 UBP in vitro.

![Diagram of dNaM - dTPT3](image)

**Figure 1.1 Unnatural base pair** (a) Chemical structure of the dNaM-dTPT3 unnatural base pair. (b) Elongation complex with 10mer RNA primer and dTPT3 or dNaM in the template strand.

### 1.2 Materials and Methods

**DNA Template Design**

Wild-type RNA pol II used for transcription was purified from *Saccharomyces cerevisiae* as described in previous studies (5, 28). The DNA template and non-template oligonucleotides were purchased from IDT, unless otherwise stated. RNA primers were purchased from TriLink Biotechnologies and radiolabeled using ($\gamma$-32P) ATP and T4 Polynucleotide Kinase (NEB).
The template strand containing dTPT3 and dNaM were provided by the Romesberg Lab. The DNA sequence of the template strand was
5’-CTACCGATAAGCAGACGXTCCTCTCGATG-3’, where X is either dTPT3 or dNaM, or A, T, G, C. X is appropriately identified for each experiment. The DNA sequence of the non-template strand was 5’-CTGCTTATCGGTAG-3’. The 10mer RNA primer used was 5’-AUCGAGAGGA-3’ and the 8mer primer used was 5’-AUCGAGAG-3’.

Pol II Elongation Complex Assembly

The pol II elongation complexes for transcription assays were assembled using established methods. An aliquot of 5’-32P-labeled RNA was annealed with a 1.5-fold amount of template DNA and 2-fold amount of non-template DNA to form the RNA/DNA bubble scaffold in elongation buffer (20 mM Tris-HCl (pH = 7.5), 40 mM KCl, 5 mM MgCl2). An aliquot of annealed scaffold of RNA/DNA was then incubated with a 4-fold excess amount of pol II, E.coli RNAP, or T7 RNAP at room temperature for 10 min to ensure the formation of a pol II elongation complex. Final reaction concentrations after mixing were 25 nM scaffold, 100 nM pol II, E.coli RNAP, or T7 RNAP, 5 mM DTT, 5 mM MgCl2, 40 mM KCl, 20 mM Tris-HCl (pH = 7.5), and various concentrations of NTP. Reactions were quenched at various times by the addition of one volume of 0.5 M EDTA (pH 8.0).

In vitro pol II transcription assays

The pre-incubated polymerase/scaffold complex was mixed with an equal volume of solution containing 40 mM KCl, 20 mM Tris-HCl (pH = 7.5), 10 mM DTT, 10 mM MgCl2, and 2 mM or 1mM individual (ATP, UTP, GTP, CTP) or mixture of nucleotide
triphosphates (NTPs), or rTPT3 or rNaM. Final reaction concentrations after mixing were 25 nM scaffold, 100 nM pol II, or E.coli RNAP, T7 RNAP, 5 mM MgCl₂, and 1 mM or 500 µM NTP in elongation buffer unless stated somewhere else. Reactions were quenched at various times by addition of one volume of 0.5 M EDTA (pH = 8.0). For the TFIIS or GreB treatment, the elongation complex is mixed with an equal volume of solution containing 40 mM KCl, 20 mM Tris-HCl (pH = 7.5), 10 mM DTT, 10 mM MgCl₂, 2 mM or 1 mM NTP, and 2 µM or 1 µM TFIIS/GreB. Final reaction concentrations after mixing were 25 nM scaffold, 100 nM pol II, 5 mM MgCl₂, 1 mM or 500 µM NTP, and 1 µM or 100 nM TFIIS/GreB in the elongation buffer. Reactions were quenched at various times by the addition of one volume of 0.5 M EDTA (pH 8.0) and the quenched products were analyzed by 16% denaturing urea/TBE PAGE and visualized using a storage phosphor screen and Pharos FX imager (Bio-Rad). Nonlinear-regression data fitting was performed using Prism 6. The time dependence of product formation was fit to a one-phase association equation to determine the observed rate (k_{obs}).

1.3 Results

In order to investigate the transcriptional capabilities of the dNaM-dTPT3 UBP, a purified RNA pol II elongation complex was assembled with a mini scaffold composed of a 10mer RNA primer, annealed to the template strand containing either dNaM or dTPT3, and a non-template strand (Figure 1.1b). For comparison, E. coli RNAP and T7 RNAP elongation complexes were assembled using the same scaffold in the same manner. Nucleoside triphosphates were added to the elongation complex and incubated and quenched for varying time periods and the resulting RNA transcripts were quantified.
For single nucleotide incorporation experiments, 1 mM of individual canonical NTPs (ATP, UTP, GTP, or CTP) or either rNaM or rTPT3 were added separately to the elongation complex composed of a DNA template containing either dNaM or dTPT3. As shown in Figure 1.2a, Pol II has a strong selectivity for effectively incorporating rNaM opposite the dTPT3 template within the first few minutes, whereas all four canonical NTPs fails to be efficiently incorporated, even after prolonged incubation (24 hours). Quantitatively, the $K_{\text{obs}}$ values for rNaM incorporation is two orders of magnitude higher than the incorporation of all the four natural NTPs (Table 1.1). Intriguingly, this UBP (NaM:TPT3) selectivity is lost, when we tested the rTPT3 incorporation opposite the dNaM template (Figure 1.2b), demonstrating that the selectivity of poll II with regards to the UBP is template-specific. E. coli RNAP also demonstrated greater selectivity for the UBP over canonical NTPs, albeit less selectivity than pol II, when dTPT3 is in the template strand, preferentially incorporating rNaM and UTP. Similar to pol II, E. coli RNAP loses its selectivity bias when dNaM is present in the template, and promiscuously incorporates ATP, UTP, CTP, and rTPT3 (Figure 1.2a and 1.2b). In sharp contrast, T7 RNAP selectively incorporates rTPT3 opposite the dNaM template. Quantitatively, the $K_{\text{obs}}$ value associated with rTPT3 incorporation by T7 RNAP is similar that of the preferential rNaM incorporation by pol II into the opposite template (Table 1.1).
Figure 1.2 Single nucleotide incorporation into dTPT3 template (a) and dNaM template (b) by each of the three enzymes. The time points for these transcription assays were 1min, 10min, 1hr, 3hr, 6hr, 24hr.
Table 1.1 Single nucleotide incorporation efficiency by each enzyme

