INVESTIGATING THE PHYSIOLOGICAL EFFECTS OF MUTATIONS IN THE PROPOSED BACKTRACK SITE OF YEAST RNA POLYMERASE II

By
JOSHUA CALIXTERIO MOSTALES

A THESIS

Presented to the Department of Chemistry and Biochemistry
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2019

AN ABSTRACT OF THE THESIS OF

Joshua Calixterio Mostales for the degree of Bachelor of Science
in the Department of Chemistry and Biochemistry to be taken June 17, 2019

Title: INVESTIGATING THE PHYSIOLOGICAL EFFECTS OF MUTATIONS IN THE PROPOSED BACKTRACK SITE OF YEAST RNA POLYMERASE II

Approved: _____

Professor Diane K. Hawley

Transcription, the first step of gene expression, is a process fundamental to all known forms of life. In eukaryotic cells, the enzyme RNA polymerase II (Pol II) executes transcription by moving forward along the DNA and transferring the encoded genetic information to messenger RNA. However, Pol II also backtracks on the DNA, causing transcription to become arrested. When backtracking occurs, Pol II slides backwards on the DNA, displacing nascent RNA from the active site into a proposed "backtrack site," comprising residues in the Pol II subunits Rpb1 and Rpb2 that interact with the RNA. The resulting stable "arrested complex" must be reactivated for elongation to continue. While backtracking has been implicated in numerous processes essential for regulating gene transcription, its physiological relevance is not yet certain. Using *Saccharomyces cerevisiae* (Baker's yeast), we have engineered individual and combinations of mutations in the Rpb1 region of the backtrack site to disrupt the protein-RNA interactions that arise from backtracking. Through phenotypic and growth comparisons between wild-type and mutant strains, we examine how impairing the binding of RNA to the backtrack site affects yeast fitness and various Pol II functions *in vivo*, providing further insight into the possible functions of Pol II during backtracking and arrest.

ACKNOWLEDGEMENTS

The author expresses his deep appreciation to Professor Diane Hawley for introducing him to scientific research and for guiding him throughout this project. As a member of her lab, I have experienced the joy of discovery, learned to quickly recover from moments of confusion and frustration, and cultivated the mental discipline, independence, and curiosity crucial for success in science and in life. I shall be forever grateful for her laboratory knowledge and skills, her patience in teaching, and her help in troubleshooting after failed experiments. I would like to thank the two other members of my thesis committee: Professor Louise Bishop for her perspective, wisdom, and much-needed encouragement and Vi Truong of the McKnight Lab for her kindness and willingness to serve on my thesis committee. I am truly grateful to be surrounded by wonderful and supportive friends: Erica Pledger for our impromptu dinners and trip to California and for enduring my endless complaints and pedantry; Adrian Gordon for our tennis "matches" and our shared appreciation for Chopin's piano compositions; Eden Brush for her advising and podcasting skills and our mutual need for caffeine; Ashley Huynh for our gossip and conversations about the medical field; Nathan Nguyen for his silliness and our long nights of "productive" studying; and Soilo Valeros for our long phone calls and for demonstrating empathy and support to me since middle school. I would also like to thank my fellow Peer Wellness Advocates at the Duck Nest Wellness Center for their passion and energy. Finally, I would like to especially thank my entire family for their never-ending support and unconditional love: my mother Jen, father Joey, stepmother Kate, stepfather Kevin, grandmother Ruth, grandfather Rio, Nana Fe, my siblings Naomi, JC, Jasmine, Jeremiah, and Chloe, and all my relatives back home in the Mariana Islands.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Gene expression is fundamental to all known forms of life, from simple bacteria to complex multicellular organisms like humans. When genes are expressed, the instructions encoded in DNA are converted to a biologically functional product. This product, generally a protein, may generate a set of observable traits called a phenotype. In the "central dogma" of molecular biology, gene expression is divided into two consecutive steps: the transcription of DNA to RNA and the translation of RNA to proteins (Figure 1). This process is analogous to cooking a dish – that is, if we think of DNA as a cookbook, genes as individual recipes, and proteins as prepared dishes, then we can cook a recipe in two steps: first, by writing down the recipe onto a sheet of paper (transcription) and, second, by cooking the dish using the written instructions (translation). Our interest in gene expression stems from its implications for human health and the complex, highly regulated mechanisms underlying this vital process. Here, I explore how the first step of gene expression – transcription – is regulated by the enzyme RNA polymerase II.

During gene transcription, a highly complex molecular machine copies the genetic information in DNA to a structurally similar molecule called RNA. In eukaryotes (organisms whose DNA is contained within a cell

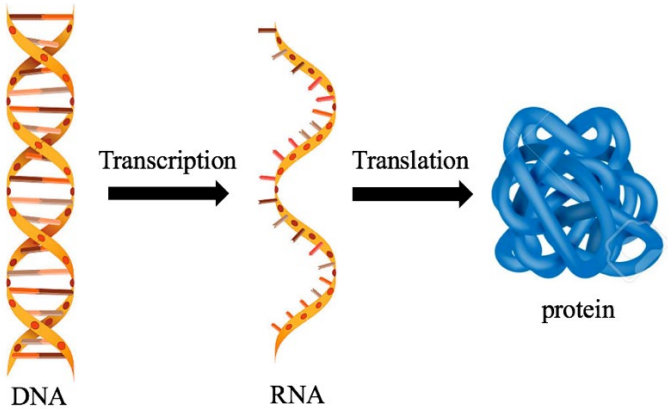


Figure 1. A depiction of the central dogma of molecular biology. Genetic information flows from DNA to RNA (transcription) and from RNA to protein (translation).

nucleus), RNA polymerase II (Pol II) forms the core of this molecular machine since it executes the polymerization of RNA in a manner that accurately reflects a particular genetic sequence of DNA. Pol II performs many functions critical to transcription; however, we typically regard transcription as a sequential process occurring in three main stages: initiation, elongation, and termination. During initiation, Pol II assembles with other proteins near the beginning of a gene. Afterwards, elongation occurs, during which Pol II reads a single strand of DNA, one base at a time, to create an RNA molecule from complementary nucleotides. Finally, during termination, the newly synthesized RNA is released, and Pol II disengages from the DNA (Figure 2).

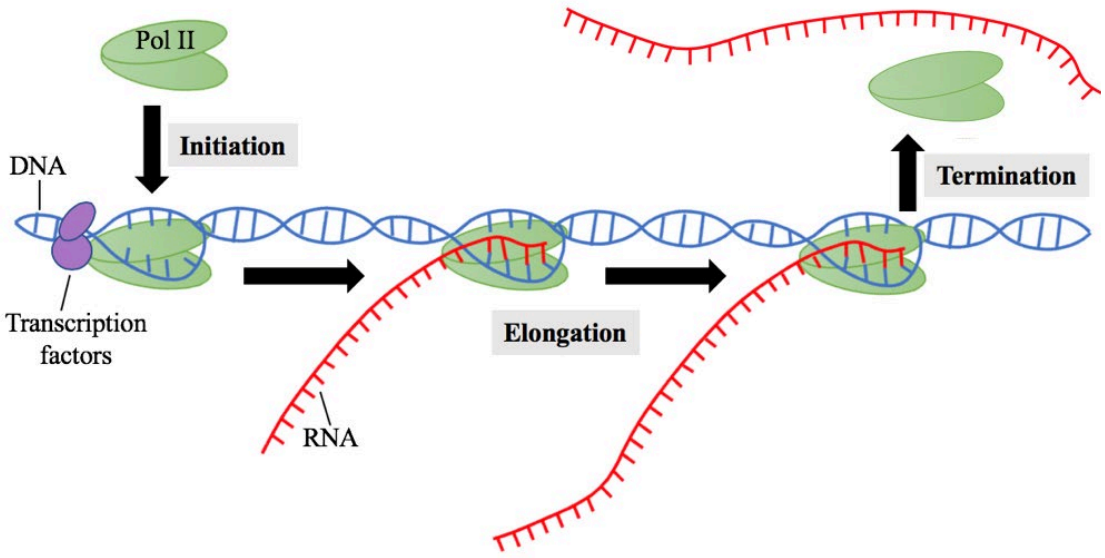


Figure 2. A simplified schematic of eukaryotic transcription and its three main stages: initiation, elongation, and termination. During initiation, Pol II binds to DNA with the help of transcription factors. Next, during elongation, Pol II travels along the DNA strand and synthesizes RNA. Finally, during termination, Pol II detaches from the DNA, and the newly synthesized RNA is made ready for translation.

Throughout the different stages of transcription, Pol II switches between many different activities, allowing it to carry out numerous functions essential for regulating and accomplishing transcription. Regulatory proteins called transcription factors interact

with Pol II to help orchestrate transcription in general or to activate or repress the transcription of specific genes. While traversing the DNA, Pol II not only shifts its association with these different proteins but also changes its own conformation, speed and even direction. Due to the highly versatile nature of Pol II and the variety of transcription factors with which it interacts, the process of transcription is remarkably dynamic. The result of this complexity, however, is that many of the mechanisms by which transcription is regulated – even by Pol II – remain poorly understood.

When visualizing the process of transcription, we typically imagine Pol II moving forward on the DNA, faithfully reading the sequences of DNA bases as it concurrently synthesizes RNA. During elongation, however, Pol II can also move backwards on the DNA, a phenomenon referred to as backtracking (Figure 3). When backtracking occurs, Pol II slides backwards on the DNA, displacing the growing end of the RNA from the active site (where new nucleotides are incorporated into the RNA) into a proposed "backtrack site." Based on structural information provided by X-ray crystallography, the backtrack site comprises amino acid residues from the Pol II subunits Rpb1 and Rpb2 that were observed to form stable interactions with the RNA[1]. These protein-RNA interactions cause Pol II to cease transcription, or arrest, while transcribing DNA. To escape arrest and resume elongation, Pol II must be reactivated by the transcription factor TFIIS.

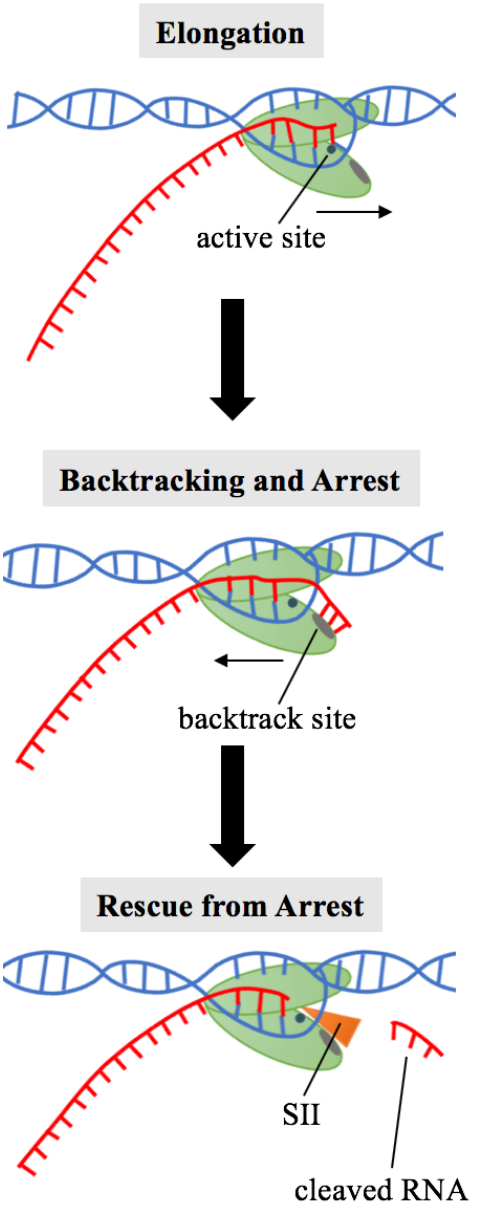


Figure 3. A simplified schematic of Pol II backtracking. Prior to backtracking, Pol II facilitates elongation by adding nucleotides to the growing RNA strand. When backtracking occurs, nascent RNA binds to the Pol II backtrack site, leading to transcription arrest. SII must then cleave RNA from the backtrack site to rescue Pol II and allow elongation to resume.

Backtracking is a conserved property in homologous RNA polymerases from all domains of life. However, given that Pol II backtracking leads to transcription arrest, one might expect backtracking to be counterproductive to the function of Pol II. We can therefore deduce that backtracking is critical for some conserved function in RNA polymerases since this phenomenon would have otherwise been eliminated by evolution. Indeed, backtracking has been implicated in many processes essential for gene regulation, from modulating the speed of Pol II to ensuring proper termination and transcription fidelity (accuracy)[2].

To better understand the importance of individual residues in the Pol II backtrack site, previous lab members and I have created a DNA library of single mutations in yeast. Each single mutation hypothetically disrupts the interaction between RNA and a specific residue in the backtrack site, which allows us to characterize how these mutations affect the

various functions of Pol II *in vivo*. My thesis, in particular, focuses on the residues that contribute to the Rpb1 region of the backtrack site and investigates their physiological role in transcription. To expand our DNA library of mutants, I have not only created a new mutation in Rpb1 but also engineered combinations of single mutations in the Rpb1 region of the backtrack site. Our motivation for combining mutations comes from our desire to examine the physiological effects of severely disrupting the binding between RNA and the backtrack site – an aspect of our research that has hitherto been unexplored. We hypothesize that our new single mutation will impair RNA binding in a way that would generate a phenotype distinguishable from wild-type. Moreover, we hypothesize that our combined mutations will impair RNA binding even more and severely compromise the ability of Pol II to form a stable arrested complex. Such a mutant strain, we hypothesize, would not only produce a more deleterious phenotype but also enable us to better address what Pol II function might be compromised and its effects on the fitness of the cell.

