Title
Single Molecule and Synthetic Biology Studies of Transcription

Author
Zamft, Bradley Michael

Publication Date
2011

Peer reviewed|Thesis/dissertation

Single Molecule and Synthetic Biology Studies of Transcription

by

Bradley Michael Zamft

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos Bustamante, Chair
Professor Jasper Rine
Professor Jan Liphardt

Spring 2011

Single Molecule and Synthetic Biology Studies of Transcription

Abstract

Single Molecule and Synthetic Biology Studies of Transcription

by

Bradley Michael Zamft

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Carlos Bustamante, Chair

The horizons of biology are ever expanding, from the discernment of the detailed mechanisms of enzyme function, to the manipulation of the physiological processes of whole organisms and ecosystems. Single molecule studies allow for the characterization of the individual processes that comprise an enzyme's mechanochemical cycle. Through standardization and generalization of biological techniques, components, and knowledge, synthetic biology seeks to expand the scale of biological experiments and to usher in an age of biology as a true engineering science, in which those studying different hierarchical levels of sophistication need not start from the fundamental biochemical principles underlying all biological experiments. Here we report our findings on the processes governing transcription and its role in gene expression through the use of both single molecule and synthetic biology methods.

We have established a promoter-free, factor-free method of initiation of transcription by the mitochondrial RNA polymerase in *Saccharomyces cerevisiae*, Rpo41 through the use of synthetic oligonucleotides to imitate the hybridization geometry of Rpo41 during active transcription. Using this system, we have established that a sub-micromolar NTP concentration is appropriate for non-saturating transcriptional runoff assays. We have optimized the transcription buffer and found that 10 mM $MgCl_2$, 40 mM KCl, and 10 mM DTT are sufficient for robust transcription. Stability studies show that Rpo41 loses approximately 30% of its activity during each freeze-thaw cycle, and that the pre-formed elongation complex loses transcriptional activity with a half-life of 7.4±1.5 hr.

Through the use of optical trapping techniques, we have established a method to monitor the transcription of individual Rpo41 molecules in real time. This has allowed us to measure the kinetic rates of nucleotide incorporation by the enzyme: K_m = 22±13 μM^{-1} and v_{max} = 25±2.5 bp/s. Both of these rates are more similar to those of the main nuclear RNA polymerase in the same organism, RNA Polymerase II (Pol II) than to that of the T7 RNA polymerase, despite the fact that Rpo41 is a single-subunit RNA polymerase with

homology to those of the T-odd bacteriophage and no discernable homology to Pol II. Furthermore, like Pol II and the *E. coli* RNA polymerase, transcription by Rpo41 consists of periods of processive transcription interspersed with periods of pausing. We have also observed retrograde motion of Rpo41 during pauses, termed backtracking, a process that has not been reported in phage-like RNA polymerases.

We have performed single molecule assays of transcription by both Pol II and Rpo41 on templates of differing base pair composition and found that, in general, the characteristics of pausing are attenuated in templates of higher GC content. Specifically, the frequency of pausing is decreased in GC-rich templates, as is the average pause duration. The distribution of pause durations is correspondingly shifted to shorter pauses on GC-rich templates.

We discuss two mechanisms by which template composition may affect pausing: (1) movement of the backtracked transcription bubble is affected by differences in the base stacking energies from the disrupted/created DNA/DNA and RNA/DNA base pairs at the ends of the bubble, and (2) secondary structure of the nascent RNA upstream of the backtracked transcription bubble imposes an energetic barrier to its backward movement. We give *in silico* evidence that it is the latter mechanism. Incorporation of this secondary structure energy barrier (an "energy penalty") into a model of transcriptional pausing by backtracking allows for statistical fits of the mean pause densities, mean pause durations, and the distribution of pause durations for each enzyme on each template. Furthermore, incorporation of the energy penalty allows for fitting of the pause characteristics for a given enzyme using a single, enzyme specific hopping rate, k_0, that is independent of template, and a single, template dependent energy penalty term, ΔG_{RNA}, which is enzyme independent. For Rpo41, we find that k_0, the hopping rate of the backtracked enzyme along DNA without RNA secondary structure, is 5.4±1.8 s^{-1}, while it is 2.9±0.3 s^{-1} for Pol II. Furthermore, the average energy penalty due to the nascent RNA, ΔG_{RNA}, on the AT-rich template used in this study is 0.7±0.1 kT, while it is 0.8±0.1 kT for random DNA and 1.0±0.1 kT for GC-rich DNA.

In order to confirm that it is the secondary structure of the RNA that is the cause of the energy penalty, we performed the same single-molecule transcription assays in the presence of RNase A, an enzyme that digests unprotected RNA in both single-stranded and double-stranded form. The pausing characteristics of all traces on all templates in the presence of RNase A are statistically indistinguishable from those on AT-rich DNA without RNase, indicating that the RNase digested enough of the nascent RNA to disrupt any secondary structure. Protection of the 5' region of the nascent RNA by steric interactions between the polymerase and the RNase prevented full degradation of the RNA, and thus allowed for some backtracking. This strongly supports the new model, presented here, of modulation of transcriptional pausing by secondary structure of the nascent RNA.

In contrast to the detailed and isolated nature of single-molecule transcription, we also performed a synthetic biology project involving Rpo41. The intent of this project was to investigate the plausibility of the creation of a transcriptionally independent mitochondrion, and by extension a minimal cell, by movement of the mitochondrial transcriptional machinery from the nuclear to the mitochondrial genome. Thus we

performed *in vivo* mitochondrial transformation of yeast cells with a synthetic construct containing the gene encoding for Rpo41. We report that we have successfully integrated said synthetic gene into the mitochondrial genome, and have seen its expression to the transcriptional level. Furthermore, we are fairly confident that the full, intact mRNA of the synthetic gene is being created within the mitochondrial matrix.

We have not been able to detect expression of the protein product of the integrated synthetic construct, nor have we been able to isolate a strain that exhibits its expression in the absence of the wild-type, nuclear copy. Because the length of Rpo41 is longer than any other protein synthesized within the mitochondrial organelle, we have begun experiments to determine the maximal polypeptide length able to be translated by the mitochondrial ribosome and associated cofactors.

To my parents, who could not have supported me more,
my grandfather, who gave me science,
and Aiwa, whose love made even the lowest days grand

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

5FOA	5-Fluoroorotic acid
AA	Amino acid
bp	Base-pair
Da	Dalton, equivalent to 1 gram/mole.
ΔG_{RNA}	Energy penalty to backward backtracking
ΔG_{SS}	Energy of 30 bp of secondary structure behind transcribing polymerase
DNA	Deoxyribonucleic acid
EMS	Ethane methyl sulfonate
F	Force
k	Boltzmann constant
kB	Kilobase pairs (1000 bp)
k_b	Backward hopping rate during backtracking
kDa	Kilodalton (1000 Da)
k_e	Rate of elongation
k_f	Forward hopping rate during backtracking
K_m	Michaelis-Menten rate for nucleotide addition during transcriptional elongation
kT	Thermal energy
LHON	Leber's hereditary optic neuropathy
MELAS	Myopathy, encephalopathy, lactic acidosis, and stroke-like symptom
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA (the mitochondrial genome)
NARP	Neuropathy, ataxia, retinitis pigmentosa
NTPs	Nucleotide triphosphates
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
Pol II	RNA Polymerase II of *Saccharomyces cerevisiae*
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPM	Revolutions per minute
Rpo41	Mitochondrial RNA polymerase *Saccharomyces cerevisiae*
RT	Room temperature
RT-PCR	Reverse-transcriptase polymerase chain reaction

SEM	Standard error of the mean
STD	Standard deviation of the mean
T	Temperature
tRNA	Transfer RNA
UTR	Untranslated region
v_{max}	Pause-free velocity at saturating NTPs
v_{np}	Pause-free velocity
WLC	Worm like chain
WT	Wild-type

IT TAKES A VILLAGE: ACKNOWLEDGMENTS

Getting this Ph.D. has been the hardest thing I've done in my life this far. It is that simple. Perhaps this speaks to the gifted life I have had: rather than complain about my child's distended stomach, or long days on the unemployment line, or the pain of, say, a severe case of gout, I can firmly say that being paid (little) to think and to explore for nine (long!) years has been the challenge of my life. But it is simply the truth. Furthermore, as the title of this section shows, I did not even do this Ph.D. on my own. Rather, I *could not* have done it on my own. The age-old saying that it takes a village to raise a child holds true when raising a scientific child: it takes dozens to turn a graduate student into a Ph.D.

It would be a shame, then, to not whole-heartedly and sincerely thank those who helped me build this dissertation. It would be a shame to stifle my gratitude for space or time considerations; I owe you more. Thus I unapologetically present below, at length, my heartfelt thanks to all those who helped me in this dissertation and the underlying research it presents.

Of course I could not have done this work without the support of my advisor, Carlos Bustamante. Sure, he paid me, and for that I am thankful, but he did so much more.

One day I entered his office for an 8am meeting, frazzled and fried. I had been working on one of those "trivial" steps in science that should take a few days but sometimes ends up taking months, years, or never. I sat down defeated, worried that he would be disappointed in me, concerned that he would think it my fault, or a fault of not working hard enough, even though nothing could be further from the truth.

"Brad," he said, "go home, have a beer, and go to sleep."

Any advisor could find the money to pay me to get a Ph.D., but the confidence Carlos had in me, the support he gave to me, was invaluable and essential to my success. Graduate school is often lonely and frustrating, feelings which could be exacerbated by an advisor. In contrast, Carlos' support made my time at Berkeley easier.

It is no secret, however, that the vast majority of hours doing graduate research occur not in the presence of the advisor, but rather by working with others in the lab. Once again, I was lucky to work in the Bustamante lab, one which was filled with helpful and friendly people. I thank you all, collectively, for the welcoming atmosphere that made it easy to come to lab early and to stay late.

There were a few people, however, who went above and beyond in their kindness and assistance. They have been canonized in my mind, for their charity with their time and effort was simply extraordinary. There was St. Lacramioara Bintu and St. Toyotaka Ishibashi, who, overwhelmed with the successes and stresses of their own projects and careers, did not laugh or shout when they turned around from their computers one day to find me, a physicist who had "gone rogue" to play in strict biology for a number of years, asking them to participate (i.e. help) in the biophysical characterization of another RNA polymerase. They did not dismiss me outright, though they admit now that initially they were skeptical and not completely devoid of dread when confronting the notion of beginning an entirely new project. Instead, they embraced me and my project. What has

resulted was a number of late nights, early mornings, fascinating conversations and novel scientific insights. I have learned so much about science and about how to do science from them. The work presented of the second and third chapters of this dissertation would simply not exist without their contributions.

The work of the fourth chapter would not exist either, were it not for the help of Anton Vila-Sanjurjo. Anton and I started the synthetic biology branch of the Bustamante lab together. We set up the laboratory space, purchased all of the equipment, and performed all of the initial experiments. Anton and Carlos (with my input) effectively established the core theoretical underpinnings of the project, while Anton and I worked out the experimental details together. He and I have gone through all of the ups and downs of establishing an ambitious project *de novo*; for his camaraderie and support, I thank him.

The synthetic biology division of the Bustamante lab was and is the home of St. Alyssa Rosenbloom, a graduate student that joined the project three years after its inception. As the only two graduate students working in that space, we helped each other through the frustrations and stresses of attempting to get a Ph.D. We even managed to have a few laughs from time to time. She has been a dear friend and confidant of mine during my time at Berkeley.

Jean-Charles Blouzard has taken over my project to create a transcriptionally-independent mitochondrion. He has provided some help with the results presented in this dissertation, specifically those involving RT-PCR. I wish him luck and godspeed. Likewise, Mariana Leguia was a postdoc in the synthetic biology division of the Bustamante lab. She taught me much, especially that the added expense and time of making your own stock of something far outweighs the time lost and expense paid of a failed experiment. Javier Ceja-Navarro, Celina Vila-Sanjurjo, and Lourdes Dominguez-Gerpe contributed the atmosphere within the lab as well.

It takes a village to make a graduate student, not just a family. There have been many scientists outside of the Bustamante lab that have made this work possible. First and foremost is Professor Jasper Rine, who has given me so much advice on yeast biology, the scientific method, and the politics of academia. He has also showed an astounding amount of generosity with the resources of his own lab, specifically the brilliant, friendly, and helpful people that comprise it. Laura Lombardi, Erin Osborne, and Jen Gallagher have given much of their time and thought to the synthetic biology portion of this dissertation. Dr. Gallagher was especially generous in this regard, having performed the northern blotting technique with me.

Other Berkeley professors who have helped with this dissertation include John Dueber, who has been as much a friend as a source of technical advice; Bob Jacobson, who gave down-to-earth advice; and Jan Liphardt, who, as the chair of my qualifying exam committee and member of my thesis committee, has made the two most stressful components of graduate school much less so. Professor Tomas Fox and his talented technician Christine Butler opened their lab to me and taught me the mitochondrial transformation procedure; he, along with Professors Peter Thorsness of University of Wyoming, Pawel Golik of the University of Warsaw, and Gerald Shadel of Yale University have provided crucial advice about yeast mitochondrial biology. Thank you all for your time and help.

There have been a slew of undergraduates who have taken time away from their studies and social lives to help us in our research. First and foremost is Joseph Marlin, who was a friend, hard worker, and great assistant. Soon to be an M.D./Ph.D., I offer him my thanks and congratulations. Other undergraduate students include Courtney Lane, Sandy Lao, DJ Cummings, Milton To, Anaar Eastoak-Siletz, Natalie Kolawa, Bernadette Hapsari, Richard Wang, Andrew Chen, and Natalie Benadum.

Thank you to those who took the time to read this manuscript and strengthen it with your comments. Chapters 2&3 have been extensively reviewed by Lacra and Toyotaka; Chapter 4 has been vetted by Alyssa, Jean-Charles, Jen Gallagher of the Rine Lab, and Anton (multiple times).

Though the days in the lab were long, I did manage to spend some time out of it. How I spent that time was just as crucial to my success in graduate school as my time on the bench. My friends at the University of California Martial Arts Program (UCMAP) – an astonishing collection of smart, funny, and friendly people – gave me a place to blow off steam after defeats and successes in science alike. No matter how tired, sad, depressed, or angry I was, I always knew that at 6pm I had to get to the gym. Miraculously, the moment my bare feet hit the mat, all the ill feelings melted away, and life was simply good. UCMAP was my haven during my time at Berkeley, and it is one of the things that I miss most. Thank you Fil Vrnak for having a thick skull and no sense of pain, Steve Brown for making us run up Cedar, and too many others to mention for creating the most welcoming atmosphere in any dojo in the world. I cannot overemphasize the importance you have had on my life and on my graduate studies.