The absolute $K_{obs}$ values for each enzyme and each nucleotide substrate for the (a) dTPT3 template and the (b) dNaM template. (Units are in hr$^{-1}$)

<table>
<thead>
<tr>
<th></th>
<th>ATP</th>
<th>UTP</th>
<th>GTP</th>
<th>CTP</th>
<th>rNaM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Pol II $K_{obs}$</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.002</td>
<td>1.33</td>
</tr>
<tr>
<td>E. coli RNAP $K_{obs}$</td>
<td>0.18</td>
<td>3.63</td>
<td>0.20</td>
<td>0.13</td>
<td>6.94</td>
</tr>
<tr>
<td>T7 RNAP $K_{obs}$</td>
<td>0.50</td>
<td>1.37</td>
<td>0.47</td>
<td>0.47</td>
<td>1.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ATP</th>
<th>UTP</th>
<th>GTP</th>
<th>CTP</th>
<th>rTPT3</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Pol II $K_{obs}$</td>
<td>0.71</td>
<td>0.11</td>
<td>0.15</td>
<td>0.07</td>
<td>0.96</td>
</tr>
<tr>
<td>E. coli RNAP $K_{obs}$</td>
<td>5.61</td>
<td>4.97</td>
<td>0.16</td>
<td>1.04</td>
<td>2.09</td>
</tr>
<tr>
<td>T7 RNAP $K_{obs}$</td>
<td>0.17</td>
<td>0.17</td>
<td>0.13</td>
<td>0.19</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Our results suggest the recognition and selection of this UBP by RNA polymerases is asymmetric and RNA polymerase specific. In the comparison of the relative $K$ values for each polymerase when dTPT3 is in the template, pol II has the strongest selectivity among the three enzymes (Figure 1.3a and 1.5a). Conversely, when dNaM is in the template, T7 RNAP possesses the most selectivity for preferential incorporation of rTPT3 among the three enzymes (Figure 1.4c and 1.5b). This data indicates there are asymmetric patterns between the templates and the incoming nucleotide substrates for different types of RNA polymerases. These patterns that arise may be due to differences in the active sites of single and multi-subunit polymerases, which might be explored in future structural studies.
Figure 1.3 Quantitative analysis of single nucleotide incorporation into dTPT3 template (a) Pol II demonstrates selectivity for rNaM, while E. coli RNAP (b) and T7 RNAP (c) are less selective.

Figure 1.4 Quantitative analysis of single nucleotide incorporation into dNaM template Pol II (a) and E. coli RNAP (b) are less selective than T7 RNAP (c) in the selective incorporation of rTPT3.
Figure 1.5 Relative K values for each enzyme shows template dependent nucleotide selectivity (a) For the dTPT3 template, pol II demonstrates the most selectivity for the matched rNaM, followed by E. coli RNAP and then T7 RNAP. (b) For the dNaM template, T7 RNAP demonstrates the most selectivity for the matched rTPT3, followed by pol II and then E. coli RNAP.
In order to investigate the incorporation efficiency of the single addition of rNaM opposite different nucleotides in the template, an elongation complex was assembled with a template strand containing either A, T, G, C, or the matched dTPT3 (Figure 1.6a).

Figure 1.6 Single nucleotide incorporation of rNaM opposite different templates (a) Elongation complex with template strand containing either dTPT3, or A, T, G, C. (b) rNaM can be incorporated opposite various templates. The concentration of rNaM added was 1mM. The time points for these transcription assays were 1min, 5min, 20min, 1hr, 3hr, 20hr.

rNaM was very efficiently incorporated opposite the cytosine template by pol II, whereas rNaM can be incorporated opposite both thymine and dTPT3 templates with a similar efficiency by E. coli RNAP. In contrast, T7 RNAP can selectively incorporate rNaM opposite the dTPT3 template (Figure 1.6b). In parallel, the single addition of
rTPT3 was also investigated opposite templates containing either A, T, G, C, or the matched dNaM (Figure 1.7a). Both Pol II and E. coli RNAP can efficiently incorporate rTPT3 opposite thymine and dNaM template, whereas T7 RNAP is able to incorporate rTPT3 opposite adenine, guanine, and dNaM in a similar manner (Figure 1.7b).

\[
\begin{align*}
5' & \quad CTGCTTATCGTAG & \text{NTS} \\
3' & \quad GTAGCTCTCCTXGAGACGAATAGCCATC & \text{Template} \\
5' & \quad AUCGAGAGGA & \text{rTPT3} \\
\end{align*}
\]

\[X = \text{dNaM or A, T, G, C}\]

**Figure 1.7** Single nucleotide incorporation or rTPT3 opposite different templates (a) Elongation complex with template strand containing either dNaM, or A, T, G, C. (b) rTPT3 can be incorporated opposite various templates. The concentration of rTPT3 added was 1mM. The time points for these transcription assays were 1min, 5min, 20min, 1hr, 3hr, 20hr.

The $K_{\text{obs}}$ values are summarized in Table 1.2.Taken together, the data suggests that both rNaM and rTPT3 can be incorporated opposite various template nucleotides. However, the incorporation of rTPT3 and rNaM into the unmodified DNA template is
several orders of magnitude lower than canonical, natural NTPs. Therefore, we expect very low incorporation of rTPT3 and rNaM opposite natural templates due to the high competition in the presence of natural cognate NTP.