To investigate the behavior of Pol II, we use the fungus *Saccharomyces cerevisiae* (Baker's yeast) as our model organism (Figure 4). As one of the most widely studied eukaryotic model organisms, yeast have contributed much to the pioneering research in genetics and molecular and cell biology. Several features make yeast an ideal model organism, including their accessibility, low cost of maintenance, rapid cell cycle, and ease of genetic manipulation. With a short doubling time of 1.5 to 2 hours, yeast can be grown rapidly in culture[3]. Cultured yeast can then be used for further manipulation, experimentation, and storage, making them highly versatile and sustainable in the lab. Through DNA transformation, a common technique in which exogenous DNA is

transferred into a cell, we can easily add or delete genes from the yeast genome. Yeast can be transformed in two ways, either by introducing circular DNA molecules (plasmids) into cells or by integrating DNA fragments into the yeast genome while eliminating undesired sequences (homologous recombination).

A. B.

Figure 4. Yeast as the model organism in our study. (A) Scanning electron microscopy image of *Saccharomyces cerevisiae*, or Baker's yeast. Budding yeast cells are noticeable, and visible scars on the surface of some cells are indicative of recent budding off of daughter cells after cell division. (Source: Thesis by Rasha Aref, University of Greifswald[4]). (B) Yeast colonies grown on agar plates. Colonies typically appear grey-white with a smooth surface. (Source: Department of Veterinary Disease Biology, University of Copenhagen[5]).

Fortunately, the entire genome of *S. cerevisiae* has been determined, and its complete genomic sequence is stored in a public online domain called the Saccharomyces Genome Database (SGD). In the yeast genome, a subset of genes, including those encoding various Pol II subunits like Rpb1, are highly homologous to genes found in other eukaryotes like humans. In addition, an online archive called the Yeast Protein Database (YPD) compiles the complete set of proteins expressed by yeast and their corresponding structures, functions, and homologies to other proteins. The sequence homology between yeast polymerases and the polymerases of other eukaryotes is yet another reason for *S. cerevisiae* being our model organism of choice. Using such

databases as SGD and YPD, we are not only able to compare wild-type and mutant sequences of yeast genes but also extrapolate information from the yeast genome and proteome to other eukaryotes like plants, animals, and humans.

Experimental Approach

The overarching goal of this project is to gain a broader understanding and appreciation of Pol II backtracking. In particular, we set out to investigate the physiological consequences of mutating the Rpb1 region of the Pol II backtrack site. When backtracking occurs, Rpb1 and Rpb2 residues bind to nascent RNA in a way that causes transcription to become arrested. Therefore, we can infer that mutating Rpb1 residues in the backtrack site would disrupt the protein-RNA interactions that arise from backtracking. In this project, we have created a new single mutation and engineered combinations of mutations in Rpb1 with the intent to extensively obliterate the protein-RNA backtracking interactions. Since backtracking is essential for transcription, interfering with the event would likely influence Pol II function and, by extension, transcription as a whole.

Our experimental approach has two main phases, starting with the creation of mutant *rpb1* yeast strains and culminating with the characterization of our mutants *in vivo.* The gene encoding Rpb1 is rather large, containing an open reading frame that is 5,199 DNA base pairs (bp) long[6]. Due to the inconvenient size of the *Rpb1* gene, we first used plasmids encoding a portion of *Rpb1* that includes the backtrack site. These plasmids contain either wild-type sequences (used to generate single mutations) or mutant sequences (used to engineer combinations of mutations) that were generated by previous lab members. To introduce a mutation, we employed site-directed mutagenesis, a technique that enables us to make specific changes to the DNA sequence in a way that

alters the encoded protein sequence. In our case, we altered one or two bases on the DNA encoding the backtrack site to replace a particular Rpb1 amino acid residue with another. We induced these genetic changes such that the chemistry of the resulting residue is changed (see Figure A1 in the Appendix for a complete list of amino acids), allowing us to reduce or eliminate the interaction between the Pol II backtrack site and the RNA. Our final step in creating our mutants involves introducing the mutated sequence into yeast in the context of the entire *rpb1* gene.

After mutagenesis, we added our mutant *rpb1* sequences into yeast through homologous recombination. It is difficult to predict the effects of a mutation on an organism without scientific examination, so we proceeded to examine the effects of our mutations by conducting several *in vivo* tests. While most mutations tend to be neutral – that is, neither harmful nor beneficial – some mutations may generate a noticeable effect that either enhances or diminishes the fitness of an organism. The latter effect is more likely to occur, especially when considering the notion that an organism has evolved a set of genes best-suited for its environment. Under ideal laboratory conditions, yeast harboring mutations will often appear unaffected. However, under more stressful growth conditions, yeast mutants are more likely to show growth defects relative to wild-type. These differences in growth sensitivities can be extrapolated to changes in overall fitness or, for our purposes, to changes in Pol II function. In this project, we have characterized our mutants using *in vivo* assays that target a particular function of Pol II, such as its ability to proofread RNA and terminate transcription.

Review of Literature

Many biological processes essential to life on earth, such as cell division and differentiation, metabolism, growth, and homeostasis, are dependent on the regulation of gene expression. Each step in the flow of information from DNA to RNA to protein is intricately regulated by a vast network of proteins, and any perturbation to this high level of control poses a threat to the health and survival of the organism affected. Therefore, uncovering the mechanistic details of gene regulation and their potential implications to human health drives not only this project but also much of the research in genetics and molecular and cell biology.

As the first step of gene expression, transcription is a fundamental, highly regulated process that results in the production of RNA. In eukaryotic organisms, Pol II plays a central role in accomplishing transcription because it catalyzes the synthesis of messenger RNA (mRNA) using the information encoded in DNA. For a particular gene, the level of mRNA transcripts synthesized by Pol II typically corresponds to the amount of protein product. Gene expression, therefore, depends on the extent to which transcription is regulated. Since Pol II is the core enzyme involved in transcription, it must be tightly controlled so that transcription is carried our efficiently and faithfully. Transcriptional misregulation can lead to a broad range of diseases and syndromes in humans[7], and given the central role of Pol II in eukaryotic transcription, many research groups continue to investigate the enzymology and biochemistry of Pol II. Such research provides key information crucial to this study, including the structure, function, and regulation of Pol II and the underlying mechanisms by which it transcribes DNA.

RNA POLYMERASE II STRUCTURE AND FUNCTION

RNA polymerase II (Pol II) is composed of 12 subunits (Rpb1 to Rpb12) and has a combined mass greater than 550 kilodaltons (kDa)[6]. The amino acid sequences of all Pol II subunits in yeast have been determined[6]. Additionally, several X-ray crystal structures of Pol II have been elucidated, providing the structural basis for understanding transcription initiation and elongation and the various interactions between Pol II and its many transcription factors. Throughout the various stages of transcription and during instances when it binds other proteins, Pol II adopts a unique structural conformation. It is during these fortuitous moments that X-ray crystallography can provide meaningful structural information and emerging insights into the mechanism of transcription by Pol II.

A high-resolution X-ray crystal structure of a 10-subunit yeast Pol II reveals that the cleft at center of Pol II has four mobile elements – namely, a "core," "clamp," "shelf," and "jaw lobe" (Figure 5)[6]. The core is the

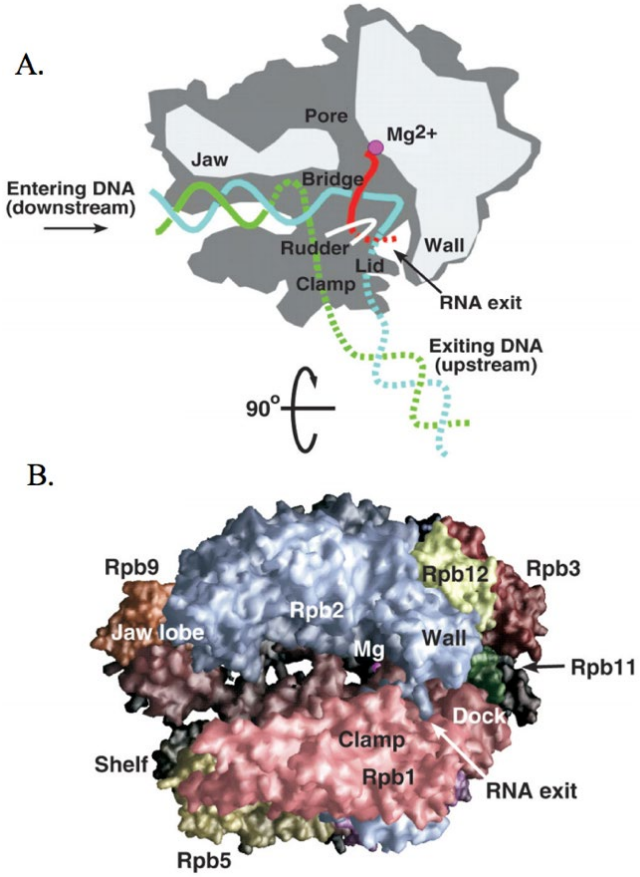


Figure 5. The 10-subunit yeast Pol II structure (from review by Steven Hahn[8]). (A) Diagram showing the top view of Pol II during elongation, containing DNA (blue and green), RNA (red), and the active site (magenta). (B) Space-filling model showing the side view of Pol II with the mobile elements and various subunits labeled by color.

largest of the four elements, containing Rpb1 and Rpb2 regions that form the active site and Rpb3, Rpb10, Rpb11, and Rpb12 subunits that are involved in Pol II assembly. During initiation, the clamp element, which comprises regions of Rpb1, Rpb2, and Rpb3, moves with a swinging motion over the cleft and active site, causing Pol II to adopt either an open or closed conformation. In the open conformation, the clamp swings open, allowing single-stranded DNA to insert into the cleft to reach the active site. The subunit Rpb7, which is not present in the aforementioned crystal structure, contributes to the closed conformation by locking the clamp in place. The other two elements, the shelf and jaw lobe, are made up of regions of Rpb1, Rpb5, and Rpb6 and Rpb1, Rpb2, and Rpb9, respectively. In contrast to the core and clamp, the shelf and jaw lobe elements rotate parallel to the cleft and generate relatively little movement. Such rotations, however, have been proposed to act as a "helical screw," allowing DNA to advance toward the Pol II active site.

For RNA synthesis to occur, substrate nucleotide triphosphates must enter the Pol II active site through a narrow, funnel-shaped channel[9]. In the active site, two magnesium ions and a "trigger loop" catalyze the incorporation of the nucleotide substrate to the growing end of the RNA. The trigger loop positions itself in the active site only when the correct substrate nucleotide (complementary to a particular DNA base) is present, sequestering the nucleotide in the active site so that it can be accurately incorporated to the nascent RNA[10]. Catalysis, as a result, is coupled to the selection of the correct nucleotide, allowing Pol II to synthesize RNA with high fidelity. After Pol II adds the correct nucleotide, it rapidly incorporates the next complementary nucleotide without having to disengage from the DNA strand. Altogether, Pol II and its consortium of

transcription factors accomplish transcription with remarkable accuracy, speed, and processivity.

Pol II Backtracking: Causes, Mechanism, and Function

During transcription, Pol II moves forward along the DNA and synthesizes mRNA. Transcription, however, is not a completely uniform process, as Pol II also backtracks, pauses, and arrests on the DNA. When extensive backtracking occurs, Pol II becomes arrested and transcription ceases. Various events have been linked to the onset of backtracking and arrest, such as the scarcity of nucleotides, the presence of nucleosomes and other obstacles on the DNA, torsional stress on the DNA due to supercoiling, DNA lesions, purine-rich segments of DNA, and nucleotide misincorporation[11].

The proposed mechanisms by which Pol II backtracks depend on the extent of backtracking. Backtracking by one to two nucleotides typically occurs as a result of the erroneous nucleotide incorporation, causing transcription to pause[12-13]. Fortunately, Pol II can quickly resolve issues related to nucleotide misincorporation by virtue of its intrinsic nuclease activity that cleaves incorrect nucleotides from the growing end of the RNA. Such intrinsic activity requires the Pol II subunit Rpb9[14] and, oftentimes, the transcription factor SII (TFIIS)[15]. Thus, intrinsic cleavage is essential for Pol II to proofread its work and to ensure that transcription is accomplished with high fidelity.

Pol II backtracking beyond two nucleotides is averted by favorable RNA base-stacking interactions and by a "gating tyrosine" residue on Rpb2[1]. However, if RNA base-stacking interactions are weak, backtracking beyond two nucleotides may occur, allowing nascent RNA to bypass the gating tyrosine and intrude further into the backtrack site where it ultimately binds. Such extensive backtracking inhibits elongation and causes Pol

II to become arrested. Once arrested, Pol II cannot rely solely on its intrinsic cleavage activity to resume transcription; instead, it requires TFIIS to cleave the backtracked RNA. TFIIS rescues Pol II from arrest by displacing nascent RNA from backtrack site and inducing the cleavage of backtracked RNA, thereby generating a new RNA end in the Pol II active site and allowing transcription to continue (Figure 6).