The director of the UCMAP Yongmudo program, Dr. Norman Link, has taught me as much as any professor or colleague in Berkeley. In addition to teaching me 90% of my martial arts knowledge, he taught me about the balance between personal ambition and family, about how to be a father, a scientist, and a social citizen at the same time. He taught me self-defense, but also that 50 is the new 18, that you only get old when you want to, and that fun doesn't end when your hair falls out.

Carlos has a poignant story of when he knew he loved science. As a child, his father had given him a microscope, and he discovered for himself that the presence of soap residue on his slide caused his bacteria to lyse. He concluded that that their membrane must have been made up of lipids. Decades later he is still playing with microscopes, albeit a bit more expensive.

I do not have a particular moment in which I remember learning I loved science. Instead, I have a series of memories of my grandfather, telling of his times as an engineer at the Cooper Union and on the railroad, giving me chemistry books copyright 1934, and explaining the world through observation, intuition, and calculation. Everyone else in my family studied a social science; it is only my grandfather and I with degrees in science and/or engineering. Thank you, grandpa, for sharing.

There is, of course, no more important of an influence on my life than my parents. Yes, they raised me. But they did so much more. I am so lucky to have parents that nictitate at

the thought of my coming home, that love nothing more than to feed me and help me in any way that they can. I am a scientist today instead of in jail because my father let it be clearly known that education was the most important thing. I will be a Ph.D. because they pushed me on the hardest days, and housed and fed me while I wrote this dissertation. Their love can be considered the very definition of unconditional and boundless. I love you too.

And finally there is Aiwa, the love of my life. Each day as I left lab, be it at 5pm or 5am, I would realize that walking home to her made it the best part of my day. Her smile as she opened the door for me made each day precious. Thank you for listening to countless hours of whining, for laughing at my jokes, and for simply being you.

Graduate school is hard. It is harder than running a marathon, or fasting for days. It is the ultimate exercise in humility and futility, as one tries to wrestle mother nature for her secrets and academia for the right to be called a colleague. It takes a village to make a Ph.D. out of a graduate student, to bathe him when he plays in the mud, to shade him from the sun and shelter him from the rain. I could have asked for none better. Thank you.

Chapter 1: INTRODUCTION

There is a revolution well under way in the field of biology, from the descriptive, observational science of Linnaeus' taxonomy, to the predictive, causal, and explanatory sciences of, for example, biochemistry and molecular biology. Biology has expanded, now, from merely categorizing nature, to understanding the underlying principles of its organization.

The central principle of molecular biology, first articulated by Francis Crick in 1958[1], is that DNA encodes for RNA, which in turn encodes for protein. This seemingly simple statement, commonly referred to as the "central dogma of molecular biology" forms the core of the revolution within biology (note: like any good dogma, the central dogma is violated in some cases. These violations are beyond the scope of this dissertation). It encapsulates the entire paradigm of the chemical nature of life. DNA, the archival master plans of the cell, is protected from damage and mutation in the nucleus. This DNA is copied into a working copy of the plans – a similar but less stable molecule called RNA – through the process known as transcription. This working copy is then translated to a completely different polymer - protein - through the process aptly known as translation. It is this polymer - made of amino acids that, when linked together, can take on three dimensional structures with physical, mechanical, and chemical properties - that makes up much of the structure and machinery of living systems.

Within the central dogma lies the intricacies of life, including the rules that govern it: chemistry and physics. The beginning of the revolution purportedly began in 1828, when Friedrich Wöhler synthesized urea[2], showing for the first time that the building blocks of life – organic compounds – could be synthesized in the lab, just as chemists had been doing for inorganic compounds since the times of alchemy. This was not lost on the scientific establishment, as is evidenced by Wöhler's obituary by chemist August Wilhelm von Hofmann:

> *The synthesis of urea is in the true sense of the word an epoch-making discovery. With it was opened to investigation a new field of which chemists did not hesitate to take possession. The present generation, which daily reaps rich harvest on this field conquered for it by Wöhler, can transport itself only with difficulty to those remote times when the formation of an organic compound in the body of the plant or animal seemed to be brought about in a mysterious manner by the vital force, and it is able therefore hardly to realize the sensation produced upon their minds by the building up of urea from its elements… [it] removed at a single blow the artificial barrier which had been raised between organic and inorganic chemistry*[3]*.*

This revolution in biochemistry continued with the discovery of the first enzyme[4], and moved to greater scales and more sophisticated assays, such as *in vitro* transcription, translation, replication, and even biogenesis. These experiments were often done as chemists in the decades before them did them – in bulk, with millions or billions of

molecules, in solution. The difference, of course, was that the molecules themselves were larger and more complex than any "pure" chemist had dealt with, but the knowledge gained was similar: time and ensemble averages of behavior. These statistical measurements, such a reaction rates ("kinetics") and equilibrium concentrations ("thermodynamics") have served the scientific community very well. For example, we can now, to a high level of certainty, calculate how many molecules of oxygen a cell uses to convert sugar into energy[5], or how fast *on average* a motor protein can carry its cargo[6]. The contribution biochemistry has given to our understanding of the central dogma is unparalleled. It has allowed us to understand which proteins participate in turning DNA into RNA, and RNA into protein, as well as the rates of such processes. But it has fallen short of fully answering the fundamental mechanisms by which these processes occur; biochemistry, genetics, and molecular biology can give many hints to the processes of life, but the most intricate of details, down to single molecule interactions between the biological players, is reserved for biophysics.

There is a second revolution going on in biology, one that is much more recent, and at a completely different scale than biophysics. This field of "synthetic biology" was purportedly named by the polish geneticist Waclaw Szybalski

> *Let me now comment on the question "what next." Up to now we are working on the descriptive phase of molecular biology…. But the real challenge will start when we enter the synthetic biology phase of research in our field. We will then devise new control elements and add these new modules to the existing genomes or build up wholly new genomes. This would be a field with the unlimited expansion potential and hardly any limitations to building "new better control circuits" and…. finally other "synthetic" organisms, like a "new better mouse…." I am not concerned that we will run out exciting and novel ideas...*[7].

While biophysicists aim to reduce biological processes down to the single molecules that perform them and the fundamental forces that govern them, synthetic biologists aim to design, characterize, and formalize biology from the molecular to organismal level, and wrap up all of those processes into proverbial "black boxes" of varying complexity. Other scientists, then, can take those black boxes for granted, much as a computer engineer takes for granted circuit boards, a circuit board engineer takes for granted microchips, and a microchip engineer takes for granted the solid state physics upon which transistors and microchips are built. Each person in the chain of design can ignore the intricacies of he who is below (or above) him in the chain, and merely has to rely on the specifications of the parts given him from below, or define such specifications for the parts he must give above.

Similarly, the emerging field of synthetic biology aims to characterize and formalize the rather mature tools of molecular biology, biochemistry, and metabolic engineering, and turn them into a true engineering science, in which the user need not "reinvent the wheel," or characterize every enzyme in his pathway, in order to perform a novel biological function.

The purpose of this dissertation, then, is to span the spectrum between these two revolutions. In Chapter 2, we characterize the single enzyme named Rpo41 – a machine of one molecule – that is responsible for copying DNA into RNA within the mitochondrial organelle. To do this we act as biophysicists, utilizing the momentum carried by light, working with lasers and optics, to pull on Rpo41 while it is transcribing, and to see how it pulls back. Using the physical quantities of velocity and force, we can in turn gain insight into the fluctuations and motions the molecule undergoes while doing its job.

In Chapter 3, we utilize the techniques discussed in Chapter 2 to address a more fundamental question regarding transcription in general. We extend single molecule transcription assays by Rpo41 to multiple templates, and do the same for its nuclear counterpart – RNA polymerase II. In doing so we discover a fundamental physiochemical principle that imparts a sequence dependence on the pausing of transcription: the newly formed RNA molecule itself forms structures that prevent the polymerase from moving backward, aiding in its own synthesis.

In Chapter 4 we shift from studying transcription by striving for ever defined detail and ever increasing resolution – "zooming in" – to understanding it as a tool within a greater system – "zooming out." We attempt to understand the broad principles of gene expression, mitochondrial transcription *in vivo*, and the fundamental genetic principles underlying the living state. We explore the possibility of taking established "black boxes" within the field of mitochondrial and yeast biology, and use them to "tinker" with mitochondrial transcription and, indeed, mitochondrial biogenesis as a whole. We move the gene *RPO41*, which is normally present in the nucleus of the yeast *Saccharomyces cerevisiae* and whose product (the protein Rpo41) is normally transported into the mitochondria, to the mitochondrial genome, upon which it normally performs transcription. We do this as synthetic biologists in that we consider this the first step in the creation of a synthetic cell (from the mitochondria), one which is highly specified by the fact that it is rationally designed. Such a high level of specification will aid in the projects of other synthetic biologists, who often grapple with the noise of their host organism's metabolic processes. Furthermore, it will lead to the elucidation of the very many processes we will need to engineer inside the mitochondria, will help answer the oft-asked question of what minimally defines life, will aid metabolic engineers in designing pathways, and will be the first step towards the creation of whole synthetic organisms - Szybalski's "new better mouse."

Chapter 2: Characterization of the Yeast Mitochondrial RNA Polymerase, Rpo41

Introduction

The first step of the central dogma of molecular biology, and in some sense the first step in what makes beings out of DNA, is transcription. This is the duplication, of one polymer – poly-deoxyribonucleic acid (DNA) – into another – poly-ribonucleic acid (RNA). RNA polymerase (RNAP) is the core enzyme that, along with molecular assistants, performs this complicated task.

RNA polymerases have been isolated throughout nature, present in disparate forms in viruses, bacteria, mammalian cells, and everything in between. While the specific molecular compositions and configurations of these polymerases differ, as do the mechanisms regulating their behavior, the ultimate process they perform is the same. Thus detailed characterization of these machines at the molecular level has provided insight into the fundamental principles underlying not just transcription, but also regulation of gene expression and, in turn, the spectacular process of turning physics into chemistry into life.

NTP-dependent DNA copying fractions were first isolated from rat liver nuclei in 1959[8]. This led to an era of transcription research in which RNA polymerases from all phyla were characterized by both biological and crystallographic means. The necessary and sufficient elements of various transcriptional systems were defined, their kinetic properties discerned, and their structures revealed. These assays became more and more sophisticated as instrumentation and technology kept pace with the hard work, imagination, and understanding of those in the field.

The 1990's saw the use of single molecule techniques to study various biological processes, including transcription[9,10]. This allowed the field to move beyond ensemble averages and to explore characteristics of transcription that were previously masked in bulk experiments by asynchrony and variation. The *Escherichia coli* RNAP[9,11,12], RNA polymerase II from *Saccharomyces cerevisiae* (Pol II)[13-15], and to a lesser extent the RNAP from bacteriophage T7[16-18] have all been studied using single molecule techniques. This has led to an understanding of these enzymes as molecular motors, and a characterization of them likewise.

Here we aim to characterize yet another RNAP – the mitochondrial RNAP of *S. cerevisiae,* Rpo41. It is present in the same cell as the well-characterized Pol II, and thus must play a role in the coordination between mitochondrial processes and the rest of the cell. Yet it transcribes a genome that is much higher in AT content. Furthermore, it bears no homology to its nuclear neighbor, nor does it resemble that of the α-proteobacterial progenitor of the mitochondrial organelle. Rather, Rpo41 has homology to the T-odd bacteriophage RNAPs[19]. These single-subunit RNA polymerases (ssRNAPs) are evolutionarily distinct from the multi-subunit ones extensively characterized by single molecule methods[20].

This schizophrenic nature of Rpo41 – it is a phage-like enzyme present in an organelle of α-proteobacterial origin within the eukaryotic cell – begs the question: how does Rpo41

transcribe? Does the homology of the primary amino acid sequence of the enzyme with that of the phage dictate that it transcribe in the same manner, or, has it evolved over the millennia to match the transcriptional characteristics of Pol II?

The study, then, of transcription by Rpo41 at the single-molecule level, is really the study of molecular motors, transcription, and evolution. Furthermore, because the single molecule analysis of T7 RNAP has only been done at low resolution, the study of Rpo41 could be considered the first higher-resolution study of transcription by a single-subunit RNAP. Thus the characterization of Rpo41 provides a key missing piece in the puzzle of understanding transcription amongst the phyla.

TRANSCRIPTION

Transcription, in general, is composed of three stages: initiation, elongation, and termination. While the particular details of each of these steps differs by organism (and organelle), we give here generalizations of each process.

Initiation, as its name implies, is the process in which transcription begins. It involves the binding and unwrapping of the template DNA at a promoter sequence by the core RNA polymerase and/or transcription initiation factors[12,21-26]. Polymerization of RNA from individual substrate NTPs begins after alignment of the catalytic center of the RNAP with the first transcribed nucleotide of the promoter (the +1 nucleotide). This form of polymerization is often times abortive, resulting in a number of short (2-8 bp) transcripts for each fully transcribed primary transcript[27]. Single molecule studies have discerned that "scrunching" of the DNA occurs within the polymerase during initiation. The stored energy in the scrunched DNA has been proposed to facilitate promoter escape, and the transition from initiation to elongation[28,29].

Unlike initiation, elongation is a processive and relatively robust process, in which the RNA polymerase engages in RNA synthesis until the end of the primary transcript, i.e. termination[18,30,31]. This is necessary because, unlike DNA replication, in which later DNA polymerases can re-start polymerization where previous ones left off, transcribed RNA is no longer associated with the template after RNAP dissociation. This is not to say that elongation is incessant. Both bulk and single molecule optical trapping experiments on the *E. coli* RNA polymerase and the nuclear RNAP responsible for mRNA transcription in yeast (Pol II) have shown that elongation is comprised of translocation interspersed with distinct pausing events[9,13]. The kinetics of pausing have further revealed that pausing kinetically competes with translocation, indicating that it is an off-pathway process. Further details of pausing during elongation are given in Chapter 3.

Polymerization of RNA occurs through the nucleotide incorporation cycle. We begin with a RNAP in the post-translocated state, in which the 3' end of the RNA is 1 bp upstream of the NTP binding site of the polymerase. A substrate NTP can bind to a region in the active site of the post-translocated RNAP, upon which time the formation of phosphodiester bond between the 3'end of the RNA chain and the incoming NTP can occur. This results in the release of one molecule of pyrophosphate and the addition of one nucleotide to the RNA. The RNAP is then considered to be in the pre-translocated state, in which the 3' end of the RNA obstructs the NTP binding site and prevents further substrate from entering the core of the polymerase. Brownian-driven movement of the polymerase

one base pair – translocation – clears the 3' end of the RNA from the NTP binding site, resulting in a post-translocated RNAP, and completing the nucleotide addition cycle[32].