**Table 1.2 Single nucleotide incorporation efficiency opposite different templates** The absolute $K_{obs}$ values for each enzyme and each nucleotide substrate for the (a) dTPT3 template and the (b) dNaM template. (Units are in hr$^{-1}$)

<table>
<thead>
<tr>
<th></th>
<th>29-11A</th>
<th>29-11T</th>
<th>29-11G</th>
<th>29-11C</th>
<th>dTPT3</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Pol II $K_{obs}$</td>
<td>0.26</td>
<td>4.80</td>
<td>0.32</td>
<td>36.62</td>
<td>1.32</td>
</tr>
<tr>
<td>E. coli RNAP $K_{obs}$</td>
<td>0.79</td>
<td>11.47</td>
<td>0.24</td>
<td>4.15</td>
<td>11.19</td>
</tr>
<tr>
<td>T7 RNAP $K_{obs}$</td>
<td>0.71</td>
<td>0.66</td>
<td>0.22</td>
<td>0.50</td>
<td>1.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>29-11A</th>
<th>29-11T</th>
<th>29-11G</th>
<th>29-11C</th>
<th>dNaM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Pol II $K_{obs}$</td>
<td>0.32</td>
<td>1.02</td>
<td>0.30</td>
<td>0.69</td>
<td>0.85</td>
</tr>
<tr>
<td>E. coli RNAP $K_{obs}$</td>
<td>0.61</td>
<td>5.07</td>
<td>0.44</td>
<td>1.28</td>
<td>2.03</td>
</tr>
<tr>
<td>T7 RNAP $K_{obs}$</td>
<td>0.56</td>
<td>0.58</td>
<td>0.21</td>
<td>0.38</td>
<td>0.69</td>
</tr>
</tbody>
</table>

NTP chase experiments were conducted in which the pol II elongation complex was incubated with either 1mM rNaM or rTPT3 for two hours, and then chased by the addition of 1mM mixture of canonical NTPs. For all three polymerases, the full length transcripts could be achieved, indicating that pol II, E. coli RNAP, and T7 RNAP can continue elongation after bypassing the UBP (Figures 1.8).
Figure 1.8 NTP chase experiments (a) Elongation complex with dTPT3 or dNaM template and 2 hr incubation of rNaM, then chase with 1mM NTPs. Full length transcripts can be achieved by all three polymerases for (b) dTPT3 template and (c) dNaM template. After 2hr incubation of initiation nucleotide substrate, the time points for the NTP chase were 1min, 5min, 20min, 1hr, 3hr.
In order to investigate the efficiency of pol II transcriptional bypass of the UBP, an elongation complex was formed with a shorter, 8mer primer annealed to the template strand, with either dNaM or dTPT3 three base pairs away (Figure 1.9a). In the presence of 500µM mixture of only canonical NTPs, pol II is able to efficiently extend to the 10mer transcript and stalls before the dTPT3 site, most likely due to ineffective incorporation of the natural NTPs. However, in the presence of both 500µM NTP and 500µM rNaM, pol II is able to efficiently bypass the strong dTPT3-induced pausing site to achieve the full length transcript (Figure 1.9b), presumably as rNaM becomes preferentially incorporated opposite dTPT3 and pol II is then able to incorporate natural NTPs to achieve the full length transcript. In a parallel experiment using dNaM template, a similar pol II stall was observed immediately before the dNaM site in the presence of canonical NTPs. However, pol II has difficulty bypassing the dNaM site, even in the presence of both 500µM NTP and 500µM rTPT3, as neither the natural NTPs nor rTPT3 can be efficiently incorporated (Figure 1.9c).

TFIIS is a known transcription elongation factor that can rescue pol II pausing caused by various translocation barriers by cleaving the backtracked transcript (37-40). In order to test whether TFIIS is able to relieve dNaM-induced pol II arrest, we added 100nM TFIIS to the elongation complex along with the NTP and rTPT3 mixture. As shown in Figure 1.9c, TFIIS was ineffective in rescuing dNaM-induced pol II arrest. This further confirms the preference of pol II to selectively incorporate rNaM opposite dTPT3, rather than rTPT3 opposite dNaM, which is consistent with the previous single nucleotide incorporation data.
Figure 1.9 Elongation of the UBP by pol II (a) Elongation complex with either dTPT3 or rTPT3 in the template strand 3 base pairs away from the primer. (b) The pausing caused by dTPT3 is bypassed by pol II only in the presence of rNaM, and the addition of TFIIS does not seem to increase bypass efficiency. (c) The pausing caused by dNaM is difficult to bypass, even in the presence of rTPT3 and TFIIS. The time points for these transcription assays were 1min, 5min, 20min, 1hr, 3hr, 6hr, 24hr.
The dTPT3 template poses a strong pausing for E. coli RNAP, in sharp contrast to the dNaM template (Figure 1.10b and 1.10c). With prolonged incubation with canonical NTPs, E. coli RNAP shows some bypass of dTPT3-induced pausing, likely due to misincorporation of UTP. Efficient bypass of the pausing caused by the dTPT3 site was observed after the addition of the matched rNaM, suggesting the preferential incorporation of rNaM opposite dTPT3. In contrast, E. coli RNAP bypass of dNaM is quite efficient, likely due to efficient misincorporation of ATP or UTP opposite the dNaM site (Figure 1.10c), which is consistent with the single nucleotide addition data (Figure 1.5b). Similar to RNA pol II, we found that E. coli RNAP also demonstrates asymmetric incorporation of the dNaM–dTPT3 UBP. We also tested the effect of GreB, a TFIIS counterpart that can rescue arrested E. coli RNAP elongation complexes via cleavage of the 3’ end of the nascent RNA transcript (41). As shown in Figures 1.10b and 1.10c, addition of GreB does not improve overall bypass efficiency of the paused E. coli RNAP.
Figure 1.10 Elongation of the UBP by E. coli RNAP (a) Elongation complex with either dTPT3 or rTPT3 in the template strand 3 base pairs away from the primer. (b) Bypass of the pausing caused by dTPT3 is slightly enhanced in the presence of rNaM. (c) Pausing caused by dNaM is short-lived due to bypass via misincorporation of NTPs.