Backtracking has been implicated in multiple processes essential for regulating and accomplishing gene transcription. Through backtracking, Pol II can regulate gene transcription by inducing regulatory pauses and arrests, modulating the rate of elongation, ensuring transcription fidelity, and contributing to cotranscriptional RNA folding and transcription termination[2]. Our study of Pol II, in particular,

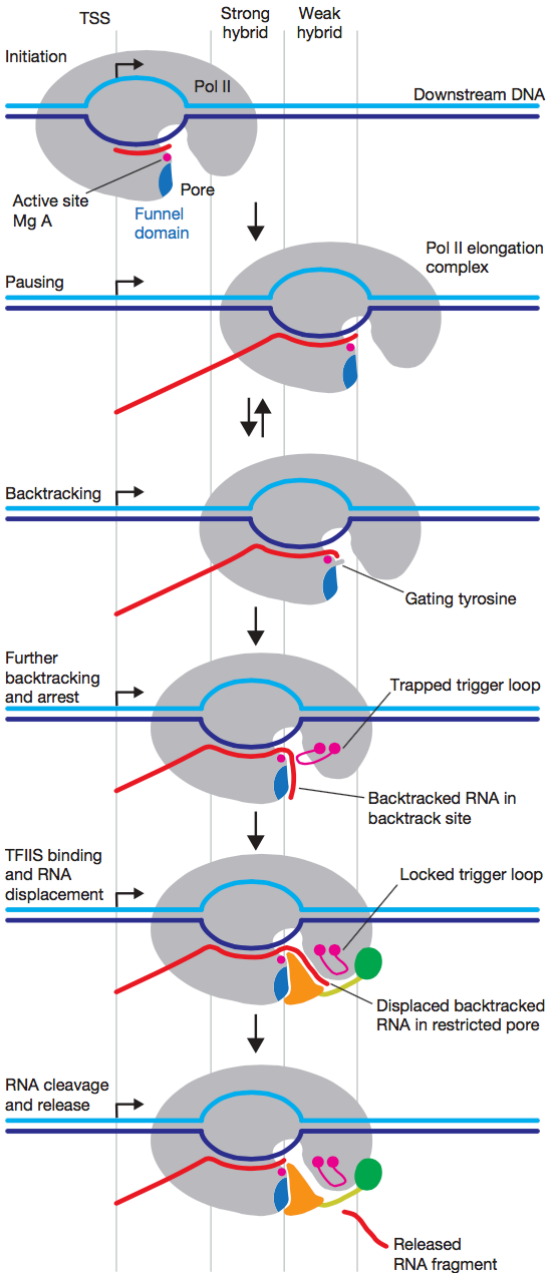


Figure 6. A more detailed schematic of Pol II pausing, backtracking, arrest, and reactivation (from Cheung and Cramer[1]). Pol II (grey) adds nucleotides to the end of the RNA strand (red) at the active cite (magenta). When extensive backtracking occurs, nascent RNA binds to the Pol II backtrack site (blue). TFIIS (orange) cleaves backtracked RNA to rescue Pol II from arrest.

investigates the importance of the backtrack site in regulating transcription fidelity, elongation speed, and termination *in vivo*.

THE POL II BACKTRACK SITE

Backtracking causes Pol II to switch between two different conformations – that is, from its "elongation complex" to an "arrested complex" – thereby causing transcription to cease. An X-ray crystal structure of the Pol II arrested complex has been solved, providing the structural basis for understanding the protein-RNA interactions that form during backtracking[1]. Structural comparisons between the elongation and arrested complexes of Pol II offer significant insights into Pol II's backtracking activity relative to its primary function of catalyzing RNA synthesis.

The Pol II backtrack site is composed of Rpb1 and Rpb2 regions that form stable interactions with backtracked RNA. However, the backtrack site is quite extensive and comprises five functional elements: the "bridge helix," "trigger loop," "fork loop," "hybrid binding domain," and "pore and funnel domain" (Figure 7). The bridge helix, trigger loop, and pore and funnel domain are parts of Rpb1, whereas

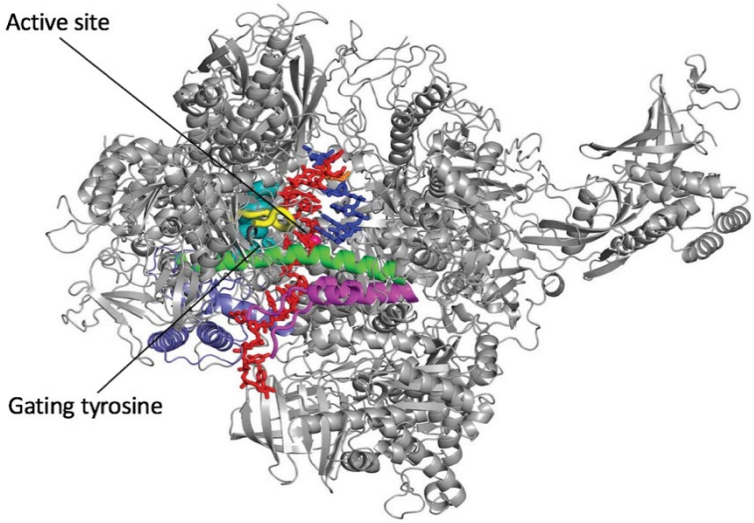


Figure 7. Structure of the Pol II arrested complex. Interactions between backtracked RNA (red) and the functional elements of the backtrack site are displayed: bridge helix (green), trigger loop (purple), fork loop (yellow), hybrid binding domain (teal), and funnel domain (lavender) (from Cheung and Cramer[1]).

the fork loop and hybrid binding domain are parts of Rpb2. When backtracking occurs, nascent RNA is extruded from the active site into the backtrack site, where it forms stable interactions with residues constituting all five functional elements.

The structure of the aforementioned arrested complex reveals molecular interactions between the residues comprising the backtrack site and eight nucleotides of backtracked RNA (Figure 8a). The first backtracked RNA nucleotide (+2) interacts with residue T827 on the bridge helix and E529 on the fork loop and stacks between the +1 nucleotide and residue Y769 (gating tyrosine) on the hybrid binding domain (Figure 8b). Backtracking beyond two nucleotides causes the backbone between nucleotides +2 and +3 to kink, and this kinked backbone contacts the R766 residue on the hybrid binding domain. The +3 nucleotide contacts the residues Q1078 and T1080 on the trigger loop. RNA that has backtracked beyond the +3 position interacts exclusively with the funnel domain of Rpb1. The backbone between nucleotides +3 and +4 contacts residue K752,

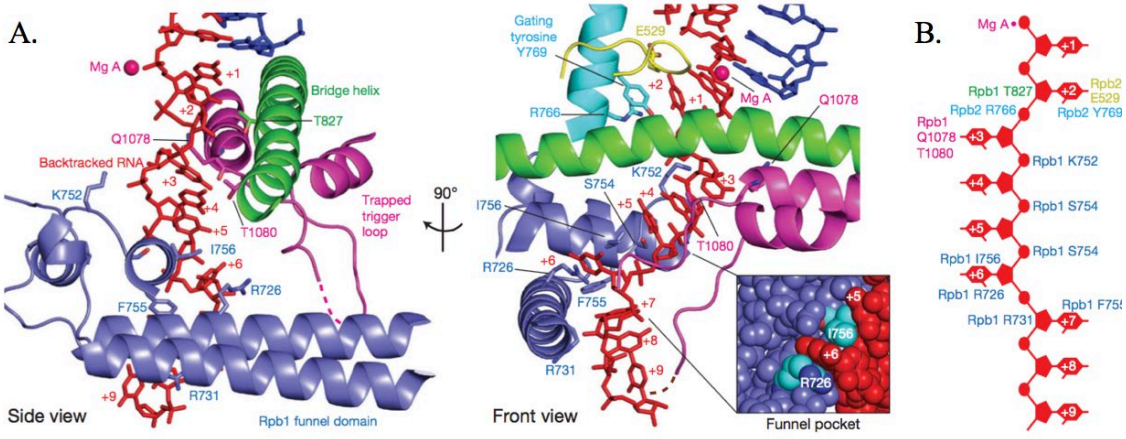


Figure 8. Structure of the backtrack site in the Pol II arrested complex and the proposed interactions between Rpb1 and Rpb2 residues and backtracked RNA. (A) Side and front views of Pol II backtrack site residues contacting the backtracked RNA (color designations similar to Figure 7). (B) A simplified schematic of the interactions between backtrack site residues and backtracked RNA (from Cheung and Cramer[1]).

and the backbones between nucleotides +4 and +5 and between +5 and +6 both contact residue S754. The base of nucleotide +6 inserts into a funnel pocket containing residues R726 and I756. The backbone between nucleotides +6 and +7 kinks again, causing nucleotide +7 to contact residue R731 and to base stack between residue F755 and nucleotides +8 and +9. In this project, we focus on three Rpb1 residues comprising the backtrack site: R726, T827, and K752. We mutated the DNA encoding these residues in order to disrupt their interactions with RNA and to assess their effects on Pol II function and cell fitness.

MATERIALS AND METHODS

Growth medias

Escherichia coli was cultured in Luria-Bertani (LB) broth containing 1% tryptone, 0.5% yeast extract, and 1% NaCl at pH 7.5. Ampicillin was added to a final concentration of 100 μg/ml in LB broth containing 1.5% agar to select for resistant bacteria. *Saccharomyces cerevisiae* was cultured in synthetic complete (SC) dropout media broth containing 0.67% bacto-yeast nitrogen base and 2% glucose and lacking amino acids used for selection[16]. Media for the blue/white screen contained 2% galactose instead of glucose. Transformations, serial dilutions, and yeast grown for the colony lift assay were plated on media made from SC dropout broth containing 2% bacto-agar.

rpb1 mutagenesis

In this project, four *rpb1* mutants were generated: one single mutant (K752T), two double mutants (R726S + K752T and K752T + T827A), and a triple mutant (R726S + K752T + T827A)[*]. Two other *rpb1* single mutants (R726S and T827A) were obtained by previous lab members (Table 1). We also engineered two restriction sites into the mutant containing the R726S mutation. The DNA sequence of the Rpb1 backtrack site is provided in the Appendix (Figure A2). A complete list of the yeast strains, plasmids, and primers used in this study are summarized in Table 2.

[*] The standard mutation nomenclature for amino acid changes are described in three parts, in the following order: a one-letter abbreviation for the original amino acid, a number denoting the position of the amino acid in the peptide sequence, and a one-letter abbreviation for the new amino acid.

TABLE 1. *RPB1* MUTATIONS IN THIS STUDY

Mutation	Method of Mutagenesis	Created by
R726S[#]	Random, obtained in screen	Yerim Lee[17]
K752T	Site-directed	Joshua Mostales
T827A	Site-directed	Mandi Severson[18]
R726S + K752T	Site-directed[^]	Joshua Mostales
K752T + T827A	Site-directed[^]	Joshua Mostales
R726S + K752T + T827A	Site-directed[^]	Joshua Mostales

[#]Sequence containing the R726S mutation was engineered in this study to have two restriction sites flanking the mutation.
^Site-directed mutagenesis was performed using plasmids already containing *rpb1* mutations.

Creation of single mutation K752T and introduction of restriction sites near mutation R726S. To produce the single mutation K752T, site-directed mutagenesis of the DNA sequence encoding the Rpb1 backtrack site was performed by PCR amplification of the region in a TOPO vector coding for wild-type Rpb1 (pDH442) using primers DHO473 and DHO436. The mutagenic forward primer (DHO473) introduced the mutation K752T into the amplified sequence (Figure 9a). Mutagenic PCR fragments and the TOPO vector were purified, digested with the restriction endonucleases *Bam*HI and *Xba*I, and ligated together prior to transformation into *E. coli* (Figure 9b). Colonies resulting from the transformation were isolated, and bacterial minipreps were performed to extract the DNA from the bacteria. To confirm the presence of the mutation in the recombinant plasmid, the DNA was sent for sequencing using the primer DHO436 which anneals close to the mutated sequence.