Crystallographic[32] and biochemical[33] assays of T7 RNAP has found that pyrophosphate stabilizes the pre-translocated state, and thus translocation likely happens after pyrophosphate release[32]. The biochemical assays also found that the pre- and post-translocated states are in positional equilibrium, indicating that T7 RNAP translocates through a "Brownian ratchet" mechanism. In this model, the RNAP fluctuates between the pre- and post- translocated states until release of pyrophosphate, which results in conformational changes of the RNAP that energetically favor the post-translocated state. This is in contrast to a "power-stroke" model, in which the free energy released through catalysis results in conformational changes that causally result in translocation. Single molecule studies have found that RNAP sensitivity to opposing force is increased at lower NTP concentrations, excluding all but the most complicated power stroke models and strengthening the Brownian ratchet model further[16,34]. Similar studies have shown that Pol II[35] and the *E. coli* RNAP[36] elongate using the same type of rectified ratchet mechanism.

Termination is the mechanism by which transcription ends, and has a crucial role not only in defining the ends of mRNAs, but in regulation as well. There are at least two different mechanisms for termination in *E. coli*, one involving the *trans*-acting termination factor Rho[37], and the other resulting from sequence-specific conformational changes of the nascent RNA within the transcribing RNAP. Both share a similar proposed mechanism for transcriptional termination: perturbation of the DNA/RNA contacts within the core of the transcribing RNAP. In Rho-dependent transcriptional termination this is done by ATP-dependent hydrolysis and translocation by Rho, which binds a C-rich region of the nascent RNA near the exit channel of the transcribing RNA polymerase[29]. In Rho-independent termination, a (relatively unstable) sequence of poly-U occupies the hybrid within the polymerase, while a hairpin forms upstream. This destabilizes the hybrid, and leads to dissociation of the RNAP[38]. The Rho-independent termination mechanism is also present during transcription by T7 RNAP, and is so similar to that of *E. coli* RNAP that some bacterial terminators are recognized by the T7 RNAP[39].

Single molecule studies of termination of *E. coli* RNAP have shown that forces on the nascent RNA have little effect on termination. Thus any forces generated by Rho or through the thermodynamics of hairpin formation must be aided by or mediated through allosteric interactions[40].

Transcriptional termination in eukaryotes, like that in prokaryotes, is generally initiated by *cis*-acting elements near the 3' end of the nascent transcript. For mRNAs, this includes the poly-adenylation signal, which is recognized by a cleavage/polyadenylation complex consisting of the polyadenylation factor CPF, and cleavage factors CFIA and CFIB[41].

The details of the above discussion on transcription in general have mainly resulted from studies using Pol II, *E. coli* RNAP, and T7 RNAP. Transcription and gene expression is much less well characterized for Rpo41. We discuss the specific processes which govern gene expression in yeast mitochondria, with a focus on transcription, below.

GENE EXPRESSION IN YEAST MITOCHONDRIA

The enzyme responsible for transcription in yeast mitochondria is the Rpo41/Mtf1 heterodimer. Both proteins of the complex are encoded in the nucleus, translated by cytoplasmic ribosomes, and post-translationally transported to the mitochondrial matrix. Rpo41 (153 kDa) is the catalytically active RNA polymerase. It is a single-subunit enzyme, and has high homology not to the RNA polymerases in the nucleus, but rather to those in the T-odd bacteriophage family[19]. Rpo41 is capable of initiation of transcription by itself, albeit in a nonspecific manner, as well as specifically from a promoter in which as few as three nucleotides have been pre-melted by introducing a lack-of-homology on the non-template strand, a so called "bubble" promoter[42,43]. A control bubble without the consensus promoter sequence resulted in no transcription. In this chapter, we describe a novel method to initiate transcription by Rpo41 alone in a promoter-independent manner.

The accessory factor to Rpo41, Mtf1 (43 kDa, also known as sc-mtTFB) has often been likened to the bacterial sigma factors, due to the fact that it interacts with both the core RNA polymerase and the promoter DNA[22], unwinds the DNA to form an open complex[25,44], decreases nonspecific initiation[22], and dissociates from the complex during the transition to elongation[45]. However, crystallographic studies have shown that Mtf1 is structurally homologous not to bacterial sigma factors, but rather the rRNA methyltransferase ErmC' [46]. The fact that Rpo41 can initiate transcription from a pre-melted promoter without Mtf1 indicates that promoter specificity is intrinsic to Rpo41, while role of Mtf1 is to melt the promoter and stabilize the open promoter complex[43]. This further distinguishes Mtf1 from bacterial sigma factors, which comprise "most if not all" of the determinants for promoter recognition[47].

Transcriptional initiation in yeast consists of interactions between Rpo41 and the template DNA, Mtf1 and the template DNA, and Mtf1 and Rpo41. The interactions between Rpo41 and the template DNA provide all of the catalytic activity, as well as most of the promoter specificity. The presence of Mtf1 during transcription initiation attenuates aberrant initiation[22], which is purportedly mediated through the interactions between it and Rpo41. In contrast, Pol II alone consists of 12 subunits, and requires association with several other complexes, including five "general transcription factors" and the "mediator" complex, consisting of at least 26 proteins[48,49].

Transcriptional elongation, on the other hand, is a simpler process in that there are only two elements: Rpo41 and the template DNA. Mtf1 dissociates from the complex somewhere between nucleotide four and thirteen[45], after which time Rpo41 proceeds to processive elongation alone. To date there have been no published studies investigating the properties of transcriptional elongation by Rpo41. Similarly, the mechanism of transcriptional termination in yeast mitochondria is currently unknown[50].

Unlike most other systems, most regulation of expression in yeast mitochondria is not done at the transcriptional level, but rather at the translational level, as is discussed in Chapter 4. Nonetheless, there are a few known mechanisms of transcriptional regulation within the mitochondrial organelle (the reason for this transcriptional regulation is unclear), examples of which we give here. Differences in promoter sequences can result in as much as a 20-fold difference in promoter strength[51,52], as can elements upstream of the promoter[53]. Downstream elements – especially the first and second nucleotides after the transcriptional start site – can modulate transcriptional activity as well. Most primary

transcripts in mitochondria contain multiple genes – they are polycistronic – and require post-transcriptional processing for the formation of individual mRNAs (discussed below)[54]. Positions of particular genes within these polycistronic transcripts result in differences in mRNA levels, with promoter-distal sequences being less abundant[52]. This indicates a mechanism of attenuation of transcriptional elongation and/or selective degradation of these distal transcripts[54].

Finally, environmental conditions affect transcript levels via nucleo-mitochondrial interactions, as well as mechanisms internal to the mitochondria. For example, the phenomena of glucose repression, in which steady-state levels of mitochondrial RNAs are up to six-fold lower in strains grown on glucose than on galactose[54], has been indirectly linked to modulation of mitochondrial transcription via the nuclear Snf1 protein. Snf1 is involved in derepression of nuclear genes, but does not directly alter the abundance or properties of Rpo41 or Mtf1[55]. Rather, growth on glucose results in high rates of glycolysis, which in turn results in high levels of ATP. ATP inactivates Snf1, resulting in repression of many nuclear genes, including those in the respiratory chain, which in turn results in lower ATP concentrations in the mitochondrial matrix. This matrix ATP concentration is "sensed" by the initiating Rpo41/Mtf1 heterodimer (due to promoter-specific kinetics at the +1 nucleotide, which is always an A in mitochondria), and consequently results in gene-specific modulation of transcription[56,57].

There is a lack of correlation between the rates of transcription for particular transcripts and the abundance of their protein products[58], indicating that regulation of expression occurs post-transcriptionally, at the RNA processing and/or the translation steps. Implication of the RNA processing step, specifically the balance between transcription and RNA degradation, is strengthened by the fact that partial loss-of-function mutations in the mitochondrial degradosome can be rescued by partial loss-of-function mutations in Rpo41[59]. The mitochondrial degradosome, mtEXO, is a haloenzyme composed of the RNase Dss1 and the helicase Suv3[60]. It is the main ribonuclease involved in RNA turnover in mitochondria; it, along with others involved in mitochondrial RNA processing, are discussed in a recent review[50].

Regulation of translation in yeast mitochondria is much better understood than RNA processing. Translational activators for all mitochondrially-encoded integral proteins have been isolated and studied in recent years including, for example, Pet309, Pet111, Pet54, Pet122, and Pet494. These proteins recognize the 5' untranslated regions (UTRs) of specific mRNAs (*COX1* for Pet309, *COX2* for Pet111, and *COX3* for Pet54, Pet122, and Pet494), and recruit the mitochondrial ribosome for translation. Because they are located on the matrix side of the mitochondrial inner membrane, they serve to localize translation, an essential step in the co-translational insertion of integral proteins into the mitochondrial inner membrane[61]. Modulation of these activators results in a concomitant modulation the protein levels of their respective mRNA targets[62]. Furthermore, overexpression of Pet111, the *COX2* translational activator, *prevents* translation of *COX1*[63], indicating a more complicated cross-talk between translational activators.

The role of Rpo41 is not limited to transcription of primary transcripts; it has been demonstrated that deletion or mutation of the *RPO41* gene has deleterious effects on replication of the mitochondrial genome. It has been proposed that the first step of the

mitochondrial DNA (mtDNA) replication involves Rpo41 via the synthesis of an RNA-primer, which serves as anchor for the DNA polymerase[64-66], though this model has recently been challenged[50]. Deficiency of Rpo41 expression leads to mtDNA rearrangements and the loss of large portions of the mitochondrial genome[50,67-69].

A Brief introduction to Optical tweezers

While a large portion of the data presented in this chapter is the result of biochemical studies of *in vitro* transcription by Rpo41, the culmination of this chapter, as well as the entirety of the next, is the analysis of transcription by Rpo41 at the single molecule level. These experiments utilize optical trapping techniques as a means to hold on to single polymerases during transcription. There are a number of excellent reviews on the subject of optical trapping and the instruments used therein ("optical tweezers")[70-72]. Thus here we provide only the most general principles of optical trapping, as well as a description of the instrument used in our experiments.

Optical trapping experiments utilize focused laser light to impart forces on micron or sub-micron sized polystyrene beads. These forces result from the gradient of light intensity around the point of focus, which is diffracted through the bead. The spatial distribution of light intensity serves to impart differential diffraction forces depending on the location of the bead in the potential well ("trap"), which in turn leads to a net force toward the trap center[73]. These forces can be in the pico-Newton range, and the beads trapped with these forces can be localized with nanometer resolution[71].

Conjugation of biomolecules to these polystyrene beads serve as a linkage between the biological and physical (optical) components of the experimental setup. Traditional single molecule optical trapping studies have used a single optical trap, while a second attachment of the biological setup (the "tether") occurs via linkage to a cover slip or another bead that is held in place on the tip of a micropipette by suction[70].

More recent tweezers setups, including the one used in this study, hold the tether by optical traps at both ends[71]. This decouples the system from the sample chamber, greatly reducing noise and allowing for sub-nanometer resolution of the differential bead position.

In the succeeding chapter we begin a journey towards the understanding of Rpo41. We describe the establishment of robust *in vitro* assays of transcription, and the utilization of those assays in the initial characterization of the enzyme. We then move to the power of optical trapping to describe the mechanical properties of the polymerase, and conclude with the kinetics of nucleotide incorporation.

Results and Discussion

Purification of Biotinylated RPO41

Molecular cloning of the biotin-tagged RPO41 expression vector

The experimental setup for single-molecule optical trapping studies has conventionally utilized biotinylated forms of the protein under study. This allows for the attachment of the

protein to a streptavidin-coated polystyrene bead through biotin-streptavidin electrostatic interactions, the strongest noncovalent interaction found in nature[74]. The bead can in turn be trapped in a potential well formed by high intensity laser light (the "optical trap") – a prerequisite for single-molecule optical trap studies.

We used a BirA-based biotinylation protocol in order to biotinylate Rpo41. The BirA enzyme attaches a biotin to a specific lysine residue of acetyl-CoA carboxylase in *E. coli*[75], and has been shown to biotinylate other proteins containing a 13 amino-acid consensus sequence[76]. Thus the strategy for producing a biotinylated version of Rpo41 is to add the consensus sequence (the "biotinylation tag") to the Rpo41 coding sequence through molecular cloning, and express the protein in the presence of BirA and free biotin.

We inserted the coding sequence for a version the above described biotinylation tag (GLNDIFEAQKIEWHE, red is the site of biotinylation) into a pre-established Rpo41 expression vector (pProExHtb-RPO41[77]) in two ways: via ligation into the NarI and StuI sites, to yield plasmid pProExHtb-Avi-Rpo41, and via ligation into the StuI site, to yield plasmid pProExHtb-Avi2-Rpo41 (see Figure 1). The use of both NarI and StuI sites allowed for insertion of the sequence coding for the biotinylation tag specifically in the same orientation as the *RPO41* ORF, whereas the use of only the StuI site required further analysis of transformants in order to select for clones containing the sequence in the proper orientation. However, the use of the single StuI site for cloning kept the entire linker region between the N-Terminal 6×histidine tag and the *RPO41* ORF intact (i.e. it only added nucleotides to the end of the linker), whereas the directional cloning method resulted in the elimination of 24 nucleotides of the linker sequence, corresponding to a deletion of eight amino-acids from the linker of the expressed protein. The elimination of this linker sequence proved deleterious to the subsequent purification steps, as the expressed protein was unable to be loaded on a Ni-NTA resin for affinity chromatography (not shown). Ni-NTA-based affinity chromatography is based on the interaction between the nickel ions present in the resin and the 6×histidine tag present in the protein. Thus the linker region is crucial for this interaction.

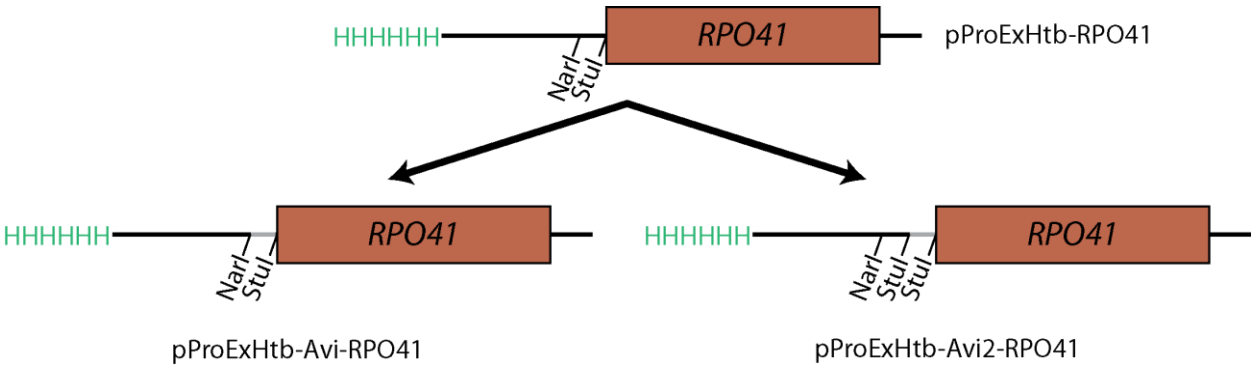

Figure 1: Insertion of a biotinylation tag into a *RPO41* expression vector. We inserted the 45-bp DNA sequence encoding a biotinylation tag (grey) into the Rpo41 expression/purification vector pProExHtb-RPO41[77] by directional cloning (bottom left) and by non-directional cloning (bottom right). The directional cloning removed a portion of the linker between the 6×histidine tag (green) and the *RPO41* ORF (brown), whereas the non-directional cloning kept the linker intact.