T7 RNAP also partially stalls at the 10mer-transcript, just before the dTPT3 site on the template strand in the presence of canonical NTPs, but can fully elongate with prolonged incubation, likely due to the misincorporation of UTP opposite dTPT3. With the addition of rNaM, the stalled T7 RNAP can be effectively rescued and bypass is quick and efficient (Figure 1.11a). However, elongation through the pausing caused by
dTPT3 in the template can also occur in the presence of NTPs and rTPT3. T7 RNAP also experiences fairly strong pausing with dNaM in the template when only natural NTPs are present. Similarly, T7 RNAP transcriptional arrest can be rescued when both NTPs and rTPT3 are present, demonstrating efficient bypass (Figure 1.11b).

**Figure 1.11 Elongation of the UBP by T7 RNAP** (a) The pausing caused by dTPT3 is bypassed by T7 RNAP very efficiently in the presence of rNaM. (b) The pausing caused by dNaM is bypassed very efficiently in the presence of rTPT3. The time points for these transcription assays were 1min, 5min, 20min, 1hr, 3hr, 6hr, 24hr.

### 1.4 Discussion

dNaM and dTPT3 are two synthetic nucleotides that form a specific unnatural base pair, which can be effectively replicated in vitro. During the pol II transcription process, the selective incorporation of these synthetic nucleotides occurs in a template-dependent manner. Pol II greatly prefers rNaM incorporation over canonical NTPs opposite the dTPT3 template. However, this UBP selectivity is abolished when dNaM is present in the template strand. E. coli RNAP demonstrates a similar preference to pol II for incorporating rNaM opposite dTPT3 in the template, whereas T7 RNAP prefers to incorporate rTPT3 opposite the dNaM template. Pol II and E. coli RNAP are able to
better recognize the dTPT3 template as it enters the active site and promotes the incorporation of the matched, rNaM substrate. It is possible that some multi-subunit RNAP-specific structural motifs, such as the trigger loop, are involved in this specific recognition, which may explain why T7 RNAP has a distinct selectivity. Further structural studies are required to further elucidate the molecular mechanism of dTPT3:rNaMTP recognition.

As we gain more insights into these synthetic nucleotides and how they can be efficiently replicated and transcribed, both in vitro and in vivo, the potential applications of synthetic nucleotides and new UBPs become more apparent. We may be able to develop new methods of DNA sequencing by utilizing UBPs, such as modification of the synthetic nucleotides to include a fluorophore or another detection method. Accordingly, site-specific attachment of various types of cargo to the UBP, that can then be incorporated into DNA, would allow genomic DNA to store and retrieve increased amounts of information. The ability of pol II, E. coli RNAP, and T7 RNAP to transcriptionally incorporate synthetic nucleotides in a template dependent manner also paves a way in the construction of semi-synthetic organisms with an expanded genetic code that can be transferred from DNA to RNA, and eventually to synthesize proteins with novel chemical properties.
CHAPTER 2: RNA Pol II transcriptional arrest and bypass of regioisomeric ethylated thymidine lesions

2.1 Introduction

DNA damage is often encountered by RNA pol II during transcription of the genome and can cause pol II arrest, or error prone or error free lesion bypass (2, 9, 18, 42). Previous studies suggest that the specific pol II response to DNA damage is dependent on the chemical properties of the lesion. RNA pol II is able to easily bypass 8-oxo-dG lesions but results in high chance of transcriptional mutagenesis, due to the preferential incorporation of ATP (16, 17). However, pol II becomes significantly stalled at 1-2 cisplatin damage lesions and at monofunctional platinum modifications, due to disrupted active site and hindered translocation, respectively (43, 44).

DNA alkylation is a major type of DNA damage, caused by both endogenous metabolism and environmental factors, that alters the structure of the DNA template (25). The major sites of DNA alkylation include N3 and N7 of adenine, O6 and O7 of guanine, and N3, O2, and O4 of thymine, each of which result in a different transcriptional response by pol II (45-51). While bulky alkylation can significantly block pol II transcription elongation, the effect of smaller alkylation is dependent upon the specific position of the alkylation on the nucleotide. Different regional positions of alkylation on thymine resulted in different transcriptional bypass efficiency by pol II (50, 51). In order to elucidate the specific mechanism of bypass, we investigate pol II transcription in response to three thymine bases, each alkylated at different positions, O2, N3, and O4 (Figure 2.1a). While pol II was able to bypass the three ethylated bases during transcription, our studies reveal that distinct pol II recognition of each of the three bases
alters different fidelity checkpoints during transcription, and ultimately results in different transcriptional bypass efficiencies. Fidelity checkpoints include the nucleotide insertion step, RNA transcript extension step, and the proofreading step (1). Through this systematic investigation of different ethylated positions, we gained further understanding of the mechanism of pol II recognition of alkylation damage, as well as the specific roles of each fidelity checkpoint during transcriptional bypass.

### 2.2 Materials and Methods

**DNA Template Design**

Wild-type RNA pol II used for transcription was purified from *Saccharomyces cerevisiae* as described in previous studies (5, 28). The DNA template and non-template oligonucleotides were purchased from IDT, unless otherwise stated. RNA primers were purchased from TriLink Biotechnologies and radiolabeled using (γ-32P) ATP and T4 Polynucleotide Kinase (NEB).

DNA sequence of the template strand was

5’-CTACCGATAAGCAGACGXTCCTCTCGATG-3’, where X is either dT, O2-EtdT, N3-EtdT, or O4-EtdT. The DNA sequence of the non-template strand was

5’-CTGCTTTATCGGTAG-3’. The 10mer RNA primer used was 5’-AUCGAGAGGA-3’ and the 8mer primer used was 5’-AUCGAGAG-3’.

**Pol II Elongation Complex Assembly**

The RNA pol II elongation complex was assembled as described above in the Materials and Methods section of Chapter 1. Only RNA pol II was used.
In vitro pol II transcription assays

The pol II transcription assays were carried out as described above in the Materials and Methods section of Chapter 1. Only RNA pol II was used and only canonical NTPs (ATP, UTP, GTP, CTP) were added.