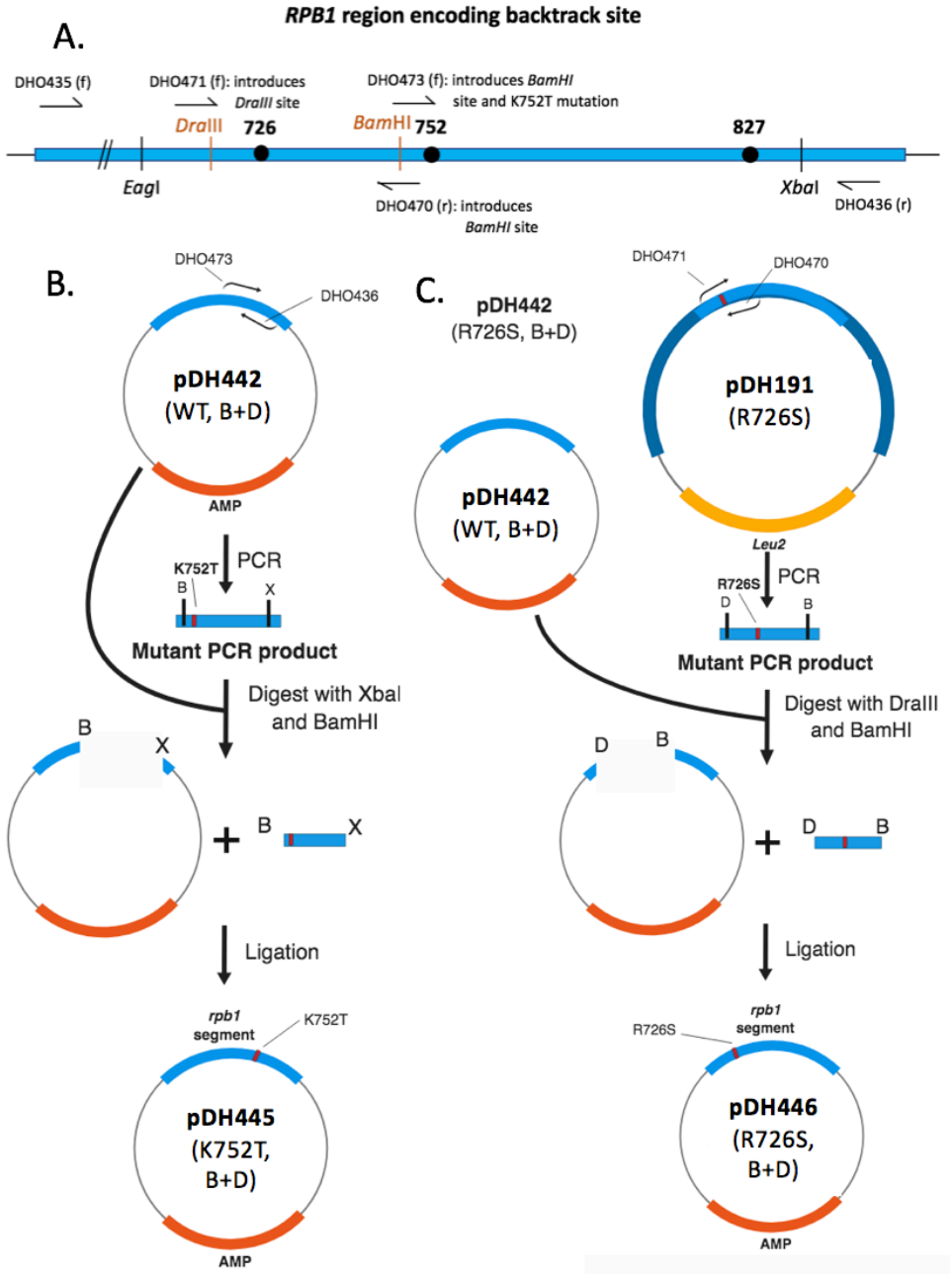


Figure 9. Creation of single mutation K752T and introduction of restriction sites near mutation R726S. (A) Schematic of the *RPB1* region encoding the backtrack site. Codons for residues of interest (R726, K752, T827) are labeled. Half arrows denote the annealing location of forward (f) and reverse (r) primers used in this study and their mutagenic effects (if any) on the wild-type sequence. Restriction sites for *Eag*I and *Xba*I occur naturally in this sequence. (B) Schematic outlining the steps taken to generate mutation K752T. (C) Schematic outlining steps taken to introduce *Bam*HI and *Dra*III restriction sites on the sequence containing mutation R726S.

To introduce the restriction sites for *Bam*HI and *Dra*III near the region encoding the mutation R726S, site-directed mutagenesis was performed by PCR amplification of the region in the pRP114 plasmid encoding *rpb1* R726S (pDH191) using primers DHO471 and DHO470 (Figure 9c). The mutagenic forward primer (DHO471) introduced the *Dra*III site, and the mutagenic reverse primer (DHO470) introduced the *Bam*HI restriction site. Mutagenic PCR fragments containing mutation R726S and the TOPO vector pDH442 were purified, digested with the restriction endonucleases *Bam*HI and *Dra*III, and ligated together prior to transformation into *E. coli.* Colonies resulting from the transformation were isolated, and bacterial minipreps were performed to extract the DNA from the bacteria. To confirm the presence of the mutation and restriction sites in the recombinant plasmid, the DNA was sent for sequencing using the primer DHO436 which anneals close to the mutated sequence.

Creation of double mutations R726S + K752T and K752T + T827A. To produce the double mutation R726S + K752T, PCR amplification of the TOPO vector encoding *rpb1* K752T (pDH445) was performed using primers DHO473 and DHO436 (Figure 10a). PCR fragments containing mutation K752T and the TOPO vector containing mutation R726S (pDH446) were purified, digested with the restriction endonucleases *Bam*HI and *Xba*I, and ligated together prior to transformation into *E. coli*. Colonies resulting from the transformation were isolated, and bacterial minipreps were performed to extract the DNA from the bacteria. To confirm the presence of the two mutations in the recombinant plasmid, the DNA was sent for sequencing using the primer DHO436 which anneals close to the mutated sequence.

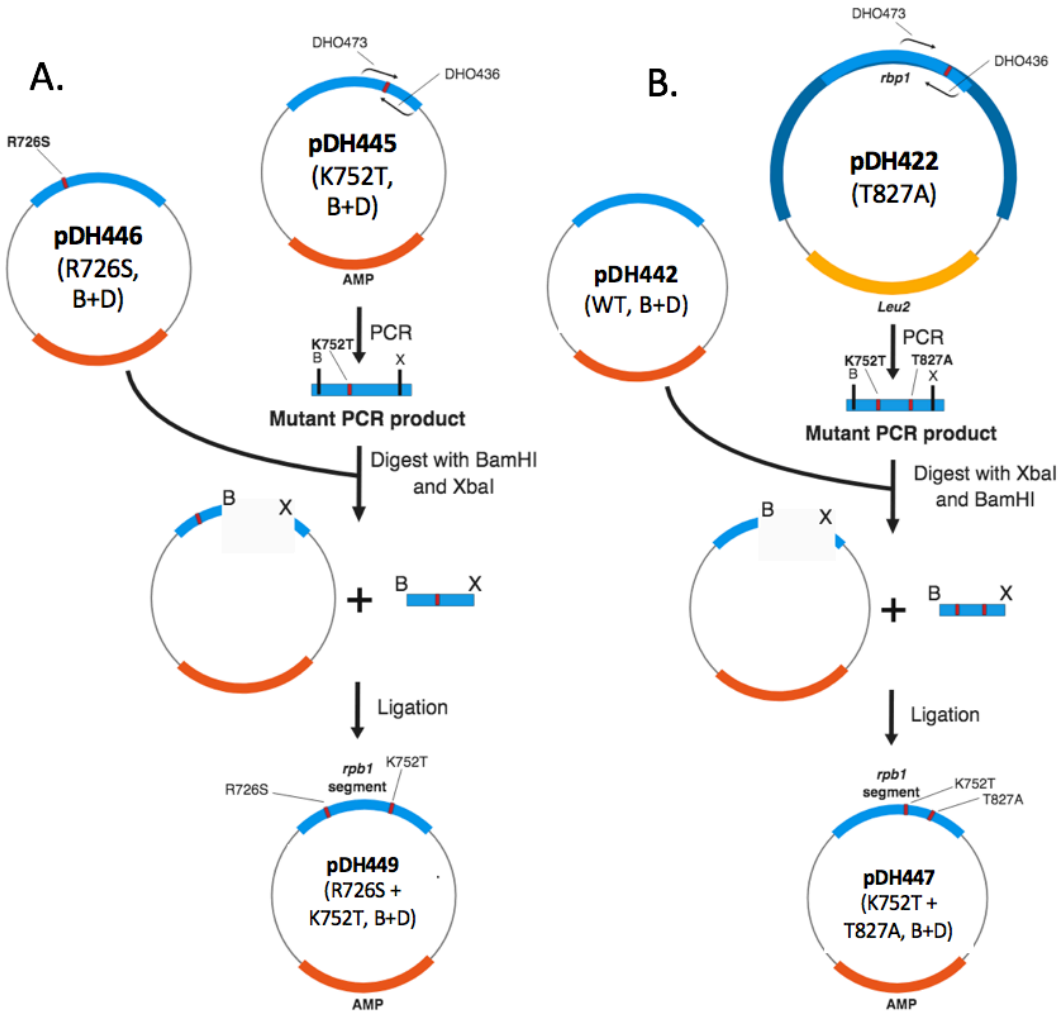


Figure 10. Creation of double mutations R726S + K752T and K752T + T827A. (A) Schematic outlining the steps taken to generate mutations R726S + K752T. (B) Schematic outlining the steps taken to generate mutations K752T + T827A. A more detailed schematic of the *RPB1* region encoding the backtrack site (shown in light blue) is provided in Figure 9a.

To generate the double mutation K752T + T827A, site-directed mutagenesis was performed by PCR amplification of the pRP114 plasmid encoding *rpb1* T827A (pDH422) using primers DHO473 and DHO436. The mutagenic forward primer (DHO473) introduced the mutation K752T and a *Bam*HI restriction site into the amplified sequence (Figure 10b). Mutagenic PCR fragments containing mutations K752T and T827A and the

TOPO vector pDH442 were purified, digested with the restriction endonucleases *Bam*HI and *Xba*I, and ligated together prior to transformation into *E. coli*. Colonies resulting from the transformation were isolated, and bacterial minipreps were performed to extract the DNA from the bacteria. To confirm the presence of the two mutations in the recombinant plasmid, the DNA was sent for sequencing using the primer DHO436 which anneals close to the mutated sequence.

Creation of triple mutation R726S + K752T + T827A. PCR amplification of the TOPO vector containing the mutation R726S and the engineered *Bam*HI and *Dra*III restriction sites (pDH446) was performed using primers DHO470 and DHO470 (Figure 11). The resulting PCR fragment containing R726S and the TOPO vector containing the mutations K752T and T827A (pDH447) were purified, digested with the restriction endonucleases *Dra*III and *Bam*HI, and ligated together prior to transformation into *E. coli.* Colonies resulting from the transformation were isolated, and bacterial minipreps were performed to extract the DNA from the bacteria. To confirm the presence of the three mutations in the recombinant plasmid, the DNA was sent for sequencing using the primer DHO436 that anneals close to the mutated sequence.

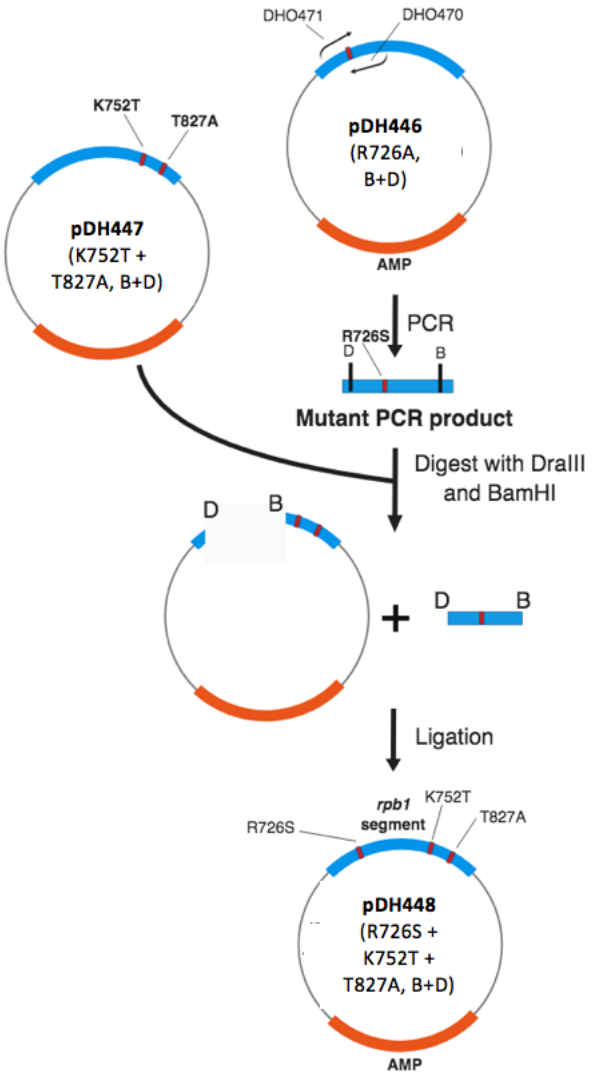


Figure 11. Creation of triple mutation R726S + K752T + T827A. The schematic outlines the steps taken to generate triple mutation. A more detailed schematic of the *RPB1* region encoding the backtrack site (shown in light blue) is provided in Figure 9a.