Expression of biotinylated Rpo41

Expression of Rpo41 in *E. coli* strain BL21 CodonPlus(RIL) has been shown to result in higher levels of recombinant protein as compared to expression in the conventional BL21 expression strain[77]. This is presumably due to codon bias – recombinant genes expressed in *E. coli* may contain a number of codons that are rare in the endogenous *E. coli* genome, leading to tRNA-limited expression[78]. BL21 CodonPlus(RIL) cells contain the CodonPlus(RIL) plasmid, which codes for overexpression of those rare tRNAs. However, the CodonPlus(RIL) plasmid is selected for using chloramphenicol resistance, the same resistance used for selection of commercially available BirA overexpression plasmids. Thus maintenance of both BirA overexpression and CodonPlus(RIL) plasmids in the same strain is not possible without further cloning. We therefore used strain AVB100, a proprietary MC1061 derivative containing insertion of the *BirA* gene in the chromosome. The *BirA* gene in AVB100 is under the control of an arabinose promoter, allowing for independent induction of expression of BirA (using L-arabinose), and Rpo41 (using IPTG). We transformed this strain with CodonPlus(RIL) to yield strain AVB100 CodonPlus(RIL). We transformed strain AVB100 CodonPlus(RIL) with plasmid pProExHtb-Avi2-Rpo41 to yield strain AVB100 CodonPlus(RIL) pProExHtb-Avi2-Rpo41.

Expression of biotinylated Rpo41 was performed by growing the cells in a similar manner as that described in the literature for the non-biotinylated version of the protein[77], using chloramphenicol and ampicillin in the growth media for selection of the CodonPlus(RIL) and pProExHtb-Avi2-Rpo41 plasmids, respectively. We found that high concentrations of chloramphenicol present in the culture during induction decreased production of Rpo41. This is presumably due to the role of chloramphenicol as an inhibitor of prokaryotic translation. Thus we grew cultures to saturation in 2× Amp/2× Cam media (200 mg/L ampicillin 50 mg/L chloramphenicol) and performed induction in 2× Amp/½× Cam media.

The solubility of biotin is 220 mg/L in water[79]. We added the powdered biotin directly to the culture in order to avoid errors in final biotin concentration due to its incomplete dissolution in a filtered stock solution. The powder was added one doubling time before induction of Rpo41, which reduced the time for any biological contaminants introduced from the powdered biotin to grow while still allowing time for the powder to dissolve in the culture. End cultures showed no sign of biological contamination.

A polyacrylamide gel of cell lysates at various stages of induction is shown in Figure 2A. An Rpo41-specific band is shown to appear at the correct molecular weight (153 kDa) even prior to induction with IPTG. This is due to "leaky" repression by the promoter controlling Rpo41 expression (the trc promoter). Nonetheless, robust expression of Rpo41 is seen after induction by IPTG. Only this expressed protein is biotinylated, as is shown by the horseradish peroxidase-conjugated streptavidin (strep-HRP) hybridization assay in Figure 2B. This "anti-biotin western blot" is identical to conventional western blot assays, but is not strictly a western blot, as the strep-HRP is not an antibody. It shows that a single band of biotinylated protein is present in the final cell lysate, and that band is of the same approximate molecular weight as Rpo41.

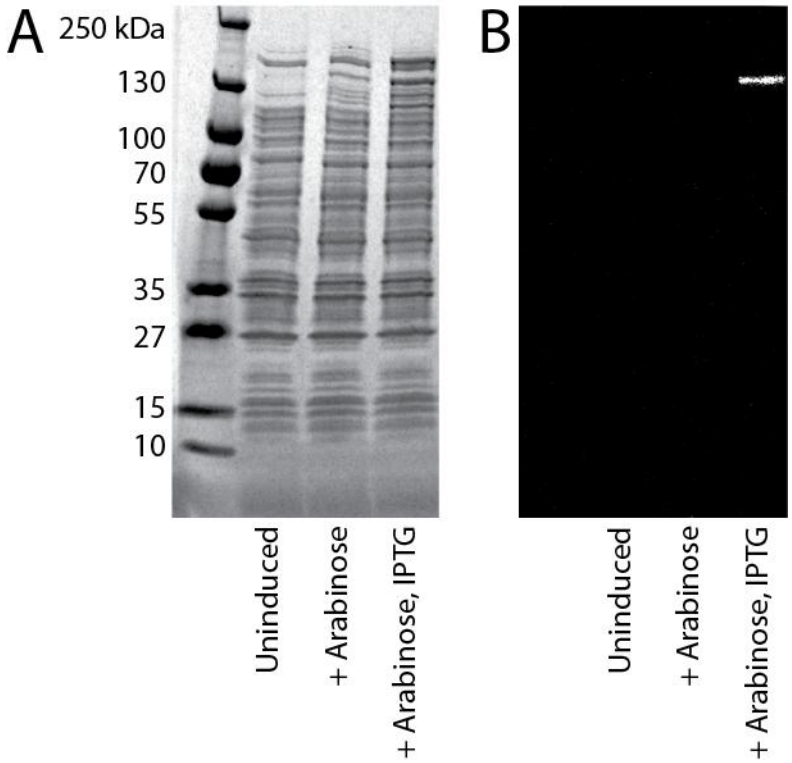

Figure 2: Expression of Biotinylated Rpo41 in *E. coli*. A. A Coomassie-stained polyacrylamide gel of whole cell lysates of AVB100 CodonPlus(RIL) pProExHtb-Avi2-RPO41 before induction ("Uninduced"), after induction of expression of BirA biotin-ligase with L-Arabinose ("+Arabinose"), and after induction of expression of Rpo41 with IPTG ("+Arabinose, IPTG"). **B.** Western blot of the same gel, indicating the presence of biotinylated protein. Transfers of the whole cell proteins to the membrane in B came from replicate lanes of the original gel shown in A.

Purification of biotinylated Rpo41

Expressed recombinant Rpo41 was purified using a protocol based on previously published methods[77], with a few modifications as described below. Cell lysates were initially incubated with an anion-exchange resin to bind nucleic acids. The flowthrough was collected and loaded onto an avidin column, which reversibly bound biotinylated proteins within the lysate. The avidin column was then washed and the biotinylated protein was eluted using many column volumes of buffer containing free biotin. This dilute eluate was then loaded onto a Ni-NTA column, washed thoroughly, and eluted using imidazole. The pooled Ni-NTA fractions were then dialyzed to remove imidazole, aliquoted, snap frozen in liquid nitrogen, and stored at -80°C.

The order of chromatography steps may seem inefficient at first, as the flow rate of the avidin column used is 6 ml/hr./ml of resin, which, for 50 mL of cleared lysate, adds over ten hours to the protocol (including loading, washing, and elution steps). However, performing the avidin chromatography step first proved to be crucial for two reasons:

1. *Trace amounts of free biotin in the sample would hinder downstream experiments*. Because the elution buffer for the avidin column contained 5 mM of free biotin, even a 1:10,000 dilution of biotin (via dialysis) would be insufficient for downstream applications. Binding the pooled avidin fractions to a Ni-NTA column, and washing said column thoroughly proved to be a much simpler, faster, and more effective means to removing free biotin from the sample.
2. *Elution off of the avidin column is not as efficient as off of the Ni-NTA column*. It took over twenty column volumes to fully elute the protein off of the avidin column. Thus the protein was very dilute in the eluate. In contrast, elution from the nickel column was possible in only a few column volumes. This served to rapidly concentrate the sample, and allowed for easier downstream dialysis steps

Polyacrylamide gels showing the protein composition of each of the chromatographic steps described above are shown in Figure 3A&B, while the final purified samples both before and after dialysis are shown in Figure 3C. These show successive Rpo41-specific purification and concentration. The anti-biotin western blot shown in Figure 3D indicates that only one species of biotinylated protein is present within the purified sample.

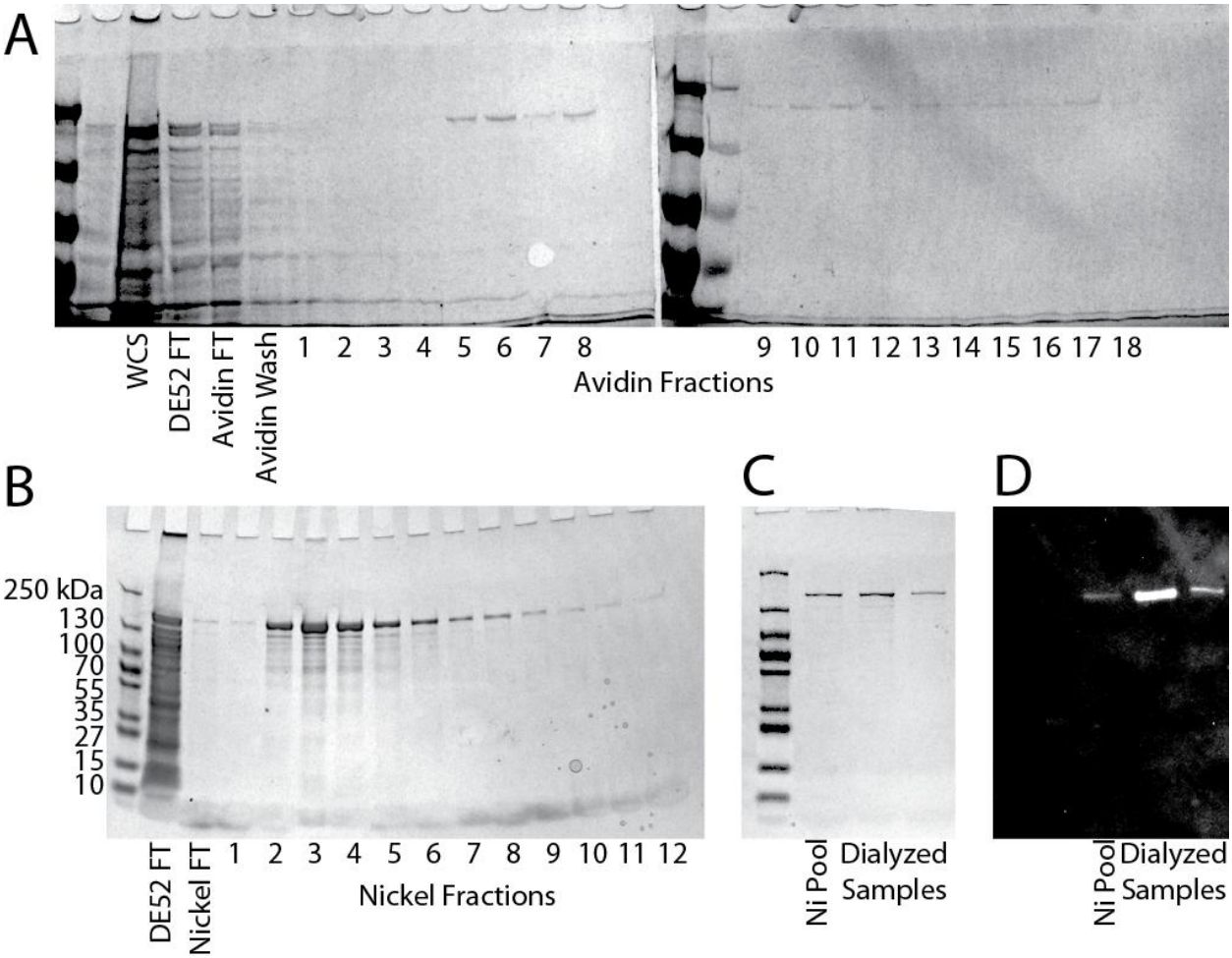

Figure 3: Purification of biotinylated Rpo41. A. The cell lysate of induced AVB100 CodonPlus (RIL) pProExHtb-Avi2-Rpo41 cells ("WCS") was incubated with DE52 anion exchange resin. The flow through was collected ("DE52 FT") and loaded to an avidin column. The avidin column was washed ("Avidin Wash") and eluted using buffer containing free biotin ("Avidin Fractions"). **B.** The pooled eluate from the avidin column was loaded to a nickel column and eluted in a step using buffer containing imidazole ("Nickel Fractions"). Samples of the flowthroughs are labeled in their respective gels as "FT." **C.** A pool of the peak nickel fractions ("Ni Pool") was split and dialyzed against storage buffer ("Dialyzed Samples"). **D.** Streptavidin conjugated HRP hybridization blot of duplicate lanes to the Coomassie-stained gel in C, indicating a high level of purity of the biotinylated samples.

Promoter Independent in Vitro Transcription by Rpo41

In vitro assays of transcription by Rpo41 have been performed previously by either using the initiation factor Mtf1[5] to unwind the template DNA, or in an Mtf1-independent manner, by using a template that has been pre-melted by either a lack-of homology (a "bubble") at the transcriptional start site, or by supercoiling[43]. In each of these studies transcription was promoter specific, i.e. the presence of a known promoter sequence was necessary for transcription from the start site. Others studying transcription at the single

molecule level, however, have developed a means for viable promoter-independent *in vitro* transcription[13,15,81] by creating a synthetic elongation bubble using an RNA primer. Because we only wanted to study the elongation phase of Rpo41 transcription, and because we wanted to compare transcription by Rpo41 with these other polymerases – especially RNA Polymerase II of *S. cerevisiae* (Pol II, see Chapter 3) – we prepared Rpo41 elongation complexes in a similar manner.

The steps taken to prepare stable elongation-competent, promoter-independent complexes using Rpo41 are shown in Figure 4. An RNA oligonucleotide of nine bases is first annealed to a longer DNA oligonucleotide named the template DNA strand (TDS). Addition of Rpo41 with this RNA-DNA hybrid results in its association with the 3' end of the RNA. This complex alone is capable of transcription, but is unstable[82]. The addition of a nontemplate DNA strand (NDS), which is a second DNA oligonucleotide that is complementary to the TDS, stabilizes the complex. This species is similar to the *in vivo* scenario of transcriptional *elongation*, in that there are no initiation factors present, there is a sizeable RNA primer already formed, and its 3' end is at or near the catalytic center of the enzyme.