2.3 Results

A purified pol II elongation complex was assembled with an 8mer primer, and three site-specifically incorporated, regioisomeric O\(^2\)-, N\(^3\)-, and O\(^4\)-EtdT lesions were introduced into the DNA template strand 3 base pairs away (Figure 2.1b). Although ethylated thymine bases alter the hydrogen bonding interactions, they do not cause a significant steric effect. To investigate pol II elongation efficiency, 1mM NTPs were added to the elongation complex. While pol II elongation proceeded smoothly through the undamaged template as expected, pol II became significantly stalled at the 10mer transcript in the presence of the N\(^3\)-EtdT lesion (Figure 2.1c). The 10mer transcript occurs right before the damaged site, indicating that nucleotide addition opposite the lesion is difficult and significantly slowed. With prolonged incubation, pol II was able to progress forward but with a minor pausing site immediately following incorporation, which suggests that the extension step from the lesion site is also weakened, but not as significantly as the addition step. After pol II bypasses the damaged site on the template, elongation resumes normally. The amount of full length transcripts of the N\(^3\)-EtdT template was significantly reduced than that of the undamaged dT template, reflecting the overall hindering effect of N\(^3\)-EtdT on transcription elongation. The N\(^3\)-EtdT lesion
seems to mainly affect the nucleotide addition and extension steps, as indicated by the major pause site at the 10mer transcript and minor pause site at the 11mer transcript.

Figure 2.1 RNA pol II transcription elongation of the damaged templates (a) Structures of the undamaged and the three alkylated thymidines. (b) The elongation complex consisting of an 8mer primer and either the undamaged or damaged template at the indicated site. (c) Major pol II pause sites occur at the 10mer and 11mer transcript lengths for the damaged templates. (d) Addition of TFIIS reduces bypass efficiency for the O²- and N³-EtdT templates. The concentration of NTPs added were 1mM and the concentration of TFIIS added was 1µM. The time points for these transcription assays were 15sec, 1min, 5min, 20min, 1hr.
In the presence of a damaged template with the O\textsuperscript{2}-EtdT lesion, pol II also stalls at the 10mer transcript (Figure 2.1c), indicating that the nucleotide addition step was difficult, similar to the N\textsuperscript{3}-EtdT lesion. Increased incubation time resulted in the decrease of 10mer transcript, but also the accumulation of 11mer product, demonstrating slow nucleotide addition and weakened extension from the O\textsuperscript{2}-EtdT site. However, the paused 11mer product was different from that of the N\textsuperscript{3}-EtdT template, suggesting that the preference for nucleotide incorporation was altered in these two different templates. The O\textsuperscript{2}-EtdT template significantly hindered pol II transcription elongation. Lastly, the O\textsuperscript{4}-EtdT template also similarly stalls pol II at the 10mer position, but bypass of the lesion site is very efficient, unlike the other two ethylated thymine bases (Figure 2.1c). Within all three of the isomeric thymine modifications, O\textsuperscript{4}-EtdT proved to be the least hindering of pol II elongation. Taken together, this data demonstrates that differential alkylation of the thymine base can cause different transcriptional bypass behaviors.

TFIIS is a known transcription elongation factor that can help rescue pol II pausing due to specific sequences, DNA lesions, and nucleosomes by promoting transcript cleavage. TFIIS was added to the elongation complex containing a damaged template to determine the effect of TFIIS on pol II transcriptional bypass. Surprisingly, the presence of TFIIS greatly reduces the bypass efficiency of pol II for the O\textsuperscript{2}- and N\textsuperscript{3}-EtdT templates as indicated by the major paused bands at the 10mer transcript. Rather than promote transcription elongation, TFIIS reduces lesion bypass in the O\textsuperscript{2}- and N\textsuperscript{3}-EtdT templates (Figure 2.1d), most likely due to the increased proofreading activity of pol II in the presence of TFIIS. This causes increased transcript cleavage, rather than lesion bypass, resulting in a majority of the pol II elongation complex to be stalled at the
damaged site. For the O\textsuperscript{4}-EtdT template, for which pol II already demonstrated efficient lesion bypass, TFIIS did not seem to have much of an effect (Figure 2.1d) and confirmed the fact that pol II greatly favors lesion bypass, rather than proofreading, for this particular alkylated thymine. Transcriptional bypass of the O\textsuperscript{2}- and N\textsuperscript{3}-EtdT lesions was difficult and further hindered by TFIIS-stimulated proofreading activity, whereas bypass of the O\textsuperscript{4}-EtdT lesion was much more efficient, with TFIIS having no significant effect.

Under the guidance of Dr. Liang Xu, single turnover experiments were conducted to quantitatively measure the nucleotide incorporation step, which also acts as the first fidelity checkpoint, opposite the lesions by adding 1mM NTPs to the elongation complex (Figure 2.2a). We found that the three different lesions demonstrated distinct selectivity preferences for nucleotide incorporation. In comparison to the undamaged template, incorporation of matched ATP was greatly reduced by \(~10^{5}\)-fold, but remained in a similar level for all three lesions, suggesting that AMP incorporation occurs in a template-dependent manner for these three lesions (Figure 2.2b and 2.2c). AMP and GMP incorporation was favored in the O\textsuperscript{2}-EtdT lesion but UMP was selectively incorporated in the N\textsuperscript{3}-EtdT lesion (Figure 2.2b and 2.2c). Surprisingly, GMP incorporation opposite the O\textsuperscript{4}-EtdT lesion was \(~10\)-fold more efficient than matched AMP incorporation. This likely accounts for the efficient pol II bypass of the O\textsuperscript{4}-EtdT lesion during the incorporation step by selectively incorporating GMP. It is apparent that different alkylation positions on the thymine base influences nucleotide selectivity during incorporation, thereby altering transcriptional fidelity.
Figure 2.2 Single nucleotide incorporation opposite the damage sites (a) Elongation complex containing the 10mer primer and the damaged template. (b) and (c) Incorporation of nucleotide substrates is lesion specific. The concentration of the NTPs added was 1mM. The time points for these transcription assays were 1 min, 5 min, 20 min, 1 hr, 3 hr, 8 hr, 24 hr.
The second fidelity checkpoint, or the extension step, in these three damaged templates was further investigated (Figure 2.3a). In the undamaged dT template, apart from the matched 11A transcript, extension from the 11G transcript was also relatively efficient. In the O²- and N³-EtdT templates, 11G extension was very inefficient (Figure 2.3b). Conversely, extension from the 11G transcript was the most efficient in the O⁴-EtdT template (Figure 2.3b). The extension efficiencies for the three different templates seem to be influenced by the different alkylation positions.