Creation of mutant* rpb1 *yeast strains. To generate mutant yeast strains, the yeast plasmid pRP114 containing the entire *RPB1* gene was used for homologous recombination (pDH443) (Figure 12). PCR was performed on the appropriate mutant TOPO vector using primers DHO435 and DHO436 to generate DNA fragments containing the mutation(s) of interest. For homologous recombination to occur, the

pRP114 plasmid was digested with the restriction endonucleases *Xba*I and *Eag*I, and the resulting cut plasmid was combined with the PCR product through transformation with competent yeast cells (DHY464), allowing the wild-type sequence encoding the Rpb1 backtrack site to be replaced with the mutant sequence contained in the PCR product. Since pRP114 contains the gene *LEU2*, transformations were done on leucine-selective media (SC-Leu) to select for cells containing the intact pRP114 that resulted from recombination. Transformant colonies were streaked onto SC-Leu media containing 5-fluoroorotic acid (5-FOA) to cure cells of the pRP112 plasmid that contains wild-type *RPB1.* Plasmids were subsequently extracted from yeast cells, transformed into *E. coli*, isolated from resulting bacterial colonies, and sent for sequencing to confirm the presence of *rpb1* mutations (using primer DHO436). Once the sequences were confirmed to have the correct mutation(s), wild-type (pDH443) and mutant (pDH450, pDH451, pDH452, pDH453) pRP114 plasmids were transformed into yeast (DHY678). Cells were plated onto SC-Leu to allow for the selection of colonies that have effectively incorporated the pRP114 plasmid. Transformant colonies were streaked onto SC-Leu media containing 5-FOA to cure cells of the pRP112 plasmid containing wild-type *RPB1.*

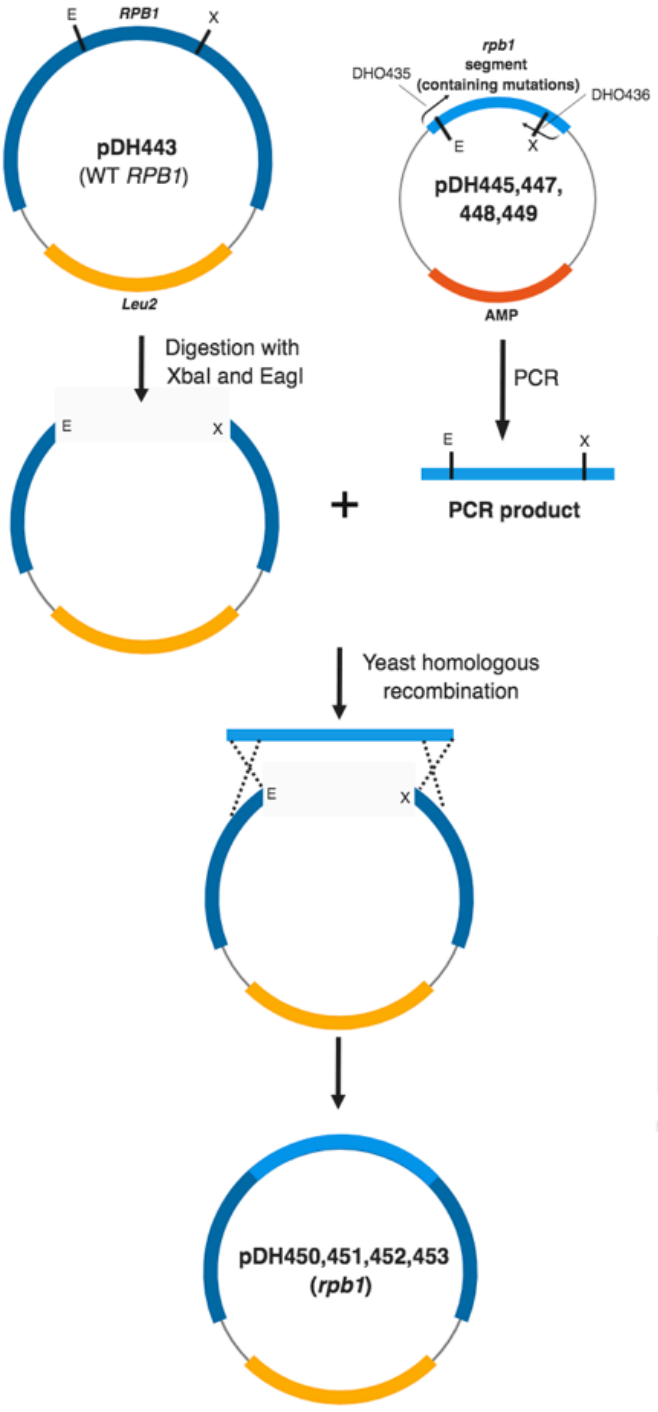


Figure 12. Yeast homologous recombination. *rpb1* mutants were created by transformation of yeast pRP114 vector (digested with *Xba*I and *Eag*I restriction endonucleases) and mutagenic *rpb1* PCR fragments.

Characterization of *rpb1* mutations.

Temperature sensitivity. The overall fitness of mutant *rpb1* strains was assayed using a spot test analysis. Mutant *rpb1* strains cured of pRP112 were grown to log-phase. Cells were diluted to an optical density (OD) of 0.6 at a wavelength of 600 nm (OD_{600} = 0.6), and then fivefold serial dilutions of log-phase cells were spotted onto YEPD media using a 48-well frogger. Cells were incubated at 18°C, 30°C, and 38°C and scored relative to wild-type after 3-7 days of incubation.

Drug sensitivity. Sensitivity to the drugs canavanine and mycophenolic acid (MPA) were tested to gain an initial assessment of Pol II fidelity and elongation speed, respectively. Mutant *rpb1* strains cured of pRP112 were grown to log-phase. Cells were diluted to OD_{600} = 0.6, and fivefold serial dilutions of log-phase cells were spotted using a 48-well frogger onto either leucine- and arginine-selective media (SC-Leu-Arg) containing 500 µg/mL canavanine or Sc-Leu media containing 50 µg/mL MPA. Strains spotted on control plates without drugs (SC-Leu) were assessed after 2-3 days of incubation at 30°C. Strains were scored relative to wild-type after 6 days of incubation at 30°C.

Blue/white screen. Termination efficiency was assayed using a colony lift colorimetric assay for β-galactosidase. Mutant *rpb1* alleles on pRP114 were introduced into yeast strains containing the pL101Btrp reporter plasmid (DHY709) through yeast transformation. Cells were plated on leucine- and tryptophan-selective media (SC-Leu-Trp) to select for colonies that have effectively incorporated the pRP114 and retained the pL101Btrp plasmids. Transformant colonies were streaked onto SC-Leu-Trp media containing 5-FOA to cure cells of the pRP112 plasmid containing wild-type *RPB1*. Cured

cells were cultured and aliquoted onto tryptophan-selective (-Trp) galactose media to induce the expression of the *lacZ.* Colonies were transferred onto filter paper, briefly submerged in liquid nitrogen to induce cell lysis, and exposed to X-gal to indicate the expression of *lacZ*. Up to 24 hours after X-gal exposure, mutant colonies were evaluated for termination deficiencies based on their color intensities relative to wild-type colonies.

TABLE 2. YEAST STRAINS, PLASMIDS, AND PRIMERS

	Name	Description	Source
Yeast Strains	DHY464	MATalpha his3Δ1 leu2Δ0 MET15 trp1Δ ura3Δ0 can1-100 cup1::HYG rpb1 Δ::NATMX4 [pRP112]	Hawley Lab
	DHY678	MATalpha his3Δ1 leu2Δ0 MET15 trp1Δ ura3Δ0 can1-100 cup1::HYG rpb1 Δ::NATMX4 [pRP112]	Hawley Lab
	DHY709	MATalpha his3Δ1 leu2Δ0 MET15 trp1Δ ura3Δ0 can1-100 cup1::HYG rpb1 Δ::NATMX4 [pRP112] [p101BTrp]	Hawley Lab
	DHY1033	MATalpha his3Δ1 leu2Δ0 MET15 trp1Δ ura3Δ0 can1-100 cup1::HYG rpb1 Δ::NATMX4 [pDH191]	Hawley Lab
	DHY1043	MATalpha his3Δ1 leu2Δ0 MET15 trp1Δ ura3Δ0 can1-100 cup1::HYG rpb1 Δ::NATMX4 [pDH450]	This study
	DHY1044	MATalpha his3Δ1 leu2Δ0 MET15 trp1Δ ura3Δ0 can1-100 cup1::HYG rpb1 Δ::NATMX4 [pDH451]	This study
	DHY1045	MATalpha his3Δ1 leu2Δ0 MET15 trp1Δ ura3Δ0 can1-100 cup1::HYG rpb1 Δ::NATMX4 [pDH452]	This study
	DHY1046	MATalpha his3Δ1 leu2Δ0 MET15 trp1Δ ura3Δ0 can1-100 cup1::HYG rpb1 Δ::NATMX4 [pDH453]	This study
Plasmids	pL101Btrp	*CEN* origin, TRP1 marker, contains ADH2 *LACZ* reporter gene	Hawley Lab
	pRP112	*RPB1, URA3, CEN* origin	Young Lab (MIT)[19]
	pRP114	*RPB1, LEU2, CEN* origin	Young Lab (MIT)[19]
	pDH191	pRP114 variant with *rpb1*-R726S	Hawley Lab
	pDH422	pRP114 variant with *rpb1*-T827A	Hawley Lab
	pDH442	TOPO vector with *RPB1* B+D	Hawley Lab
	pDH443	pRP114 variant with *RPB1, LEU*	Hawley Lab
	pDH445	TOPO vector with *rpb1*-K752T B+D	This study
	pDH446	TOPO vector with *rpb1*-R726S B+D	This study
	pDH447	TOPO vector with *rpb1*-K752T+T827A B+D	This study
	pDH448	TOPO vector with *rpb1*-R726S+K752T+T827A B+D	This study
	pDH449	TOPO vector with *rpb1*-R726S+K752T B+D	This study
	pDH450	pRP114 variant with *rpb1*-K752T B+D	This study
	pDH451	pRP114 variant with *rpb1*-R726S+K752T B+D	This study
	pDH452	pRP114 variant with *rpb1*-K752T+T827A B+D	This study
	pDH453	pRP114 variant with *rpb1*-R726S+K752T+T827A B+D	This study
Primers	DHO435	Forward 5'-TCTCAGTAAAAGACTTCACTAGTTTGG	Hawley Lab
	DHO436	Reverse 5'- CAATAGTGAGGGATCAAGGGTATG	Hawley Lab
	DHO470	Reverse 5'-TATTAATAAATGAACCCTTGGATCCTGCCAT AACCATTTGTTTCAC	Hawley Lab
	DHO471	Forward 5'-TGCTAAACATGGTATGACACTCCGTGAGTCT TTTGAG	Hawley Lab
	DHO473	Forward 5'-GCAGGATCCACGGGTTCATTTATTAATATCG	Hawley Lab

Results and Discussion

RNA polymerase II is the core enzyme that catalyzes gene transcription in eukaryotic cells. During transcription, Pol II moves forward along the DNA and synthesizes messenger RNA. Under certain circumstances not yet fully understood, Pol II backtracks on the DNA, causing transcription to arrest. When backtracking occurs, nascent RNA is extruded from the active site into a backtrack site containing Rpb1 and Rpb2 residues that are proposed to form stable interactions with the RNA. TFIIS must cleave the RNA bound to the backtrack site in order for transcription to resume. To better understand the physiological relevance of Pol II backtracking, we have created mutations in the Rpb1 region of the backtrack site and characterized yeast strains harboring these mutations *in vivo*. Since these mutations impair the proposed interactions between nascent RNA and the backtrack site, we assessed the effect of these mutations on cell viability. We also examined the growth phenotypes of mutant strains under conditions that place stress on cell fitness and on various Pol II functions, such as elongation, start site recognition, and proofreading, and we also assessed termination by Pol II using a *lacZ* reporter assay.

Creation of Mutants

In this work, we created four novel yeast strains containing mutations in *rpb1*: one single mutation (K752T), two double mutations (R726S + K752T and K752T + T827A), and one triple mutation (R726S + K752T + T827A). Two single *rpb1* mutants (R726S and T827A) were obtained by previous lab members (Table 1). Our residues of interest constitute the Rpb1 region of the Pol II backtrack site (Figure 13). In wild-type Rpb1, the arginine residue at position 726 contributes to the “funnel pocket” that contacts the fifth

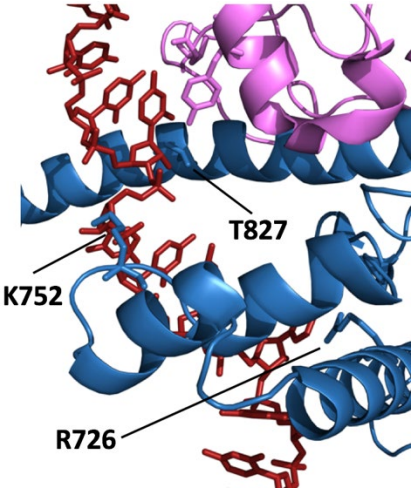


Figure 13. Structure of the backtrack site in the Pol II arrested complex containing Rpb1 (blue), Rpb2 (pink) and backtracked RNA (red). Our three residues of interest – R726, K752, and T827 – are labeled. Figure generated using PyMol software (adapted from Cheung and Cramer[1], RCSB: 3PO2).

backtracked RNA nucleotide. Since arginine is a basic side chain (which makes it positively charged under physiological conditions), our R726S mutation eliminates this positive charge and replaces arginine with a polar, partially negative serine side chain. Our second residue of interest, the lysine at position 752, is proposed to interact with the RNA backbone between the second and third backtracked nucleotide. Given that lysine is positively charged and the RNA backbone is a negatively charged, the interaction between K752 and the RNA backbone is believed to be electrostatic in nature. Our K752T mutation substitutes the lysine residue at this position with a threonine residue, which has a polar side chain that would hypothetically create unfavorable interactions with the RNA backbone and thereby impair the binding of backtracked RNA. Finally, our third residue of interest, the threonine at position 827, is proposed to interact with the first backtracked RNA by virtue of its polar side chain. Our T827A mutation alters the threonine at this position to an alanine, which abolishes any polar interactions that might arise between this residue and backtracked RNA. In addition to our single *rpb1* mutants, we also engineered combinations of mutations with the intent to significantly impair RNA binding with the backtrack site. Altogether, these *rpb1* mutations alter the chemistry of

the backtrack site in order to hypothetically disrupt the protein-RNA interactions that are necessary to form the Pol II arrested complex.

To create our mutant yeast strains, we first introduced the engineered mutations into the pRP114 plasmid through yeast homologous recombination, confirmed the presence of *rpb1* mutations on the pRP114 plasmid, and transformed these plasmids into yeast. We subsequently cured cells of the pRP112 plasmid containing wild-type *RPB1* by streaking transformants onto media containing 5-FOA. Interestingly, after plating our various strains onto 5-FOA-containing media, we obtained many colonies for our single mutants and no colonies for our double and triple mutants (Figure 14). This result strongly suggests that the double and triple mutations prevented Pol II from supporting viability in the absence of wild-type Pol II. The lethality of our double and triple mutants may be due to the disruption of the interactions between RNA and the Pol II backtrack site.