The preparation of this stable elongation complex (EC) is shown in Figure 4 and represents the first step in all of the subsequent assays discussed in this chapter and the next. Addition of three of the four nucleotides used as a substrate for Rpo41 (A/C/U, without G) results in a stalled complex, which has been utilized in previous studies as a means for selecting active complexes[13,15,81], is shown in the penultimate step of Figure 4, and is followed by runoff transcription resulting from the addition of all four nucleotides.

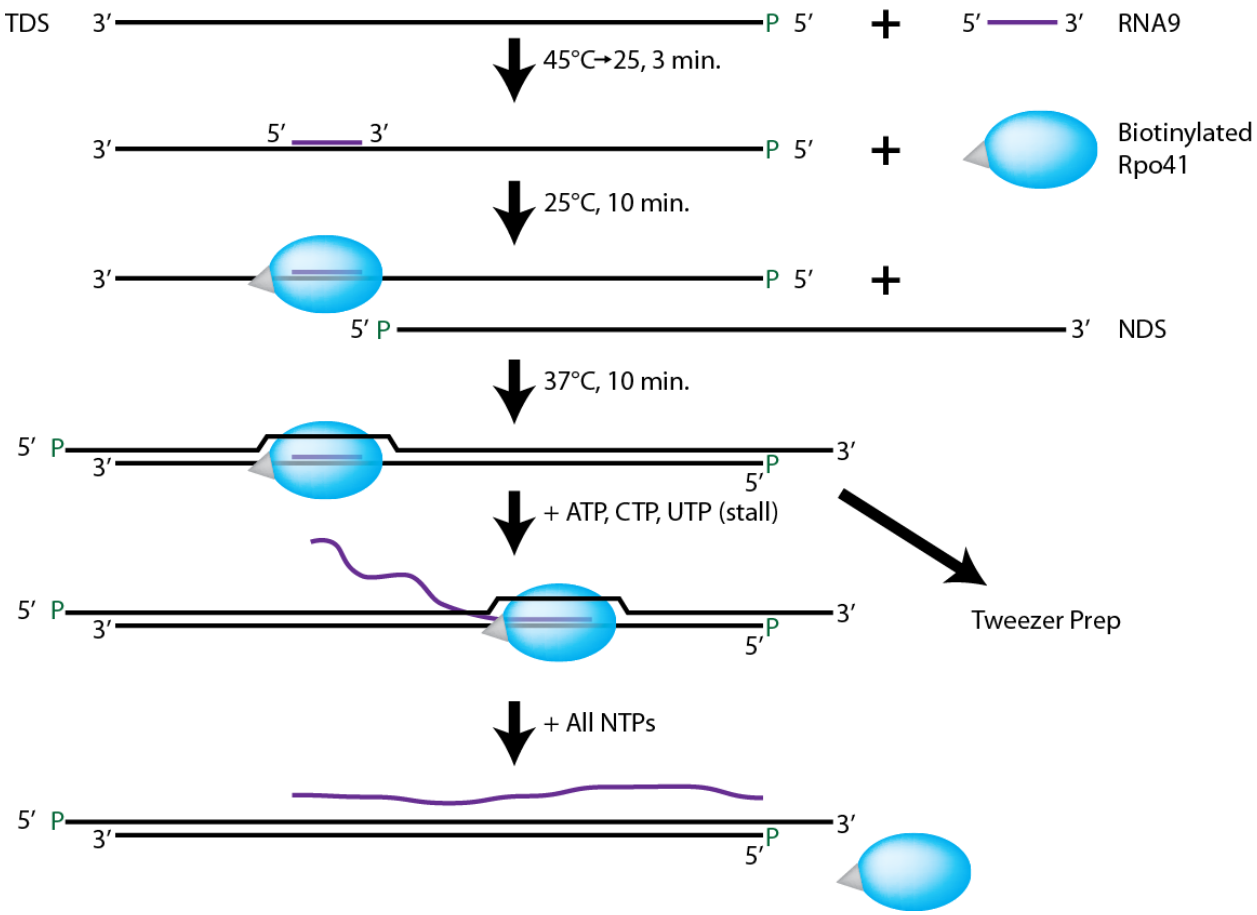

Figure 4: Preparation of stable elongation complexes using biotinylated Rpo41. All protocols involving *in vitro* transcription by Rpo41 or Pol II in this study are initially prepared by annealing a 9 base RNA primer ("RNA9" purple) to a template oligonucleotide ("TDS"). This hybrid is then incubated with the polymerase ("Biotinylated Rpo41", blue sphere with grey attachment), and then incubated with a nontemplate oligonucleotide ("NDS") to form a stable elongation complex. This complex can than either be subject to further preparatory steps for optical trapping assays ("tweezer prep"), or be analyzed directly using bulk methods. Shown here are stall and chase steps.

We performed solution biochemistry studies of transcription using the described preparation of ECs containing RNA9 end-labeled with ^{32}P. This radioactive labeling of the RNA primer allowed us to visualize the transcription products. After transcription, we loaded the sample in a denaturing polyacrylamide gel and separated the products using electrophoresis. We imaged the products using storage phosphorescence imaging.

A radiogram showing promoter-independent, Rpo41-based *in vitro* transcription and associated assays is shown in Figure 5. The first lane is from the elongation complex with no added substrate, and should give a band for RNA9 only. However, contamination of the preparation with GTP and ATP (due to the high concentrations of relatively impure ATP used in the end-labeling reaction), results in "bleed through" of even the no-NTP reaction to the +2 nucleotide, which is seen as the third band from the bottom.

Addition of ATP, CTP, and UTP to the EC resulted in a stalled complex, as loaded into the second lane in Figure 5. This resulted in a decrease in the RNA9 band and an increase in the +2 band. Once again, contamination of the reaction with trace CTP resulted in bleed through of transcription to the next CTP nucleotide at +8. Addition of all four nucleotides to the stalled complex resulted in runoff transcription, seen as the darker upper band in lanes 4 and 5.

Incomplete or aberrant annealing of any of the oligonucleotides in the EC preparation, especially the NDS, can lead to incorrect bubble formation[82]. We utilized the properties of two RNases to determine the hybridization state of the transcribed RNA: RNase T1 only digests single-stranded RNA, whereas RNase H only digests DNA/RNA hybrids. In a properly functioning elongation complex the NDS re-anneals to the TDS after passage of the polymerase (a moving bubble), and the transcribed RNA remains single stranded. Thus the radiolabeled RNA from a properly functioning *in vitro* transcription protocol should be resistant to degradation by RNase H but susceptible to degradation by RNase T1. Results from these RNase assays are seen in lanes 5 (RNase H) and 6 (RNase T1) of Figure 5 and confirm proper bubble formation in promoter-independent *in vitro* transcription by Rpo41.

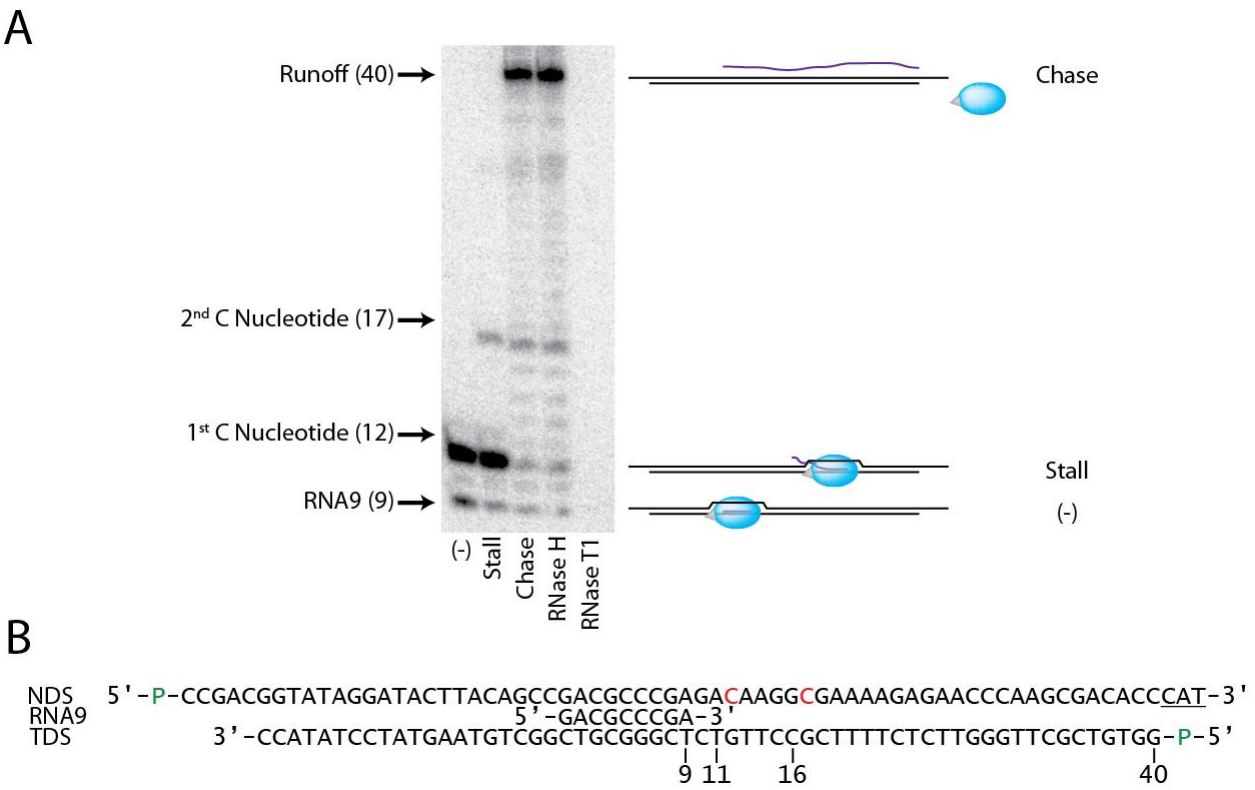

Figure 5: Promoter-independent *in vitro* transcription of Rpo41. A. Radiograms of transcriptional products using RNA9 end-labeled with ^{32}P. Elongation complexes using RNA9 (purple) end-labeled with ^{32}P were prepared as in Figure 4 and submitted to denaturing gel electrophoresis after the addition of nothing (lane 1), a stall using ATP, CTP and UTP (lane 2), or a stall and then chase with all four NTPs (lane 3). The chased product

was also subjected to treatment by RNase H (lane 4), which cuts only RNA-DNA hybrids, or RNase T1 (lane 5), which cuts only single stranded RNA. A cartoon depiction of each transcriptional state is shown to the right of the gel. The band corresponding to the stalled product in lane 1 is due to contamination of the elongation complex with other nucleotides stemming from impurities in the radioactive ATP used for end labeling, as is the third major band from the bottom in lane 2. Other bands are products stemming from stalled or aborted transcription. The locations of stall sites and the transcript runoff is given to the right of the gel, with the nucleotide positions in parenthesis. **B.** The prepared elongation complex. The length of major bands seen in A is given below the sequences. Stall nucleotides are shown in red in the NDS. 5' phosphates are shown in green. The DraIII 3' overhang used for subsequent ligation is underlined.

Bulk Characterization of Rpo41

With a functioning bulk *in vitro* assay of Rpo41-based transcription, we moved to characterizing the enzyme in a number of ways. This served to help future single molecule studies, as optimization using bulk methods is much more efficient than doing so through single molecule experiments. Furthermore, Rpo41was not very well characterized, and therefore the development of an efficient *in vitro* transcriptional system allowed us to address a number of open questions concerning Rpo41-based transcription.

We analyzed each transcription reaction by numerically integrating the intensity of the band created by the runoff transcript. This integrated spot density, or spot volume, is proportional to the amount of runoff transcript present in the particular reaction.

The first step in characterizing Rpo41 was to determine a proper working concentration of NTPs such that subsequent runoff transcription assays would be sensitive to further modification of the reaction conditions. To this end, we performed runoff transcriptional assays with titrated amounts of NTPs in the final reaction. These reactions were done in triplicate, and are shown in Figure 6A. A graph of the spot volume of each runoff normalized to the maximum spot volume for that replicate is given as Figure 6B.

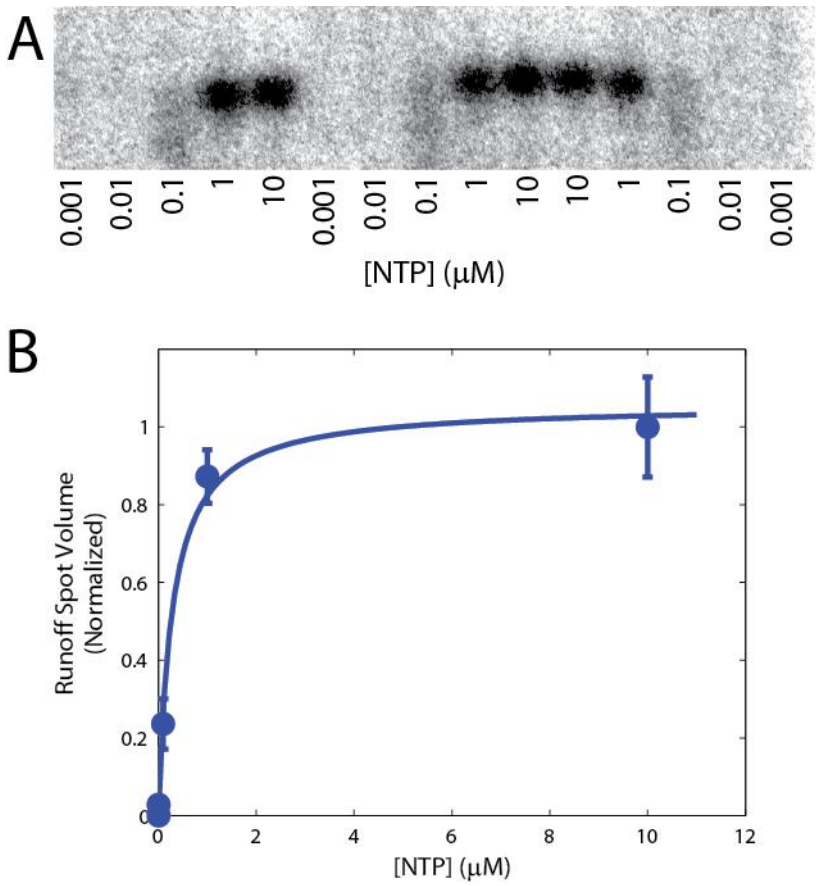

Figure 6: Transcript runoff intensity dependence on NTP concentration. A. Storage phosphor image of radiolabeled transcript runoffs for various NTP concentrations as noted. The length of the runoff reaction is 31 bp, and the reaction occurred over 1 min. **B.** Normalized spot densities of Rpo41 transcription runoffs for various final NTP concentrations. Reactions were done in triplicate. Spot volumes were normalized to the maximum integrated spot density for a particular replicate, and then averaged over replicates. Error bars are the standard deviations of the normalized spot densities for a given NTP concentration.