Figure 2.3 Extension from the damaged site is lesion specific. (a) Elongation complex with A, U, G, or C opposite the damaged site, from which extension occurs. (b) Extension efficiency is lesion specific, as extension from 11G is disfavored for the O²- and N³-EtdT lesions, but very efficient for O⁴-EtdT.
2.4 Discussion

By systematically studying the fidelity checkpoints involved in the bypass of the three alkylated thymine lesions, we can summarize the specific transcriptional bypass pathway for each of the O\(^2\)-, N\(^3\)-, and O\(^4\)-EtdT lesions. For the O\(^2\)-EtdT template, both ATP and GTP were favored in the addition step, but only extension from AMP was favored. As a result, the major lesion bypass pathway for O\(^2\)-EtdT occurs through ATP incorporation and the subsequent extension, and although bypass efficiency was weakened, transcriptional fidelity can be maintained. The N\(^3\)-EtdT template promoted lesion bypass through several pathways. ATP and UTP were favored in the addition step, followed by less efficient addition of CTP and GTP, but the extension step following CMP was the most efficient. As a result, the major lesion bypass pathways for N\(^3\)-EtdT can likely occur through ATP, UTP, and CTP addition and subsequent extension, with only GTP blocked by both addition and extension. Therefore, bypass of the N\(^3\)-EtdT lesion can result in loss of transcriptional fidelity with a potential mutation of A to U or C. For the O\(^4\)-EtdT template, the major lesion bypass pathway occurs through GTP incorporation and subsequent extension, as it was the fastest and most efficient. As a result, bypass of the O\(^4\)-EtdT lesion results in almost total loss of transcriptional fidelity, with a high frequency of A to G mutations. Taken together, the position of alkylation on the thymine base significantly alters pol II recognition of the lesion, which signals the fidelity checkpoints during transcription, and can result in either error-prone or error-free lesion bypass.

The mechanism of transcriptional bypass of the different lesions reveals the potential downstream cellular consequences. Although pol II stalled at the O\(^2\)-EtdT
lesion, transcriptional fidelity was maintained, suggesting that the presence of this lesion may not pose a severe threat to cells. Transcription through the N\textsuperscript{3}-EtdT lesion was also difficult, with the potential for mutagenesis. However, the slow and inefficient transcription of this lesion site could provide enough time for the cell to signal and recruit repair pathways. The O\textsuperscript{4}-EtdT lesion poses the most dangerous threat, as transcriptional bypass of the damage site is efficient and highly mutagenic, which could negatively alter gene expression. As pol II reads along the DNA template, specific recognition of DNA modifications or lesions directly influences the transcriptional response and maintenance of transcriptional fidelity.
CHAPTER 3: RNA Pol II transcriptional arrest by non-covalent minor groove binders: Py-Im polyamides

3.1 Introduction

Pol II frequently pauses, stalls, and backtracks during transcription elongation (3), especially when it encounters epigenetic modifications (21), DNA lesions (10, 52), small molecule binders (38, 53), DNA binding proteins (39), and other potential obstructions in our genome. As previously mentioned, stalled pol II can be rescued by TFIIS, a transcription factor that is known to rescue pol II pausing by cleaving the backtracked transcript (40, 42). Transcriptional arrest can also be rescued by other mechanisms, such as transcription coupled nucleotide excision repair, ubiquitination, and degradation (2, 42).

Pyrrole-imidazole polyamides, also known as Py-Im polyamides, are small molecules that can selectively target and bind to specific DNA sequences in the minor groove with high binding affinity (26, 27). Investigation of the effects of Py-Im polyamide binding on pol II transcription elongation revealed that a highly conserved motif of pol II acts as a minor groove sensor to promote pol II elongation inhibition, which occurs through direct interaction between pol II residues and the bound Py-Im polyamide.
3.2 Materials and Methods

*DNA Template Design*

Wild-type RNA pol II used for transcription was purified from *Saccharomyces cerevisiae* as described in previous studies (5, 28). The DNA template and non-template oligonucleotides were purchased from IDT, unless otherwise stated. RNA primers were purchased from TriLink Biotechnologies and radiolabeled using (γ-32P) ATP and T4 Polynucleotide Kinase (NEB).

DNA sequences of template and non-template strand used for TS binding orientation were 5’-
GACTCTTCTGACTTGGTCATACACACACTTCTCTCTCGTTGTTCTCTCGAT
TGTAA AGTGATGTGTCGGTAAT-3’ and 5’-
ATTACAACGACACATCACTTATGTCAGATCTACGCTC
ACGAGAGAGAGAAGTGTGTGATGACCAAGTCAGAAGAGTC-3’, respectively.

Sequences of template and non-template strand used for NTS binding orientation are 5’-
GACTCTTCTGACTTGGTCATACACACACTTCTCTCTCGTTGTTCTCTCGAT
TGTAA AGTGATGTGTCGGTAAT-3’ and 5’-
ATTACAACGACACATCACTTATGTCAGATCTACGCTC
ACGAGAGAGAGAAGTGTGTGATGACCAAGTCAGAAGAGTC-3’, respectively.

The RNA primer used in all these scaffolds was 5’-AUCGAGAGGA-3’.