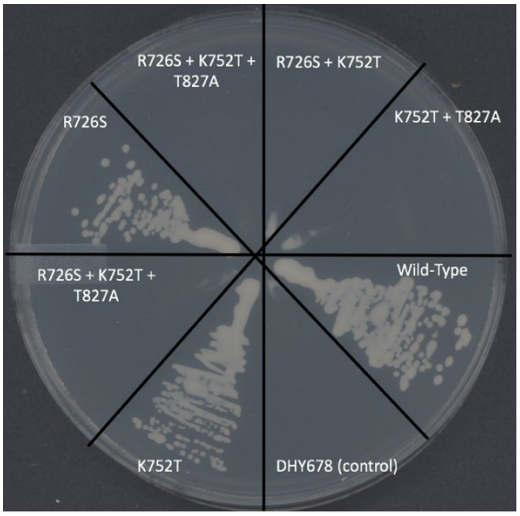


Figure 14. Various yeast strains streaked onto media containing 5-FOA to cure cells of pRP112 containing wild-type *RPB1*. Wild-type and single mutant strains (T827A not shown) produced colonies. Our double and triple mutant strains did not produce any colonies. DHY678 is the negative control strain lacking pRP112.

Altogether, these results show that cells can survive if they contain both wild-type and mutant variants of Pol II (since we obtained transformants prior to streaking on 5-

FOA), but cells cannot survive with Pol II variants alone, in the absence of wild-type Pol II (since we did not obtain colonies for our double and triple mutants on 5-FOA).

Previously, we performed the same steps to create strains of our double and triple mutants. We transferred transformant colonies onto 5-FOA to cure cells of pRP112, and intriguingly, we obtained many colonies for our single mutants and a *few* colonies for our double and triple mutants (data not shown). We transferred our double- and triple-mutant colonies from our 5-FOA plates onto SC-Leu plates to give cells harboring pRP114 the opportunity to grow even more. After obtaining colonies on SC-Leu plates, we streaked these colonies onto 5-FOA and, similar to our initial 5-FOA plates, obtained little or no colonies for our double and triple mutants (data not shown).

The few double- and triple-mutant colonies that were obtained from our initial 5-FOA plates were characterized *in vivo* using a spot test analysis at 30°C. We observed growth phenotypes similar to that of the wild-type *RPB1* strain (data not shown), suggesting that our double and triple mutants that grew on 5-FOA may have harbored undesired mutations that allowed them to appear similar to wild-type. There are several possible explanations for these observed phenotypes, and each explanation may be attributable to an overall increase in mutation rate caused by combining *rpb1* mutations. First, our double and triple mutants may have harbored mutations in the *URA3* gene, which would prevent 5-FOA selection against cells containing wild-type *RPB1* on pRP112. As a result, colonies harboring *ura3* mutations would still express wild-type *RPB1,* grow on media containing 5-FOA, and exhibit a wild-type growth phenotype on a spot test. Second, the *rpb1* gene in our double and triple mutants may have reverted back to wild-type *RPB1* (by homologous recombination between wild-type *RPB1* on pRP112

and mutant *rpb1* on pRP114). In this scenario, cells could still be cured of *RPB1* on the pRP112 plasmid, but they would still contain and express wild-type *RPB1* on the pRP114 plasmid. Third, our double and triple mutants may have harbored mutations on genes other than *URA3* or *RPB1* that might suppress the mutant phenotype. Such a mutation, although highly unlikely and difficult to ascertain, could also explain the observed phenotypes of our double and triple mutants.

Overall, these observed discrepancies between our single mutants and our double and triple mutants suggest that combining backtrack site mutations, which severely disrupt the protein-RNA interactions that arise from backtracking, might increase the overall frequency of mutations. One possible explanation for the observed lethality of our double and triple mutants is that the backtrack site is important for preventing more extensive backtracking from occurring. That is, the residues constituting the backtrack site might inhibit further backtracking by "catching" backtracked RNA and preventing Pol II from sliding backwards. Thus, if the backtrack site is significantly impaired, further backtracking can occur by many nucleotides, resulting in genomic instability and increased mutation frequency[2,19]. This would explain the growth of only a few colonies on 5-FOA, as these colonies might have incurred mutations that confer 5-FOA resistance or allow cells to exhibit a wild-type growth phenotype.

IN VIVO CHARACTERIZATION OF MUTANTS

Temperature. After establishing that our single mutants were viable and our double and triple mutants were inviable, we investigated the effects of our *rpb1* single mutations on the fitness of the cell. Growth comparisons at stressful temperatures can reveal differences in cell fitness between wild-type and mutant strains. In our study, we

introduced mutations in the Rpb1 region of the backtrack site, which hypothetically compromises the normal backtracking activity of Pol II, and examined the relative fitness our mutant strains at three different temperatures (18 °C, 30 °C, and 38°C) using a series of spot tests (Figure 15). We scored mutant growth phenotypes relative to wild-type using a scoring key outlined in Table 3. Yeast grown at temperatures near 30 °C grow optimally, while yeast grown at 18 °C and 38 °C are under cold and heat stress, respectively. At optimum temperatures (30 °C), our R726S and K752T mutants grew normally, whereas our T827A mutant grew significantly worse than the other strains (<<). Under heat stress (38 °C), our R726S and K752T mutants also grew similarly to wild-type, while our T827A mutant exhibited significantly less growth (<<). Growth sensitivities at 18 °C were more pronounced, with our K752T mutant growing much worse than wild-type (<<) and our T827A mutant exhibiting very little growth (<<<) even after 8 days. Only the R726S mutant grew similarly to wild-type at 18 °C. The data for our temperature sensitivity assay are summarized in Table 3.

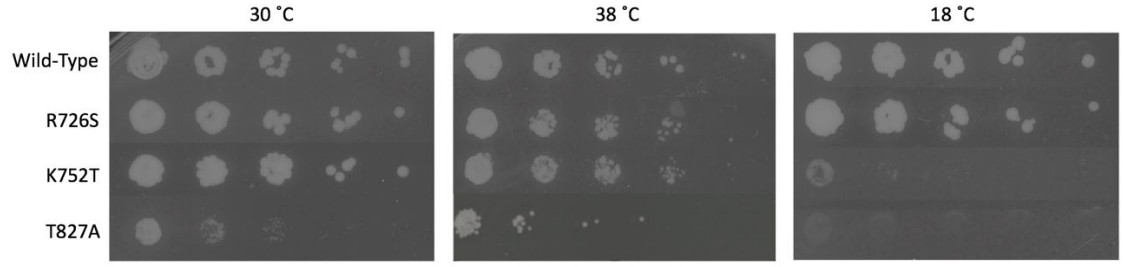


Figure 15. Cell fitness phenotypes of our *rbp1* single mutant strains at 30 °C (after 3 days), 38 °C (after 3-4 days), and 18 °C (after 7-8 days).

TABLE 3. TEMPERATURE SENSITIVITY PHENOTYPES OF *RPB1* MUTANTS

Mutation	30 °C	38 °C	18 °C
R726S	WT	WT	WT
K752T	WT	WT	<<
T827A	<<	<<	<<<

Scoring key (relative to wild-type): < slightly less growth, << significantly less growth, <<< extreme growth defect

The results of our temperature sensitivity assay suggest that under ideal growth conditions, the R726S and K752T mutations each do not reduce the overall fitness of the cell. Interestingly, even under thermal stress at 18 °C and 38 °C, the R726S mutation does not confer any phenotype that compromises cell fitness. Furthermore, the K752T mutation is detrimental to the cell only under cold stress, whereas the T827A mutation reduces cell fitness even at normal temperatures. Thus, the novel mutation created in this project, K752T, appears intermediate between R726S and T827A in terms of its effects on cell fitness.

Canavanine. To examine the effects of our *rpb1* mutations on transcription fidelity, we screened our mutants for sensitivity to canavanine. Canavanine is a toxic chemical that is structurally analogous to the amino acid arginine. Arginine and canavanine are both imported into yeast cells by an arginine permease encoded by the *CAN1* locus. The strains used in this study contain a *can1-100* gene containing a premature translation stop codon. This mutant *can1-100* gene prevents the expression of the arginine permease, thereby resulting in canavanine resistance. However, Pol II can potentially introduce transcriptional errors that allow the expression of the arginine permease encoded in the *can1-100* allele, causing canavanine import into the cell and possible lethality[14]. Errors in transcription, or reduced transcription fidelity, can be

attributed to various processes, including issues related to Pol II's intrinsic nuclease activity, backtracking and elongation speed.

Our results demonstrate that all *rpb1* mutants exhibit some level of canavanine sensitivity (Figure 16). The K752T and T827A mutants were lethal when grown with with canavanine (no growth after 7 days), while the R726S mutant exhibited considerable sensitivity to canavanine (<<). The results of our canavanine assay (summarized in Table 4) suggest that the Rpb1 backtrack site residues K752 and T827 are essential for the ability of Pol II to proofread RNA and ensure transcription fidelity. Unlike the K752T and T827A mutants, however, the R726S mutant does not display a lethal phenotype when grown with canavanine. Nonetheless, its sensitivity to canavanine suggests that the R726 residue is somewhat important for transcription fidelity.

Our control plate (SC-Leu) shows that there was an issue with plating the serial dilutions, since some mutants did not show a serial decrease in cell concentration (e.g., K752T). Furthermore, our wild-type strain grew the worst on the control plate, another indication that errors were made in plating or diluting the cells. Nonetheless, the results demonstrate that there were defects in transcription fidelity, since the K752T and T827A mutants did not exhibit any growth with canavanine. To draw more definitive conclusions about the effects of these mutations on transcription fidelity *in vivo*, this assay must be replicated without errors in plating or making serial dilutions.

Mycophenolic acid. To investigate the effects of our *rpb1* mutations on elongation speed and start site recognition, we screened our mutants for sensitivity to mycophenolic acid (MPA). MPA is a competitive inhibitor of the major form of IMP dehydrogenase (IMPDH), an enzyme involved in guanosine triphosphate (GTP)

biosynthesis. GTP is a nucleotide needed for RNA synthesis; therefore, exposure to MPA will result in decreased intracellular levels of GTP and create issues related to transcription elongation. The *IMD2* gene, which encodes IMPDH, has two start sites for transcription: an upstream start site used when GTP levels are high and a downstream start site used when GTP levels are low[20]. Use of the upstream start site results in premature termination of transcription and no production of the enzyme, while use of the downstream start site results in the expression of IMPDH. Additionally, use of the upstream and downstream start sites have been associated with a fast and slow elongation speeds, respectively. Wild-type Pol II can shift between both start sites depending on the intracellular abundance of GTP; however, some Pol II mutants have been shown to have problems shifting between the two sites or to prefer one site over the other[20].

We observed that all *rpb1* single mutants display no sensitivity to MPA (Figure 16). The results of the MPA assay (summarized in Table 4) indicate that the Rpb1 backtrack site residues R726, K752, and T827 likely do not increase the elongation speed of Pol II and/or influence its ability to recognize appropriate start sites for transcription. Previous unpublished data from the lab also indicates that the R726S variant has a preference for downstream start sites, which is consistent with the lack of MPA sensitivity shown in this study. Altogether, our results suggest that the Pol II variants under study can react to GTP depletion quite well by recognizing the downstream start site, thereby increasing the expression of IMP dehydrogenase. Our results also suggest that the Pol II variants might elongate similarly or more slowly than wild-type, since the use of downstream start sites is correlated with slow elongation speeds. However, we would

need to explicitly demonstrate that these Pol II variants exhibit slower elongation speeds and recognize the downstream start site to be certain of these conclusions.

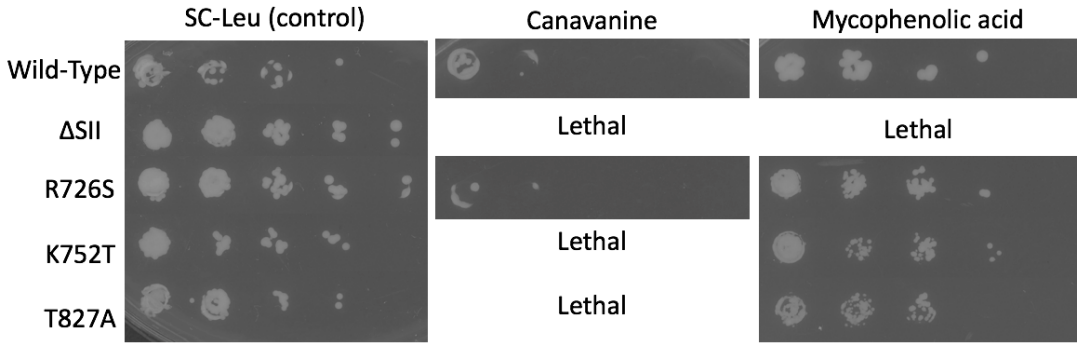


Figure 16. Drug sensitivity phenotypes of our *rbp1* single mutant strains (after 7 days). The strain ΔSII is a control strain lacking the transcription factor TFIIS.