Note that transcript runoff assays measure the number of polymerases that have reached the end of the template in a given time. This excludes not only those that were too slow to reach the end of the template DNA, but those that did not start and those that fell off or terminally paused as well. Because the distance from the end of RNA9 to the end of the template DNA in this experimental setup is 31 bp, even a polymerase transcribing at 1 bp/s would reach the end of the template in under 40 s. Thus these transcript runoff assays cannot be considered a good measure of bulk polymerase speed, but rather a measure of its convolution with the overall ability for the buffer to sustain active and stable elongation complexes.

Optimization of buffer conditions

As shown in Table 1, published studies using *in vitro* transcription of Rpo41 have used varying buffer conditions, which in turn have differed from the buffers used for single molecule characterization of Pol II. We therefore decided to optimize three components of the transcription buffer – counter-ion concentration, $MgCl_2$ concentration, and reducing agent identity and concentration. As a result of the aforementioned NTP titration experiments, all reactions were performed using limiting (0.5 μM ea.) NTPs.

Table 1: Buffers used for *in vitro* transcription.

Concentration (mM)	Rpo41 Buffer 1[43]	Rpo41 Buffer 2[25]	Pol II Buffer[14]
Tris-Cl	30	50	20
KCl	50	0	40
NaCl	0	10	0
$MgCl_2$	10	20	5
$ZnCl_2$	0	0	0.01
BME	0	0	2
DTT	0.1	10	0
EDTA	2	0	0

Graphs of the normalized integrated spot densities are shown in Figure 7A ($MgCl_2$ titration) and Figure 7B (KCl titration). Below each graph are the images of the bands used for the spot volume calculations. These assays show that $MgCl_2$ concentration does not have a significant effect on overall transcription efficiency of Rpo41, at least within the range of 5-30 mM, while KCl concentrations above 150 mM do result in a significant decrease in transcriptional efficiency. This is in contrast to Pol II, which is unaffected by KCl concentration within the same range studied here[83]. As a result of these findings, further study of transcription by Rpo41 was performed in buffer containing 10 mM $MgCl_2$ and 40 mM KCl.

We also saw an effect of reducing power on transcriptional efficiency, as shown in Figure 7C. We tested two different reducing agents – dithiothreitol (DTT) and β-mercaptoethanol (BME). Transcription was inhibited at DTT concentrations below ~ 4 mM and at BME concentrations above ~80 mM, with a total abolition of transcription at a BME concentration of 600 mM. Transcriptional efficiency was relatively insensitive to DTT concentrations ranging from 3-12 mM. Because single molecule transcriptional assays are performed at room temperature, and aqueous solutions of both BME and DTT are relatively unstable, we chose 10 mM DTT as the final concentration of the reducing agent in our transcriptional buffer. This allowed for significant degradation of the DTT while still maintaining enough reducing power to sustain robust transcription.

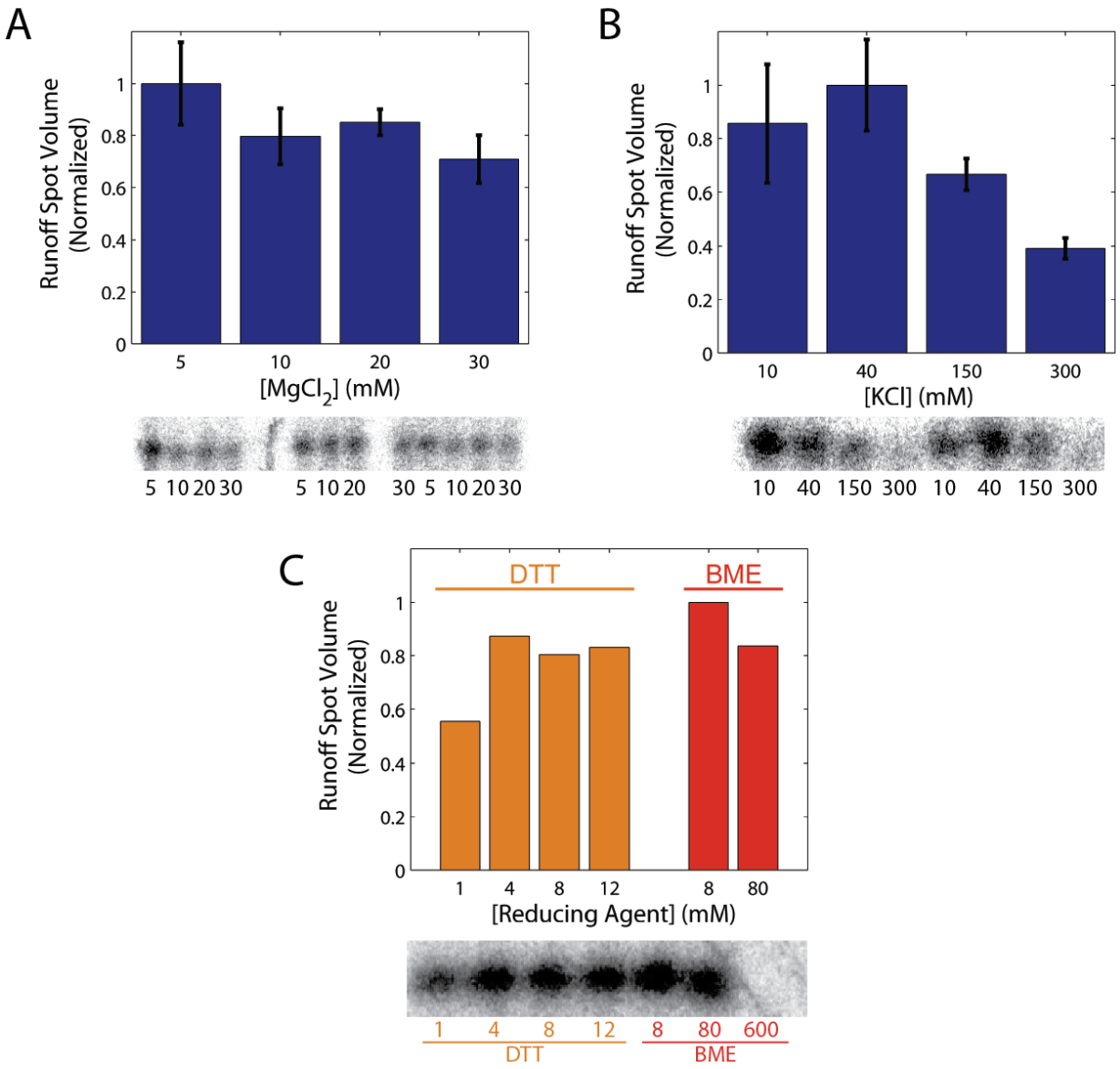

Figure 7: Buffer optimization using bulk runoff assays. Integrated spot volume of the runoff transcript with varying concentrations of buffer $MgCl_2$ (**A**), KCl (**B**), and reducing agent (**C**). Two reducing agents – dithiothreitol (DTT) and β-mercaptoethanol (BME) – were assayed in C. Below each graph is an image of the storage phosphor screen used for the spot densitometry calculations.

Determination of the robustness of Rpo41-based transcription

The assays used above to optimize buffer conditions used fresh aliquots of Rpo41, and were performed within an hour of preparation of the stable elongation complex. A sample of Pol II used in single molecule experiments, however, is stored on ice and used over the course of several days. The stability of Pol II elongation complexes greatly increases the efficiency of data collection, as the preparation of stable elongation complexes and their ligation to handle DNA (described below) takes several hours, a process that can be

avoided on subsequent days if the ECs are stable for that long. We therefore performed additional transcript runoff assays to test the stability of the Rpo41-based elongation complex. To do this, we prepared one sample of elongation complexes, aliquoted it, and stored it on ice for differing amounts of time before chasing. The spot volumes of chases performed from 5 to 31.5 hr. after the preparation of the elongation complex is plotted in Figure 8A, with the image of the storage phosphor screen placed below the plot. The time dependence of the EC transcription efficiency fits a single exponential ($R^2 = 0.99$) with a half-life of 7.5±1.5 hr. (error is the 95% confidence interval of the fit). This corresponds to the loss of activity seen when optical trapping experiments span over more than one day, and led us to prepare new ECs each day of an optical trapping experiment.

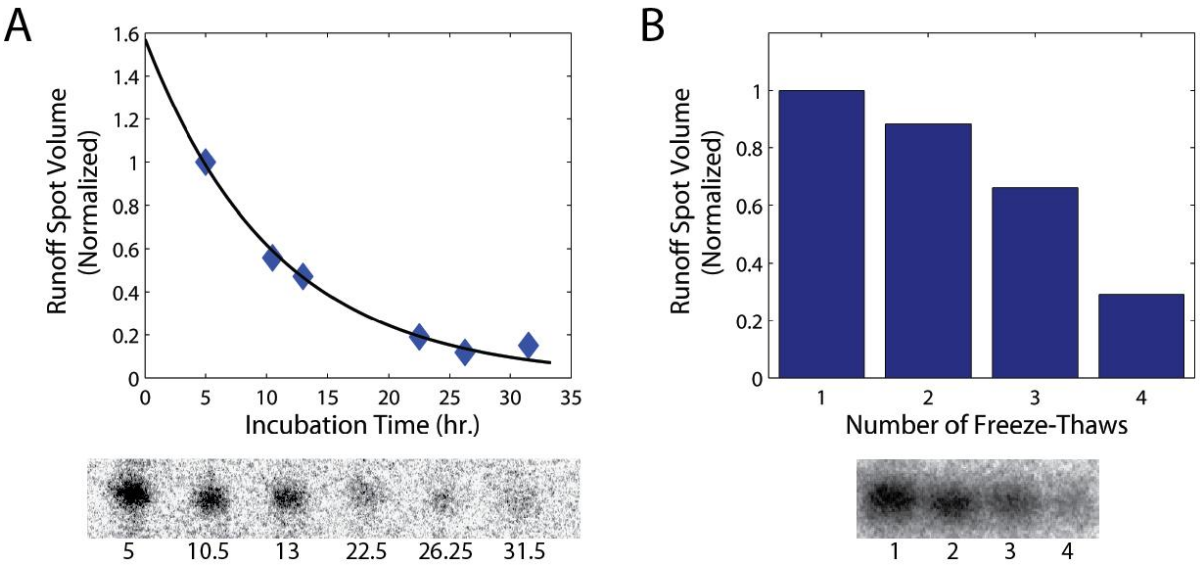

Figure 8: Stability assays of Rpo41-based transcription. A. Runoff spot volumes of elongation complexes stored on ice for varying amounts of time. The data (blue) fits an exponential (black) with a half-life of 7.4±1.5 hr. (error is 95% confidence interval of fit). **B.** Transcription efficiency after multiple freeze-thaws of the polymerase prior to EC preparation. Rpo41 loses approximately 30% of its transcriptional activity during each freeze-thaw. The image of the bands used for the spot volume calculations is shown below each plot.

As is the case with most biochemical experiments, the most precious reagent is the enzyme; oligonucleotides can be ordered, PCR amplifications are relatively simple to prepare, and most other reagents are commercially available – it is the enzyme that takes the most work and expense to procure. We therefore measured the sensitivity of Rpo41 to multiple freeze-thaws as a means to determine if we could re-use leftovers from aliquots in future experiments. Figure 8B shows that the transcriptional activity of Rpo41 is adversely affected by multiple freeze-thaw cycles. The same aliquot was subjected to different numbers of freeze-thaw cycles prior to EC preparation. Taking the ratio of successive spot volumes indicates that Rpo41 loses 30±10% of its activity during each freeze-thaw cycle (error is standard error of the mean). This indicates that bulk studies, which usually

measure relative differences in transcription efficiencies, may use preparations of Rpo41 that have been freeze thawed up to three times. Conversely, the efficiency of performing single molecule experiments is often limited by the *absolute* activity of the enzyme being studied – inactive polymerases, for example, still take up valuable experimentation time. Thus preparations of Rpo41 used in single molecule experiments should only subject to a maximum of one freezing cycle.

Mechanochemical Characterization of RPO41 Using single molecule Optical Trapping

With reasonably optimized buffer conditions and preparation protocols determined by bulk transcription, we now turn to single molecule characterization of Rpo41. Unlike bulk experiments, single molecule studies allow for elucidation of individual components of the transcription cycle.

In order to perform single molecule studies of transcription, the stable elongation complex needed to be ligated to a much longer piece of dsDNA called the handle, as shown in Figure 9. A digoxygenin molecule is conjugated to the distal end of the handle, and serves as a means for a second attachment point between the biochemical and mechanical domains of the experimental setup as shown in Figure 10. The handle also greatly increases the length of the DNA downstream of the polymerase. This allows for longer transcriptional runs, and for differentiation between valid single tethers and merely two stuck beads.

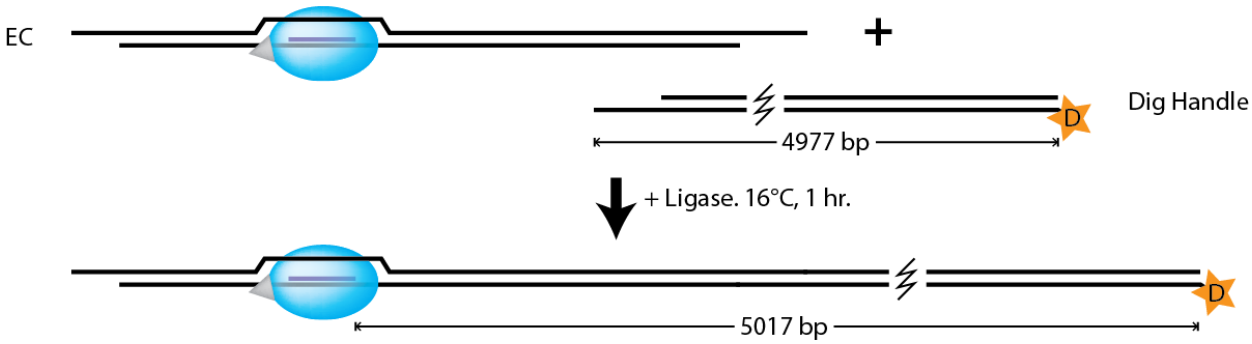

Figure 9: Ligation of elongation complexes to handle DNA. "Handle" DNA (right) is attached to the downstream end of the stable elongation products (left, prepared as shown in the first three steps of Figure 4) by a simple ligation reaction. Handle DNA is prepared by a PCR reaction containing one primer with a digoxygenin covalently attached to its 5' end (orange star).