*Pol II Elongation Complex Assembly*

The RNA pol II elongation complex was assembled as described above in the Materials and Methods section of Chapter 1. Only RNA pol II was used.
In vitro pol II transcription assays

The pol II transcription assays were carried out as described above in the Materials and Methods section of Chapter 1. Only RNA pol II was used and only canonical NTPs (ATP, UTP, GTP, CTP) were added.

3.3 Results

A purified pol II elongation complex consisting of a DNA scaffold was assembled, with a full transcription bubble upstream and a six base pair binding site downstream. Py-Im polyamides 1-4 bind to the site 5’-WGGWCW-3’ (Figure 3.1).

Figure 3.1 Py-Im polyamide structure Four Py-Im polyamides share the same 6 bp DNA target sequence. The hollow circle refers to N-methylpyrrole (Py), and the black solid circle refers to N-methylimidazole (Im).

The first scaffold contains the downstream polyamide binding site in an orientation in which pol II encounters the γ turn moiety of the hairpin Py-Im polyamide first, termed the template strand (TS) binding orientation (Figure 3.2a). The second scaffold contains the downstream binding in the opposite orientation so that pol II
encounters the C-terminal linker of polyamide first, termed the non-template strand (NTS) binding orientation (Figure 3.2b). The binding of all four Py-Im polyamides completely blocked pol II transcription elongation at specific positions upstream of their binding sites in both the TS and NTS orientations (Figure 3.2a and 3.2b). This blockage could not be bypassed even after prolonged incubation, demonstrating that Py-Im polyamide binding causes prolonged arrest of pol II at the transcription elongation phase. In the absence of Py-Im binding, pol II is able to transcribe the full length transcript. In addition, a native gel showed that the pol II elongation complex does not fall off the transcript, but rather remains in a very stable, arrested state.
Figure 3.2 Py-Im polyamides block RNA pol II transcription elongation (a) Inhibition of transcription elongation by Py-Im polyamides with a TS binding orientation and (b) with a NTS binding orientation. (c and d) Effect of TFIIS on Py-Im polyamide induced pol II arrest. The time points for these transcription assays were 10min, 20min, 30min, 1hr, and 2hr.
Py-Im polyamide induced transcriptional arrest of pol II was observed to consist of a pausing region located two to five base pairs upstream of the target binding site, as indicated by multiple pausing bands in the n-2 to n-5 region (dashed box to the right of the gel image in Figure 3.2a and 3.2b, identifies the n-2, n-3, n-4, and n-5 bands). This suggests that pol II is somehow able to sense the bound Py-Im polyamide preemptively and stalls, before the pol II active site actually reaches the binding site. Concentration dependency was measured by increasing the concentration of Py-Im polyamides, which in turn caused decreased amounts of full length transcripts and increased accumulation of the arrested bands. A single base pair mutation was introduced into the sequence specific binding site of the polyamide, which resulted in significantly diminished ligand binding affinity and abolished pol II transcriptional arrest. The ability of the Py-Im polyamide to completely block pol II during transcription is heavily dependent on the specific, matched DNA sequence between the polyamide and its binding site.

As previously mentioned, TFIIS is a transcription factor that can rescue pol II pausing or arrest at different translocation barriers (37-40). However, TFIIS was found to be insufficient in rescuing pol II arrested by Py-Im polyamide binding, as the majority of pol II remained strongly and stably arrested, regardless of binding orientations (Figure 3.2c and 3.2d). This is in direct contrast to the successful rescue of nucleosome-induced pol II arrest by TFIIS, despite the fact that the binding affinity between histones and DNA is similar to that of polyamide binding (39, 40, 54, 55). In the presence of TFIIS, the major pausing bands of pol II were shifted upstream to the n-4 or n-5 positions (Figure 3.2c and 3.2d), suggesting that pol II may adopt pre-translocation or backtracked states during arrest.
In collaboration with Dr. Liang Xu and other lab members, we identified two critical residues (R1386 and H1387) in the highly conserved Switch 1 region of the Rpb1 subunit of pol II that are proposed to have an inspective role in sensing DNA minor groove obstruction. During pol II translocation towards the Py-Im polyamide binding site, these two residues may sterically clash or directly interact with the backbones of the Py-Im polyamide (Figure 3.3). These residues may be responsible for the early inspection of the downstream DNA minor groove before unwinding of the DNA duplex, as suggested by the pausing bands upstream of the Py-Im binding site.

Figure 3.3 Molecular modeling of RNA pol II transcriptional pausing and arrest by Py-Im polyamide. The functional interplay between the two key residues, R1386 and H1387, and the polyamide-bound minor groove during RNA pol II progression. In step i, there is no steric interaction between the Py-Im polyamide and the two residues. In step ii, the polyamide and the residues come into direct contact. In step iii and iv, there is strong steric clashes between the two residues and the Py-Im polyamide, as well as the prevention of duplex unwinding, resulting in strong inhibition of the elongation complex.
3.4 Discussion

Py-Im polyamides bind to specific target sequences on the DNA minor groove and cause strong pol II transcriptional arrest, which cannot be rescued by prolonged incubation and is resistant to the effects of TFIIS. The fast association and slow dissociation of Py-Im polyamide and specific DNA sequence binding results in a strong and stably bound state of the polyamide. Pol II pauses two to five base pairs upstream of the Py-Im polyamide binding site, before the downstream transcription bubble edge has reached the binding site. Molecular modeling of the arrested pol II complex suggests that two key residues, R1386 and H1387, in the conserved Switch 1 region are responsible for sensing the Py-Im polyamide binding in the minor groove and causes the upstream stalling. The early pol II pause sites (n-4 and n-5) are likely caused by the steric clash between the two residues and the bound Py-Im polyamide during transcriptional progression of pol II towards the binding site, but before formation of the transcription bubble. The downstream pause sites (n-2 and n-3) are caused by the inability to unwind the DNA duplex.