TABLE 4. DRUG SENSITIVITY PHENOTYPES OF *RPB1* MUTANTS

Mutation	Canavanine	MPA
R726S	<<	WT
K752T	lethal	WT
T827A	lethal	WT

Scoring key (relative to wild-type): < slightly less growth,
<< significantly less growth, <<< extreme growth defect

Because of issues related to diluting or plating the cells (as revealed by the growth phenotypes on the control plate), it is difficult to ascertain that our mutants were not MPA-sensitive. Given that the wild-type strain grew worst on the control plate, it might be that the mutants are less sensitive to MPA than wild-type. In particular, the T827A mutant might be the least MPA-sensitive, since it had the worst growth phenotype among all the single mutants and exhibited significant growth defects under optimal conditions (Figure 15), yet it grew similarly to wild-type when grown with MPA. Therefore, this assay must be repeated without errors in plating or making serial dilutions in order to

draw more definitive conclusions about the effects of these mutations on elongation speed and start site recognition *in vivo*.

Blue/white screen. Using a *lacZ* reporter assay, we screened our *rpb1* mutants for defects in transcription termination. We introduced our *rpb1* mutants to strains containing the β-galactosidase reporter plasmid pL101Btrp. This plasmid contains the *lacZ* gene from *E. coli*, a yeast *GAL10* promoter, and a polyadenylation-dependent termination signal from the *ADH2* gene between the *GAL10* promoter and *lacZ* coding sequence (Figure 17). The *ADH2* terminator was placed within the intron of the gene *RP51*, which encodes a yeast ribosomal protein. Galactose (in the absence of glucose) induces transcription from the *GAL10* promoter, and Pol II will either terminate transcription due to the *ADH2* terminator or fail to terminate, resulting in the expression of the downstream *lacZ* reporter gene. The protein encoded by *lacZ* is β-galactosidase, an enzyme that turns the chemical X-gal into a blue-colored product. In this assay, the intensity of the blue color is proportional to the extent of terminator read-through – that is, any color darker than wild-type indicates defects related to transcription termination, whereas any color lighter than wild-type denotes improved termination efficiency.

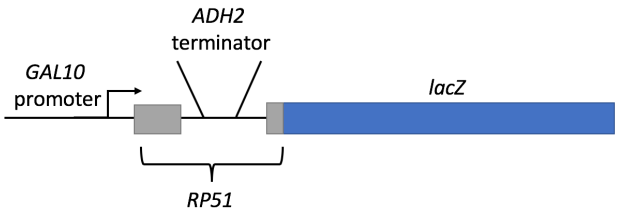


Figure 17. Schematic of the *lacZ* reporter construct used in our blue/white assay.

We observed a white phenotype for all our *rpb1* single mutants (Figure 18). These results suggest that mutating the residues R726, K752, and T827 causes increased

termination efficiency since the Pol II variants do not read through the *ADH2* terminator. Consistent with our results from the MPA assay, which suggest that our Pol II variants are slow elongating polymerases, we surmise that our mutants in the *lacZ* assay appear white because their polymerases elongate more slowly than wild-type. Slower polymerases, we hypothesize, recognize the *ADH2* termination sequence more accurately, preventing expression of the downstream *lacZ* gene. The data for our blue/white screen is summarized in Table 5.

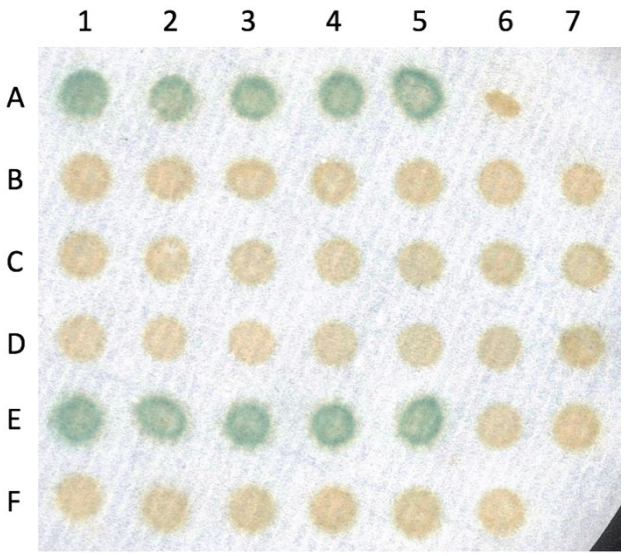


Figure 18. Blue/white phenotypes of our *rbp1* single mutant strains. Blue colonies (not shown) are indicative of *ADH2* termination read-through and *lacZ* expression, whereas white colonies are indicative of efficient termination. Wild-type colonies appear green due to a low frequency of read-through. The position of each mutant in the figure is specified in Table 5.

TABLE 5. BLUE/WHITE SCREEN OF *RPB1* MUTANTS

Mutation	Positions on Figure 18	Color
WT #1	A1-A5	Green
WT #2	E1-E5	Green
R726S #1	B1-B5	White
R726S #2	F1-F5	White
K752T #1	C1-C5	White
K752T #2	A6, B6, C6, D7, E7	White
T827A #1	D1-D5	White
T827A #2	B7, C7, D6, E6, F6	White

Altogether, although our data are rather preliminary, they indicate that residues constituting the Rpb1 region of the backtrack site have an important role in ensuring cell viability and regulating various aspects of transcription, including transcription fidelity, elongation speed, and termination efficiency. Our single mutants do not exhibit drastically different phenotypes, which is consistent with the notion that they share the same proposed role of contacting backtracked RNA and contributing to the formation of the arrested Pol II complex. By contrast, the double and triple mutants, which significantly impair the binding of RNA to the backtrack site, were inviable, suggesting that backtrack site might have an essential role in ensuring cell viability. It might be possible that the backtrack site is required to prevent Pol II from backtracking even further, since the backtrack site supposedly forms stable interactions with the backtracked RNA. In such a scenario, if the backtrack site is significantly impaired, further backtracking can occur by many nucleotides, resulting in genomic instability and increased mutation frequency[2,19]. This model, while not yet established in the field of backtracking research, provides a potential explanation for the observed lethality of our double and triple mutants.

In contrast to our double and triple mutants, our single mutants were viable, but two of them (K752T and T827A) exhibited some diminished cell fitness relative to wild-type. With our data, we can construe how each single mutation affects the ability of Pol II to accomplish transcription efficiently, particularly in the context of its backtracking activity. One possibility for the variation in cell viability across all three mutants is their relative position to one other (Figure 8b) and their likelihood of contacting backtracked RNA. The T827A mutation contacts the first backtracked RNA nucleotide (nearest the

active site), the position where misincorporated nucleotides are intrinsically cleaved by Pol II. The frequency of backtracking by one nucleotide, therefore, may be higher than more extensive backtracking, especially when taking into consideration the importance of Pol II proofreading and high-fidelity transcription for the survival of the cell[2,14,15,21]. By contrast, the mutations K752T and R726S are deeper in the backtrack site, the former contacting the backbone between the second and third backtracked nucleotides and the latter contacting the fifth backtracked nucleotide. One hypothesis for the role of the backtrack site is to prevent further backtracking from occurring and causing genomic instability[2,19]. With this notion, we can infer that residues deeper in the backtrack site may not encounter backtracked RNA as frequently because residues closer to the active site will "catch" the RNA before it backtracks even further. This idea is in line with the observed mutant phenotypes since the R726S mutation is furthest from the active site and exhibits the highest cell fitness amongst all the single *rpb1* mutants in our study.

PERSPECTIVES AND FUTURE DIRECTIONS

The results of our study offer emerging insights into the importance of the Pol II backtrack site and the physiological role of backtracking. Backtracking has been implicated in various processes essential for gene regulation, such as modulating the rate of elongation, ensuring transcription fidelity, and promoting efficient termination[2]. While our present study characterizes backtracking in these different contexts, further experiments must be conducted to further support or refute our results and to elucidate the various other roles of backtracking. In particular, the experiments performed in this study should be replicated to ensure consistency in the observed phenotypes. We can also assess transcription fidelity, elongation speed, start site usage recognition directly by employing various *in vitro* assays that assess mutation rates, elongation rates, and start site usage. We can also use other assays that target other Pol II functions, such its intrinsic nuclease activity and processivity, to gain a better understanding of the functional importance of the backtrack site.

Although we achieved our goal of combining mutations in the Rpb1 region of the backtrack site, we were unable to characterize their effects on transcription fidelity, elongation speed, and termination efficiency *in vivo*. Given the lethality of our double and triple mutants on 5-FOA, we can also explore several reasons for their lack of growth. Rather than eliminating the wild-type *RPB1* gene (on pRP112) in strains harboring *rpb1* double and triple mutations (on pRP114), we might be able to assess how having both wild-type and mutant Pol II's influences cell viability and transcription relative to strains only containing wild-type *RPB1*. For example, without eliminating the wild-type *RPB1* gene, we can determine whether our double and triple mutants harbored any mutations

that would cause them to exhibit a wild-type growth phenotype by sequencing the *URA3* and *RPB1* genes in these strains and identifying any mutations that would confer resistance to 5-FOA or revert the strain to wild-type, respectively. In strains containing both wild-type and mutant polymerases, we can also examine mutation rates to investigate the notion that eliminating the interactions between backtracked RNA and the backtrack site leads to genomic instability.

It would be interesting to introduce our *rpb1* mutations into yeast strains containing TFIIS or Rpb9 deletions. TFIIS and Rpb9 are both essential for transcription fidelity, so placing our backtrack site mutations in these genetic backgrounds would allow us to better elucidate the role of backtracking in transcription fidelity. Furthermore, we can mutate other Rpb1 and Rpb2 residues in the backtrack site using similar techniques in this study in order to characterize their physiological importance and develop a more comprehensive understanding of the backtrack site and its effects on various aspects of transcription regulation.

Altogether, the importance of backtracking and the precise mechanisms underlying this conserved phenomenon remain elusive, but we were able to demonstrate that the backtrack site is essential for cell survival and for various aspects of transcription, such as fidelity, elongation speed, and termination efficiency *in vivo*. While our results provide some insights into the physiological role of backtracking, more research must be done to further elucidate the physiological role of the backtrack site and the importance of backtracking.

GLOSSARY[22]

active site: the region of an enzyme containing the catalytic residues identified with the binding and reaction of substrate(s).

allele: any of the forms of the same gene that occur at the same place on a homologous chromosome but differ in base sequence.

amino acid: any organic acid containing one or more amino substituents, a carboxyl group, and a side-chain specific to each amino acid. A table containing all known amino acids is provided in the Appendix (Figure A1).

annealing: the pairing of complementary sequences of single-stranded DNA or RNA to form a hybrid double-stranded nucleic acid molecule.

base pair: any of the possible pairings between two bases in opposing strands of double-stranded DNA or RNA molecules. Adenine forms a base pair with thymine (in DNA) or uracil (in RNA) and guanine with cytosine, hence the number of adenine residues equals the number of thymine (and/or uracil) residues while the number of guanine residues equals that of the cytosine residues.

catalysis: an increase in the rate of a chemical reaction brought about by a catalyst (generally an enzyme in molecular biology).

conformation: the characteristic 3-dimensional shape of a protein, including the secondary, supersecondary (motifs), tertiary (domains) and quaternary structure of the peptide chain.

culture: a collection of cells, tissue fragments, or an organ that is growing or being kept alive in or on a nutrient medium.

DNA: (contrast with RNA) one of the two main types of nucleic acids consisting of a long, unbranched macromolecule formed from one, or more commonly, two, strands of linked deoxyribonucleotides; self-replicating material which is present in nearly all living organisms serves as the carrier of genetic information.

enzyme: any naturally occurring or synthetic macromolecular substance composed of protein, that catalyzes, more or less specifically, one or more biochemical reactions.

eukaryotes: (contrast with prokaryotes) any organism whose cells contain a nucleus enclosed within membranes and undergo meiosis.

fidelity (of transcription): accuracy of transcription.

fitness: the survival value and the reproductive capability of a given genotype as compared with the average of the population or of other genotypes in the population.

gene: the functional unit of heredity consisting of DNA that codes for a protein that performs the functions associated with the phenotypic expression of the gene.

genome: the whole of the genetic information of an organism.

homologous: (of sequences of residues in encoded macromolecules) having the same or similar residues at corresponding positions; (of proteins from different species) having identical or similar functions.

homologous recombination: genetic recombination between two DNA molecules of identical or nearly identical sequence when contained in the same cell.

in vivo: Latin for “in life”; occurring or made to occur within a living organism.

in vitro: Latin for "in the glass"; performed or taking place in a test tube, culture dish, or elsewhere outside a living organism.

log phase: period when the number of cells doubles at a constant, exponential rate.

media: any nutrient medium that is designed to support the growth or maintenance of a culture. Culture media are typically prepared artificially and designed for a specific type of cell, tissue, or organ. They usually consist of a soft gel (so-called solid or semisolid medium) or a liquid, but occasionally they are rigid solids.

messenger RNA: a class of naturally occurring RNA molecules that carry the information embodied in the genes of DNA to the ribosomes, where they direct protein synthesis.

mutation (of a gene): permanent alteration in the DNA sequence that makes up a gene, such that the sequence differs from what is found in most individuals in a population.

nascent: just coming into existence; newly synthesized.

nucleotides: a compound consisting of a nucleoside linked to a phosphate group. Nucleotides form the basic structural unit of the nucleic acids DNA and RNA.

open reading frame: a sequence in DNA encoding the entire protein that has a length divisible by three and begins with a translation start codon (ATG) and ends at a stop codon.

optical density: a measure of the turbidity, or density, of liquid cultures of microorganisms.