In order to observe transcription by single molecules of Rpo41, we used a dual-trap optical tweezers instrument, in an opposing force geometry, as shown in Figure 10A. Elongation complexes ligated to handles were attached to streptavidin coated polystyrene beads via the linkage between the biotinylated polymerase and the streptavidin on the bead. These beads were then flowed into our reaction chamber, and trapped with one of the optical traps. After trapping a single bead, anti-digoxygenin coated polystyrene beads were flowed into the chamber, and trapped in the other optical trap. The beads were then rubbed together until the magnitude of the force between the beads increased when the beads were pulled apart. After confirmation of the

correct tether length, corresponding to a single tether, and 30 s of calibration time, NTPs were flown into the chamber. Active polymerases would then begin transcription, resulting in a shortening of the tether. We kept the trap positions fixed in our experiments, which eliminated noise caused by force-feedback, and resulted in an increasing of the force as the tether length decreased (and the beads moved away from the centers of the traps). An example of the force time course of an entire trace, from fishing to a terminal break of the tether after transcription, is shown in Figure 10B. The length of the DNA between the beads, which decreases as transcription brings the beads together, was calculated using worm-like chain model of DNA length under force[84], and is plotted in Figure 10C. Note that, like other RNA polymerases, transcription by Rpo41 is characterized by times of processive translocation, followed by short intervals of pausing.

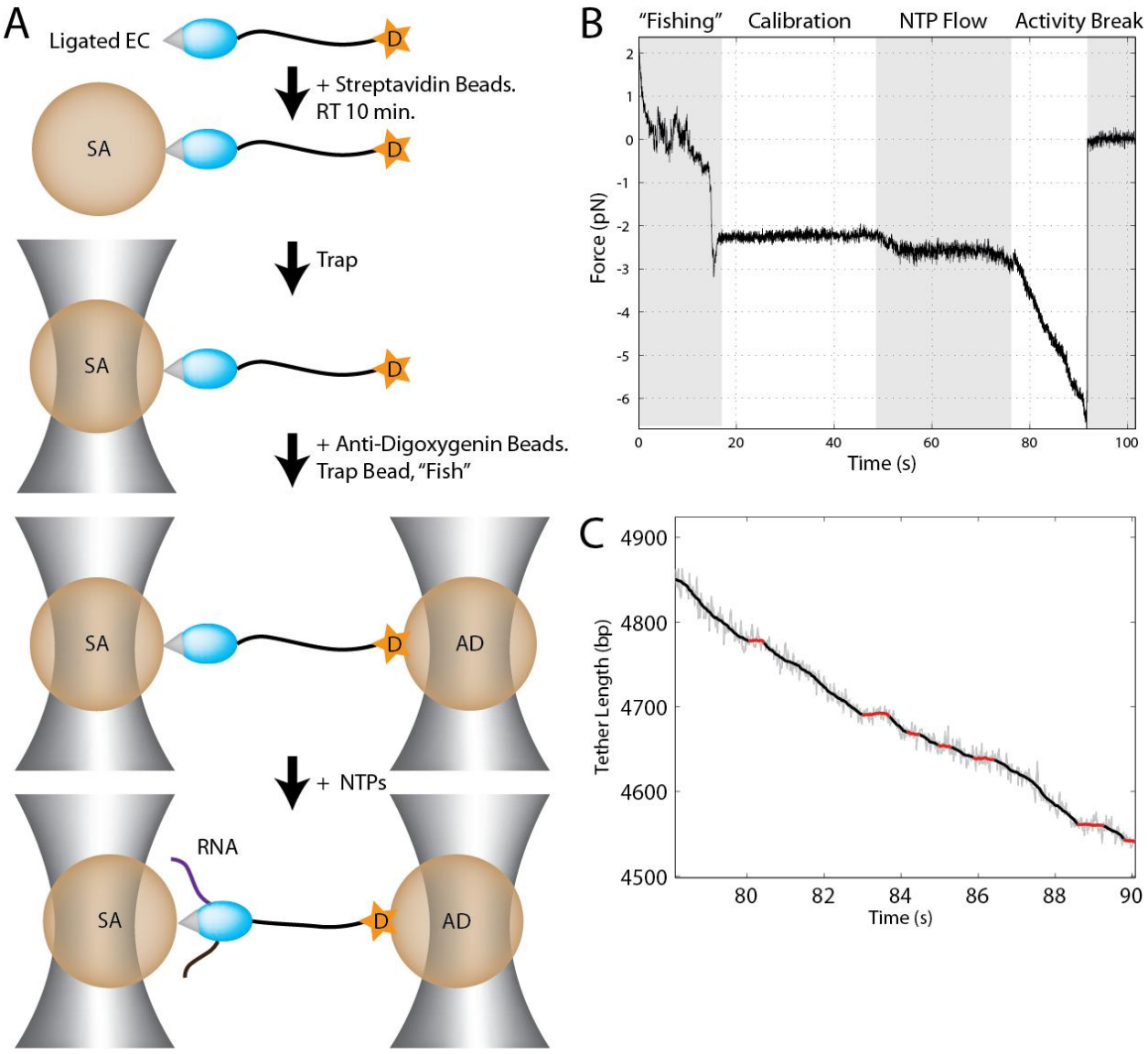

Figure 10: Optical trapping experiments. A. Preparation of the optical trapping experiment. A ligated elongation complex (prepared as shown in Figure 4 and then Figure

9) is linked to a streptavidin-coated polystyrene bead (SA), and trapped in one of the traps of a dual-trap optical tweezers apparatus. An anti-digoxygenin coated polystyrene bead (AD) is trapped in the other trap, and the two beads rubbed together until there is force between the two beads upon separation ("Fishing"). Subsequent addition of NTPs to the reaction chamber results in transcription. **B.** Force-time graph showing the steps of an optical trapping experiment. **C.** The tether length as a function of time for the same trace in B. Data averaged at 50 Hz is shown in grey, Savitsky-Golay filtered data (1 Hz) in black, and pauses in red.

Evidence of pausing and backtracking by a phage-like polymerase

As described above, Rpo41 displays clear pausing behavior. This is noteworthy, as those studying T7 bacteriophage RNA polymerase – an enzyme very closely related to Rpo41 – have stated that they have seen no indication of pausing[18], or stated that it is a rare event and given no evidence of its existence or characterization[17]. Both the finite pause duration and retrograde motion of the polymerase are shown in Figure 11B. This retrograde motion, termed backtracking, has been hypothesized to be the result of diffusion of the transcription bubble along the upstream end of the template DNA after disassociation of the catalytic center of the enzyme from the 3' end of the RNA. The distribution of backtrack distances is shown in Figure 11C. The average backtrack distance was 5.8±0.6 bp (error is standard error of the mean).

The distribution of backtrack distances for Rpo41 does not differ significantly from that of Pol II, with which we have taken data in the same way. The mean backtrack distance for Pol II was 7.2±1.1 bp. Pol II did, however, have a few pauses with very long backtracks, with the maximum being 27 bp. Nonetheless, a Kolmogorov-Smirnov test (a nonparametric goodness of fit test[85]) between the two distributions indicates that do not differ significantly ($p=0.3$).

The distribution of pause durations greater than 0.5 s (a conservative estimate of the lower limit of our pause picking algorithm) is given in Figure 11D. The pause durations follow a $t^{-3/2}$ distribution, shown as the black line, which further supports a model in which pauses are governed by diffusion of the backtracked polymerase[13]. A much more detailed examination of pausing, including comparisons of Rpo41 with Pol II, is given in Chapter 3.

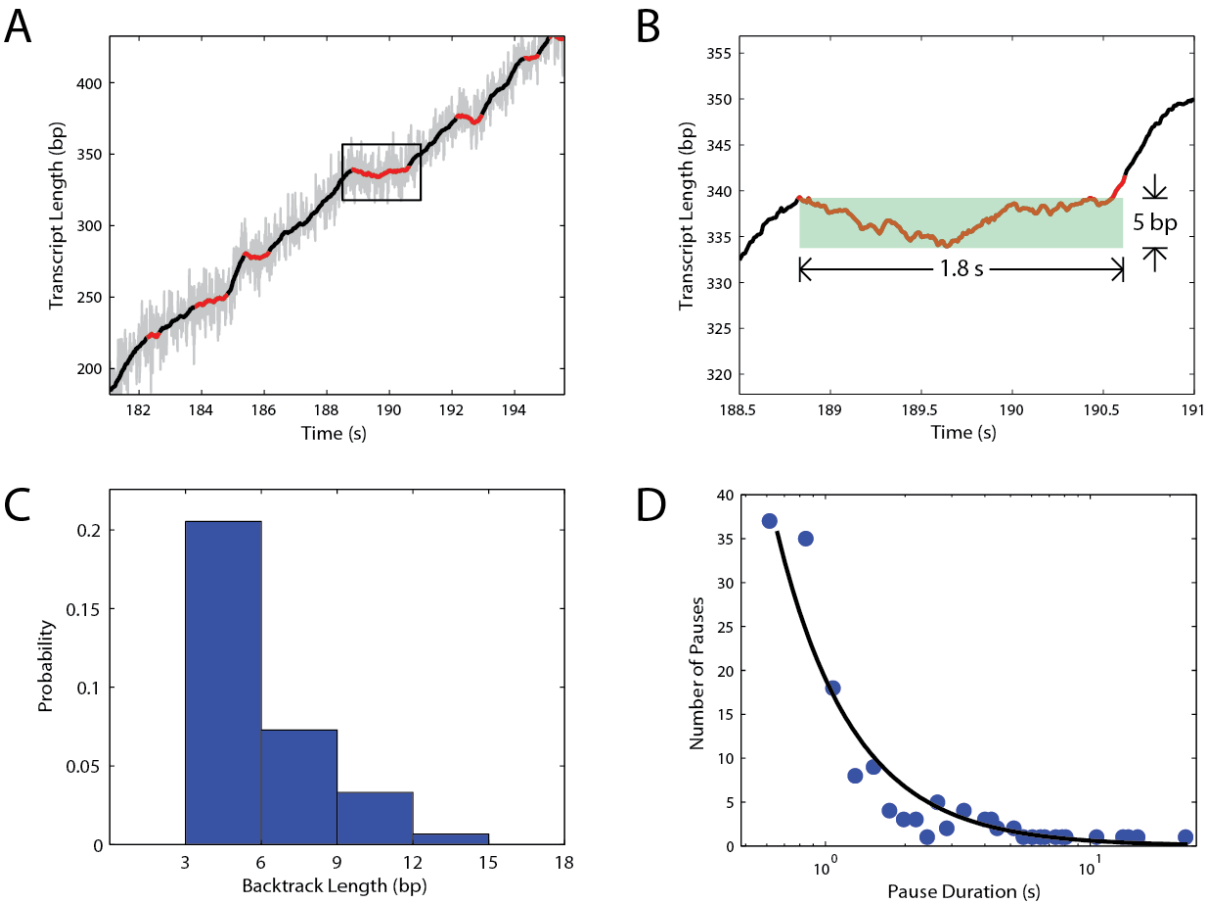

Figure 11: Pausing properties of Rpo41. A. A typical trace of transcription by Rpo41. Like other RNA polymerases, transcription by Rpo41 is governed by periods of elongation interspersed with periods of pausing. Color scheme is as in Figure 10C. **B.** Detail of boxed region in A. The green shaded box is the region used to calculate both the duration and retrograde distance ("backtrack") moved by Rpo41 during a pause. **C.** Distribution of all backtrack distances greater than 3 bp. **D.** Distribution of the pause durations greater than 0.5 s. The data (blue circles) fits a $t^{-3/2}$ distribution (black line, $R^2 = 0.91$).

Determination of the kinetic properties of Rpo41

Single molecule analysis of transcription of Rpo41 allows us to determine the canonical enzymatic properties of maximal velocity and, because we have measured such velocity at multiple NTP concentrations, the Michaelis-Menten coefficient for nucleotide incorporation during elongation. To this end, we measured pause-free velocity for each trace, calculated the mean pause-free velocity for all traces at a particular NTP concentration, plotted it verses NTP concentration, and fitted it to the standard Michaelis-Menten rate equation[86],

$$v = v_{\max} \frac{[NTP]}{K_m + [NTP]} \quad (1)$$

where v is the pause-free velocity, v_{max} is its value at saturating NTP concentration, [NTP] is the concentration of each NTP present during the transcription reaction, and K_m is the Michaelis-Menten constant specific to the enzyme.

Determination of the pause-free velocity of a trace has many of the same challenges as determination of the pause statistics of that trace – both fundamentally depend on the ability to distinguish pauses from activity. This is not often as simple as is shown in Figure 11A, and, as is the case in almost every scientific calculation, the apparent simplicity shown in Figure 11D betrays the thousands of person hours spent on pause picking algorithms by the author, members of the Bustamante Lab, and others in the single-molecule field. Traces often contain regions of "slow" activity which may either be considered a pause, or activity interspersed with many short pauses. While computer algorithms developed in the Bustamante Lab and improved upon in this work (as discussed in Chapter 3) are successful at recognizing the vast majority of pauses, a full determination of pauses would require infinitesimal resolution as well as an ability to see behind noise – a task left to the divine. These problems can largely been avoided when determining pause statistics by only considering easily identifiable pauses – those above a threshold pause duration – and extrapolating for pauses outside this finite observation window.

Determination of pause-free velocity does not benefit from the luxury of being able to only eliminate pauses *above* a certain duration; a proper determination of pause-free velocity requires removal of *all* pauses from a trace. Furthermore, slow traces are especially affected by noise, and consequently are scored as long regions of pausing interspersed with very short, very fast "pause-free" transcription that is actually just noise. This effect disproportionately affects conditions which favor slow transcription, e.g. those in which substrate NTP concentration is limiting. An example of a slow trace is shown in Figure 12A, and can be compared to a typical trace in Figure 12B. These traces must be discarded from the dataset, as any properties gleaned from the pause picking algorithm cannot be trusted.

Furthermore, traces that are slow for a particular condition may not be scored as activity for conditions in which the average velocity is already quite slow. In other words, conditions favoring slow transcription may have the lower tail of their velocity distributions cutoff as a result of the inability to distinguish activity from drift or noise, whereas the full velocity distribution is sampled for those conditions of higher mean velocity.