Prolonged pol II arrest, like in the case of Py-Im polyamide binding, may initiate ubiquitination and degradation of the elongation complex (2, 42). This is particularly useful in cancer therapy by selectively targeting particular cancer genes and promoting transcription inhibition in vivo. The mechanism of Py-Im polyamide binding could be applied to the development of other small molecules that can selectively obstruct pol II transcription in a targeted, sequence specific manner. This process of disrupting the pol II transcription process via a minor groove binder involves the highly conserved Switch 1 motif, in which two specific residues are responsible for inspecting the DNA minor
groove and can sense the obstruction. By inspecting the downstream duplex DNA environment before formation of the transcription bubble, pol II is able to detect abnormal DNA modifications or lesions in the downstream DNA template, which contributes to DNA damage recognition and transcriptional fidelity.

Chapter 3, in part, contains material as it appears in Proceedings of the National Academy of Sciences of the United States of America, 2016, Xu, Liang; Wang, Wei; Gotte, Deanna; Yang, Fei; Hare, Alissa A.; Welch, Timothy R.; Li, Benjamin C.; Shin, Ji H.; Chong, Jenny; Strathern, Jeffrey; Dervan, Peter B.; Wang, Dong, RNA polymerase II senses obstruction in the DNA minor groove via a conserved sensor motif. The thesis author was a co-author of this paper.
SUMMARY & FUTURE IMPLICATIONS

Accurate recognition of DNA damage and modification by RNA polymerase II during transcription elongation is the first step in initiating the appropriate transcriptional response. This is especially important considering that pol II encounters numerous potentially harmful lesions and healthy gene expression relies on the maintenance of transcriptional fidelity by pol II. By studying the effects of synthetic nucleotides, alkylation damage lesions, and small molecule minor groove binders on pol II transcription elongation, we can gain further insight into the biochemical mechanisms and downstream pathways that govern the accurate transfer of genetic information.

The study of the transcription of synthetic nucleotides revealed many potential implications for the future of synthetic biology and the expansion of the genetic alphabet. The unnatural base pair formed by dNaM and dTPT3 was previously shown to be replicated efficiently in vitro and our recent study has shown that this UBP can also be transcribed relatively efficiently in vitro, although some specific conditions may be required. Preliminary results revealed that pol II is able to recognize and distinguish the synthetic nucleotides with the most selectivity and efficiency when dTPT3 is in the template strand and rNaM acts as the nucleotide substrate. This template-dependent selective incorporation by pol II is an important finding that may pave the way for future studies involving pol II transcription of synthetic nucleotides. By comparing three different RNA polymerases, we found specific selective incorporation abilities of each polymerase depending on the template strand and nucleotide substrate, which can be utilized to create the optimal environment for transcription of the dNaM-dTPT3 UBP. As this is still an emerging field of research, further studies may include studying the
structures of the pol II or E. coli RNAP active site when dTPT3 is present, compared to the T7 RNAP active site, to elucidate specific recognition mechanisms between multi-subunit and single subunit RNA polymerases. In addition, further optimization of the synthetic nucleobases for more efficient transcription may also be possible.

The study of the three alkylated thymine lesions revealed the roles of the insertion and extension step in transcriptional bypass, as well as potential downstream cellular consequences based on the mechanism of pol II bypass. The insertion and extension steps act as two important checkpoints that seem to govern transcriptional bypass and maintenance of transcriptional fidelity. Although the O⁴-EtdT lesion results in highly mutagenic, error-prone transcriptional bypass, the slow insertion and extension steps caused by the O²- and N³-EtdT lesions significantly reduced transcription elongation efficiency, resulting in some potential mutagenesis in the bypass of N³-EtdT, but mostly error-free bypass of O²-EtdT. By studying the fidelity checkpoints involved in various types of DNA damage sites or lesions, as we have done with the O²-, N³-, O⁴-EtdT lesions, the downstream biological consequences and cellular responses can be analyzed.

The study of non-covalent minor groove binders, Py-Im polyamides, revealed that a conserved motif of pol II acts as a sensor to sense and inspect the minor groove of DNA. As the two key residues, R1386 and H1387, sense the obstruction caused by the bound polyamide, the elongation complex enters a state of prolonged arrest. Prolonged arrest of pol II may result in ubiquitination and degradation of the elongation complex, resulting in complete blockage of transcription, which may be useful in cancer therapeutics. Small molecule, minor groove binders may be developed with a sequence specificity that binds known cancer genes with overactive transcriptional activity in order
to inhibit cell proliferation. Furthermore, by preemptively sensing the downstream DNA environment, pol II may be able to prevent potentially harmful mutations caused by certain modifications or lesions by causing complete transcription inhibition.

These three studies are highly related as they represent some of the many types of modifications, lesions, or noncanonical structures that pol II may encounter during pol II transcription elongation. By investigating the effects of these modifications or lesions, we can biochemically portray the transcriptional response of RNA polymerase II from different angles. As previously mentioned, large bulky lesions cause pol II stalling due to reduced rates of substrate incorporation. Smaller lesions, such as the alkylated thymine lesions we examined, may cause transient pausing of pol II and transcriptional bypass may or may not result in mutagenesis, depending on the type of lesion. Variations in the DNA structure, such as synthetic or modified nucleotides, also causes pausing of the pol II elongation complex, but subsequent incorporation and elongation efficiencies depend on the specific chemical or physical structure of such variations. Obstructions on the DNA template, such as bound Py-Im polyamides or nucleosomes, can act as a barrier to pol II transcription, which may result in the recruitment of elongation factors that can rescue pol II or may result in ubiquitination and degradation.

Comprehensively, RNA polymerase II is able to sense a variety of different DNA structures, modifications, or lesions during transcription elongation. The specific mechanism of recognition allows pol II to initiate the appropriate transcription coupled cellular response, including transcriptional lesion bypass, transcriptional pausing, transcriptional arrest, DNA repair, or cell death. The sensory function of pol II allows pol II to respond to a wide variety of environmental cues or stressors, and to determine how
and when the cell’s energy and resources should be optimally utilized. Primarily, the sensory and recognition processes of pol II govern transcriptional fidelity and the maintenance of genome integrity.
REFERENCES