PCR: polymerase chain reaction; method widely used in molecular biology to make many copies of a specific DNA segment.

phenotype: the totality of the observable functional and structural characteristics of an organism as determined by interaction of the genotype of the organism with the environment in which it exists.

plasmid: a genetic structure in a cell that can replicate independently of the chromosomes; typically a small circular DNA strand in the cytoplasm of a bacterium or protozoan.

polymerase: a general name for any transferase enzyme that catalyses the formation of biological polymers, especially of polynucleotides.

primer: a short oligonucleotide sequence that provides a starting point for DNA synthesis.

processivity: the ability of an enzyme ability to catalyze consecutive reactions without releasing its substrate.

proteome: the entire complement of proteins that is or can be expressed by a cell, tissue, or organism.

prokaryotes: (contrast with eukaryotes) any organism in which the genomic DNA is not enclosed by a nuclear membrane within the cells.

residues: any of the incorporated amino acid monomers in a peptide or protein.

restriction endonucleases: any of a group of enzymes, produced by bacteria, that cleave molecules of DNA internally at specific base sequences.

RNA: (contrast with DNA) one of the two main types of nucleic acids consisting of a long, unbranched macromolecule formed from ribonucleotides.

substrate: a substance that is acted upon, especially by an enzyme; a molecule or structure whose transformation is catalysed by an enzyme.

subunit: any polypeptide component within a protein.

transformation: the transfer of genetic information by means of naked extracellular DNA in bacteria.

translation: the process by which a particular sequence of bases in messenger RNA (mRNA) determines a sequence of amino acids in a polypeptide chain during protein synthesis.

X-ray crystallography: a technique for determining the three-dimensional structure of molecules, including complex biological macromolecules such as proteins and nucleic acids, by X-ray diffraction.

APPENDIX

Figure A1. The structures, abbreviations, and chemical properties of all amino acids (taken from Dan Cojocari, Department of Medical Biophysics, University of Toronto).

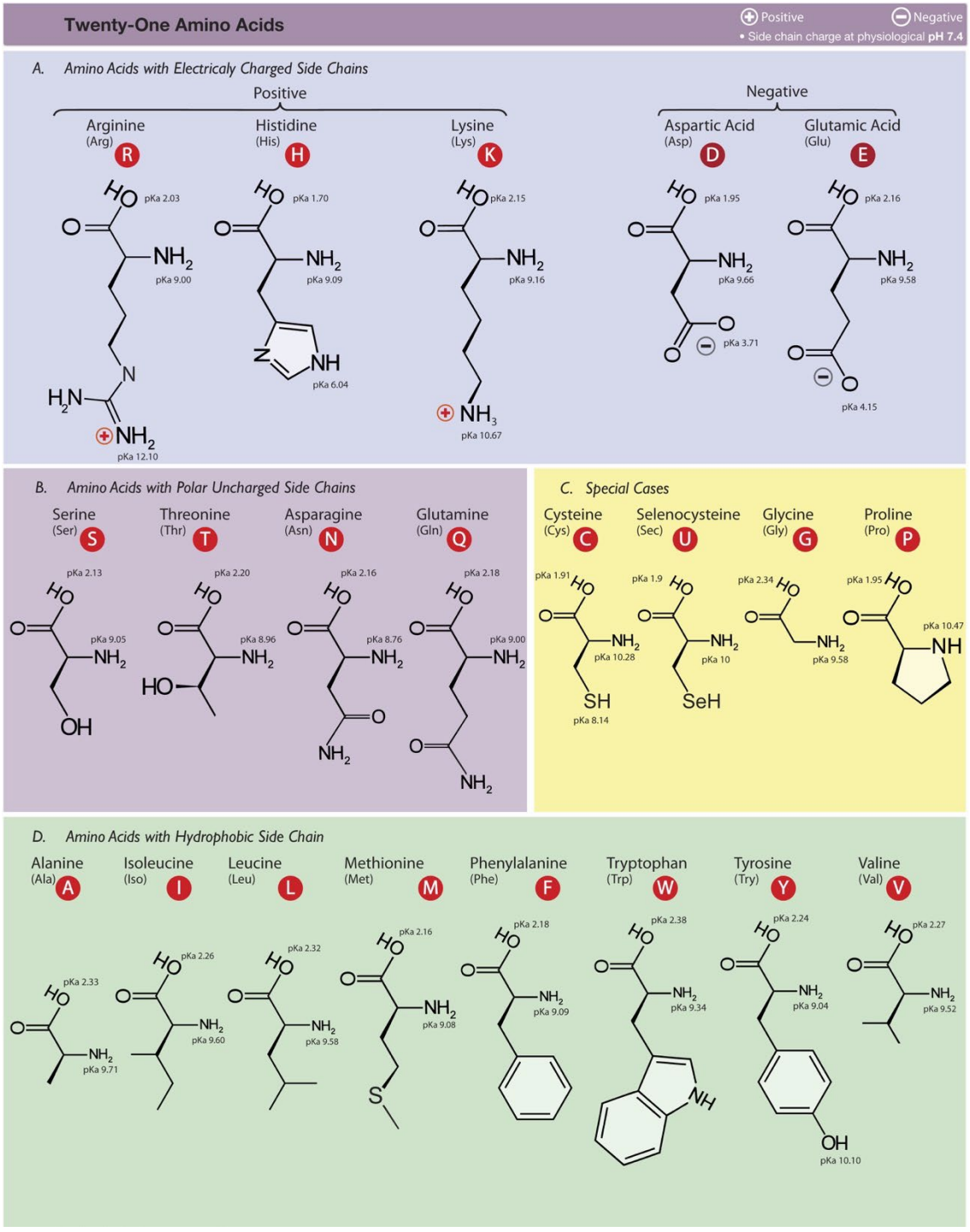
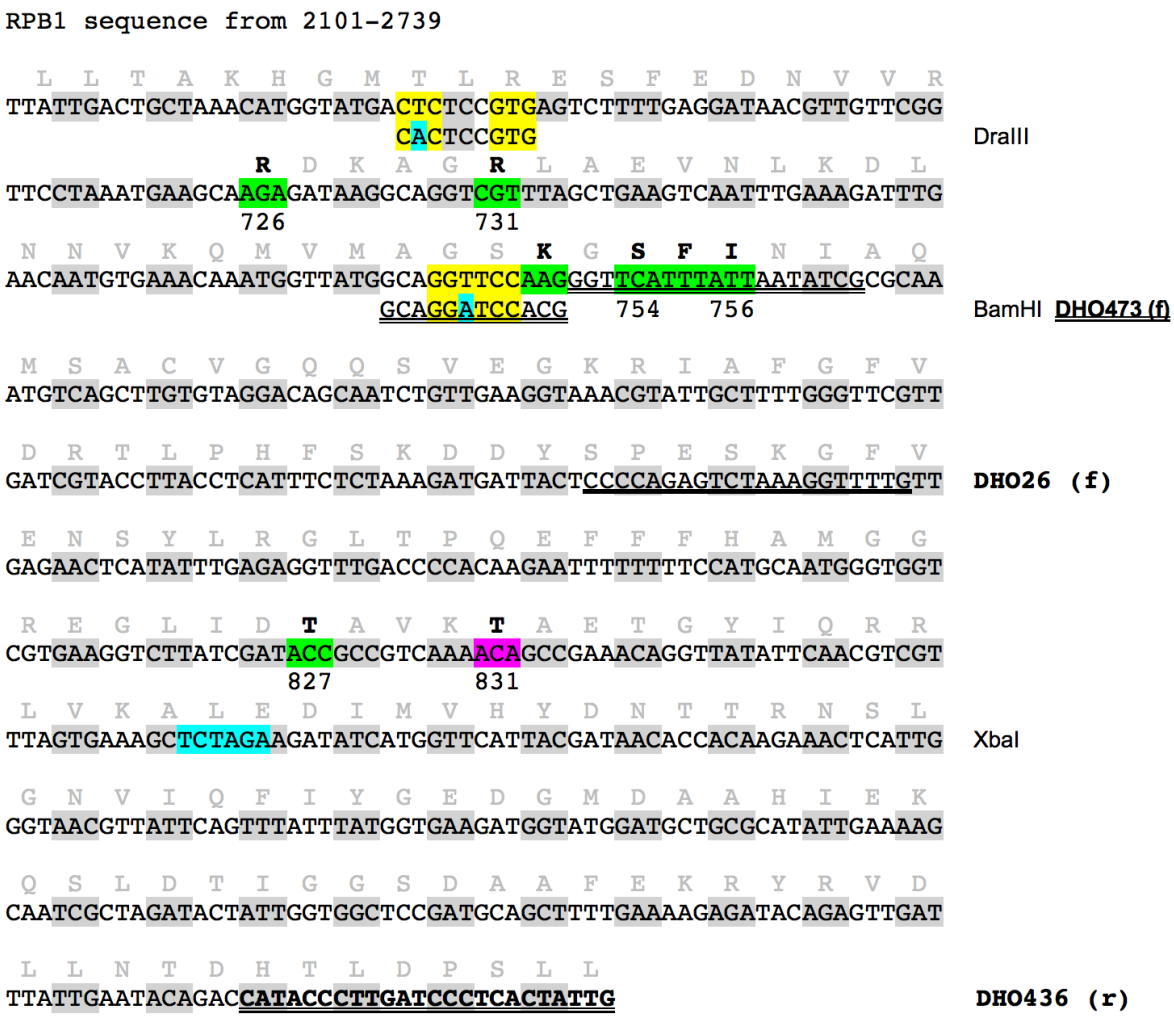


Figure A2. *RPB1* region of TOPO vector encoding residues (bolded) that are proposed to contact RNA in the arrested complex (codons for these residues are highlighted in green). The *Xba*I site and residues changed to create the *Dra*III and *Bam*HI sites are depicted (blue) together with wild-type bases that are part of the engineered restriction sites (yellow). The residue T831 (pink) is proposed to sense whether the 3' end of the DNA-RNA hybrid is frayed[23]. The forward primer DHO473 introduces the *Bam*HI site and introduces the K752T mutation. The forward primer DHO26 anneals upstream of the *Xba*I site, while the reverse primer DHO436 anneals downstream of the *Xba*I site.



BIBLIOGRAPHY

1. Cheung, A. C. M. & Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature, 471*(4):439-446.
2. Nudler, E. (2012). RNA polymerase backtracking in gene regulation and genome instability. *Cell*, *149*(7), 1438-45.
3. Herskowitz I. (1988). Life cycle of the budding yeast Saccharomyces cerevisiae. *Microbiological reviews*, *52*(4), 536-553.
4. Aref, R. (2014). Comparative analysis of repressor interaction with pleiotropic corepressors Sin3 and Cyc8 in the yeast Saccharomyces cerevisiae (PhD dissertation). University of Greifswald, Greifswald, Germany.
5. Department of Veterinary Disease Biology (2011). Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
6. Cramer, P., Bushnell, D. A., Kornberg R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science. 292*(5523):1863-76.
7. Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, *152*(6), 1237–1251.
8. Hahn S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nature structural & molecular biology*, *11*(5), 394–403.
9. Liu, X., Bushnell, D. A., Kornberg, R. D. (2012). RNA polymerase II transcription: structure and mechanism. *Biochimica et biophysica acta*, *1829*(1), 2–8.
10. Nudler E. (2009). RNA polymerase active center: the molecular engine of transcription. *Annual review of biochemistry*, *78*, 335–361.
11. Gómez-Herreros, F., De Miguel-Jiménez, L., Millán-Zambrano, G., Peñate, X., Delgado-Ramos, L., Muñoz-Centeno, M., & Chávez, S. (2012). One step back before moving forward: Regulation of transcription elongation by arrest and backtracking. *FEBS Letters, 586*(18), 2820-25.
12. Sydow J. F., Brueckner F., Cheung A.C., Damsma G.E., Dengl S., Lehmann E., Vassylyev D., Cramer P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell. 34*, 710–721.
13. Sigurdsson, S. Dirac-Svejstrup, A. B., Svejstrup, J. Q. (2010). Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Mol. Cell. 38*, 202– 210.
14. Nesser, N. K., Peterson, D. O., Hawley, D. K. (2006). RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *P. Natl. Acad. Sci. USA. 103*(9), 3268-73.
15. Jeon, C. & Agarwal, K. (1996). Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. *Proc. Natl. Acad. Sci. 93*(24), 13677-82.

16. Dymond, J. S. (2013). Chapter Twelve - *Saccharomyces Cerevisiae* Growth Media. *Methods in Enzymology*. *533*,191-2-4.
17. Lee, Y. (2010). Yeast RNA Polymerase II rpb1Mutations Conferring Termination Defects (Unpublished baccalaureate thesis). University of Oregon, Eugene, Oregon.
18. Severson, M. (2017). Creation and Characterization of Rpb1 Mutants in RNA Polymerase II in Yeast (Unpublished baccalaureate thesis). University of Oregon, Eugene, Oregon.
19. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in E. coli. *Cell. 146*(4):533-43.
20. Kaplan, C., Jin, H., Zhang, I., & Belyanin, A. (2012). Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo. *PLoS Genetics, 8*(4), E1002627.
21. Thomas, M. J., Platas, A. A. & Hawley, D. K. (1998) Transcriptional fidelity and proofreading by RNA polymerase II. *Cell*. *93*(4), 627-37.
22. Cammack, R., Atwood, T., Campbell, P., Parish, H., Smith, A., Vella, F., & Stirling, J. (Eds.) (2006). Oxford Dictionary of Biochemistry and Molecular Biology: Oxford University Press.
23. Da, L. T., Pardo-Avila, F., Xu, L., Silva, D. A., Zhang, L., Gao, X., Wang, D., Huang, X. (2016). Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nature communications*, *7*, 11244.