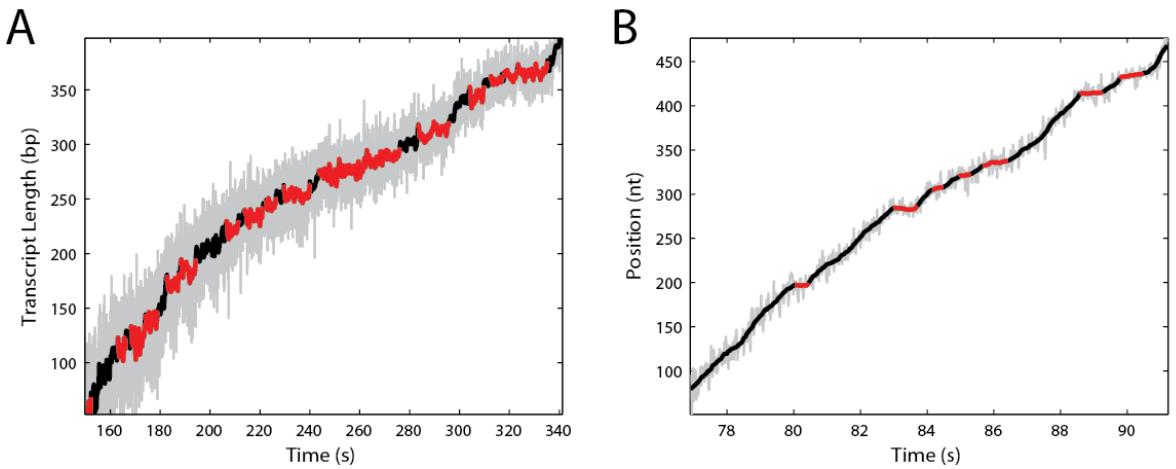

Figure 12: Complications picking pauses for slow traces. A. A trace of transcription by Rpo41 with limiting (50 μM) NTPs present in the buffer. Due to the inability for the computer algorithm (or even a human user) to distinguish pauses from noise, this trace was not used in any analysis (i.e. it was discarded from the dataset). Conditions favoring slow transcription have more of these slow traces, which must be discarded, resulting in a bias of the data. **B.** A typical trace for Rpo41 with saturating (1 mM) NTPs. Pauses are easily identifiable in such a traces. Note the differences in timescale between the two plots.

The systematic errors in the determination of pause-free velocities are exemplified in Figure 13A, which shows the pause-free velocities of Rpo41 when in the presence of differing concentrations of NTPs. In this case, the pause-free velocities of for each trace were calculated by averaging the velocities in all of the pause-free regions (the black lines in Figure 11A), resulting in one value for pause-free velocity per trace. The pause-free velocities for all traces in a particular condition were then averaged and plotted in Figure 13A. Note that the pause-free velocities at concentrations ranging from 100 to 1000 μM NTPs follow Michaelis-Menten kinetics, yet deviate significantly at 50 μM (R^2 with 50 μM data point = 0.92, 1.00 without). However, the position of the 50 μM data point relative to the other data points shifts considerably when considering overall velocities (Figure 13A).

The overall velocity data provides a much better fit to the Michaelis-Menten curve both with (R^2 = 0.95) and without (R^2= 0.98) the 50 μM data point. Note that the diffusive model of pausing predicts that overall velocity should not strictly follow Michaelis-Menten kinetics, as pauses in that model are considered off-pathway intermediates, and do not involve substrate binding. This deviation is small[83], but yet again disproportionately affects conditions favoring pausing (i.e. those with lower substrate NTP concentrations), and may be the reason that the Michaelis-Menten curve fits the data without the 50 μM data point better than that the one that contains it.

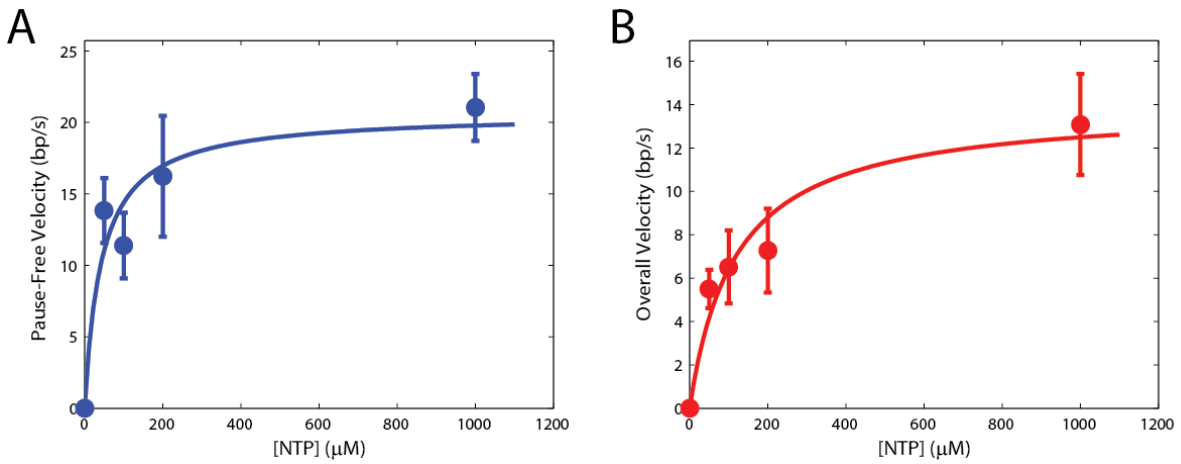

Figure 13: Systematic errors in determination of pause free velocities by picking pauses. A. Pause-free velocities for different NTP concentrations (blue circles). Pause-free velocity was determined by removal of pauses determined by the pause-picking algorithm described in Chapter 3. Errors in pause picking, especially for slow conditions, cause a deviation of the pause-free velocity from that predicted by Michaelis-Menten kinetics (solid blue line). **B.** Overall velocities (red circles), which do not depend on pause picking algorithm, provide a much better fit to Michaelis-Menten kinetics (solid red line). Errors are standard errors of the mean.

The systematic errors involved in the use of pause picking algorithms led us to investigate statistical methods of determining pause free velocities. The method we found relies on the notion that pausing is a diffusive process, and therefore the velocities during a pause arise from one-dimensional Brownian motion of the polymerase moving along the template. Consequently, the velocities of pauses should be Gaussian distributed and centered on zero velocity. We therefore composed the histogram of all velocities in a given data set, and fitted a zero-centered Gaussian to those bins with negative velocities.

The fit of the pauses (yellow solid line in Figure 14A) was then extrapolated over all data (dashed yellow line), and subtracted from the data. The leftover velocities correspond to the data cleaned of pauses, i.e. the pause-free velocity distribution. The average of this leftover data, then, is the pause free velocity of the data set.

The statistically-determined pause-free velocities for various NTP concentrations are shown in Figure 14B. They show Michaelis-Menten kinetics ($R^2 = 0.99$) with K_m = 22±13 μM^{-1} and v_{max} = 25±2.5 bp/s (errors are 95% confidence intervals of the fit).

Plugging the values of K_m and v_{max}, into Equation (1), the pause-free velocity of the polymerase at 0.5 μM NTPs has been determined to be 0.58 bp/s. This is consistent with the bulk experiments shown in Figure 6, which indicated that 0.5 μM NTPs were necessary for limiting transcription of Rpo41 over 31 bp in 60 s.

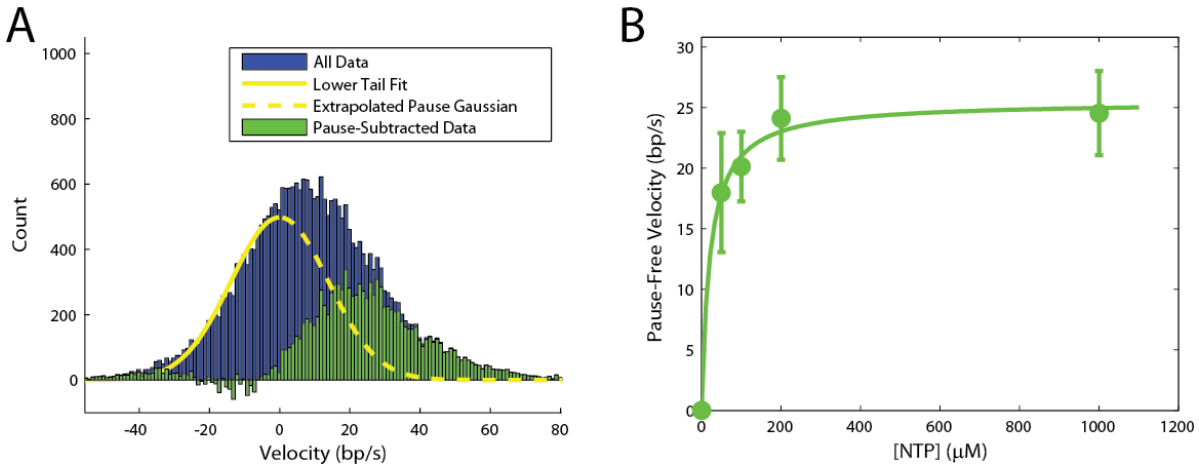

Figure 14: Determining pause-free velocity via statistical methods. A. A histogram of all the velocities of Rpo41 transcription at 1 mM NTPs (blue). The bins containing negative velocities were fit to a Gaussian distribution (solid yellow line) and extrapolated over all bins (dashed yellow line). This "pause Gaussian" was then subtracted from the velocity histogram to give the pause-subtracted data (green). The average of the pause subtracted data is the pause-free velocity. **B.** Pause-free velocities for different NTP concentrations (green circles) as determined by the statistical methods described in the text. These velocities fit the Michaelis-Menten curve (green line). Errors are 95% confidence intervals of the bootstrapped pause-free velocities.

Comparisons with other polymerases

A comparison of the enzymatic properties between Rpo41, Pol II, the *E. coli* RNA polymerase, and the RNA polymerase from the bacteriophage T7 are given in Table 2. Note that the Michaelis-Menten coefficients of Rpo41 more closely match that of Pol II than that of the phage polymerases. This is in contrast to the fact that Rpo41 is evolutionarily related to the phage polymerases, and has no discernable homology to Pol II. The similarities between Rpo41 and Pol II are maintained when considering pauses, in that the properties of pause density (average number of pauses per kB of transcription) and pause duration are statistically indistinguishable between the enzymes (the analysis of pausing is given in Chapter 3).

Table 2: Properties of RNA polymerases studied via single molecule methods

Enzyme (organism)	K_m (μM)	v_{max} (bp/s)
Rpo41 (yeast mito)	22±13	25±2.5
RNAP (T7)	174±17[16]	129±8[16]
Pol II (yeast nuc.)	99±50[83]	12.2±4.5[13]
RNAP (*E. coli*)	223±57[11]	15±4[9]

The widely accepted endosymbiotic hypothesis posits that mitochondria originated from an α-proteobacterial ancestor[87]. Yet the mitochondrial RNAP is of a single subunit (a ssRNAP), just as those of the T-odd phages, and betrays the mitochondria's eubacterial

lineage. This mosaic nature of the mitochondria has baffled evolutionary biologists aiming to decipher the life history of the organelle[88], while the discovery of an ssRNAP outside of the phage family opened new areas of research on the evolution of transcription[89]. The finding that the kinetic properties of the mitochondrial RNAP are closer to that of Pol II is sure to enrich the story further.

Just how Rpo41 came to be within the mitochondria is a topic of much debate within the literature. Other ssRNAPs were found after discovery of the single-subunit nature Rpo41[90], and it had been postulated that Rpo41 originated from those sources: phages, linear mitochondrial plasmids, or from an original nucleus-encoded ssRNAP. Further analysis using comparisons of the replicative primase-helicase in human mitochondria and the mechanism of replication in yeast mitochondria have supported the notion that the mitochondrial RNAP was acquired as part of the endosymbiont genome, as opposed to an accompanying phage-like entity[91]. This implies that the mitochondrial RNAP has been present for virtually the entire history of the modern eukaryotic cell. Similarly, Rpo41 has been separated (physically and physiologically) from the T7 RNAP since at least the initial endosymbiotic event. Though more rigorous analysis is certainly necessary, our findings that Rpo41 has more similar properties to Pol II than it does to the T7 RNAP are consistent with its coevolution with the nuclear polymerase, and its divergent evolution from that of the T7 phage.

This divergence of Rpo41 and the T7 RNAP is further supported when considering r/K selection theory[92]. In this theory, small organisms in unstable environments evolve traits that promote rapid reproduction with a low chance of offspring survival (r-selection), while larger organisms in more stable environments evolve traits that promote production of fewer but more robust offspring (K-selection). Yeasts, though almost always considered r-selected, are significantly less so than phages, which can be considered at the very r-end of the r/K selection spectrum. Indeed, the selective pressure for Rpo41 and Pol II in these terms is very similar, as they are in the same organism. It is no wonder, then, that Rpo41 has evolved to have characteristics more similar to Pol II than to T7 RNAP.

FUTURE DIRECTIONS

The establishment of robust *in vitro* assays involving Rpo41 transcription – both at the bulk and single molecule levels – opens a new door into the study of transcription in general, and mitochondrial transcription in particular. Here we discuss just a few of the additional experiments involving Rpo41 that could result in a deeper understanding of the mechanisms of transcription, the details of mitochondrial gene regulation, and the evolutionary history of the organelle.

STUDIES OF THE PUNCTUATION MODEL AND OTHER POST-TRANSCRIPTIONAL PROCESSES

As has been discussed in the introduction, there are a myriad of processes involving yeast mitochondrial transcription that have yet to be fully (or in some cases even partially) elucidated in the over 50 years of study focused on this organelle. One example of this is the punctuation model of RNA processing[93]. This model stems from the fact that many of the mature mRNAs are produced by cleavage by tRNA processing enzymes at the tRNAs

embedded within the primary transcript. The 3' termini of protein-encoding genes are further defined by the dodecamer motif 5'-AAUAA(U/C)AUUCUU-3', which is processed by mtEXO. The kinetics of transcription around both the tRNAs and the dodecamer motif may shed light on the mechanisms governing RNA processing, and could presumably be examined using the studies we have described here with little modification.

STUDIES OF INITIATION

The few *in vitro* studies that have been done on Rpo41 involve its role in transcription initiation, and have used a promoter sequence. These studies have discovered a number of processes that may govern mitochondrial gene expression, such as the role of the +1 and +2 initiating nucleotides[56], how Mtf1 affects transcription initiation[80], as well as a proposed mechanism of coupling respiration to transcription via ATP "sensing" by Rpo41[57]. These processes could be further elucidated in high spatial and temporal resolution using the single molecule methods described here. Furthermore, the development of a promoter-free *in vitro* transcription system provides an opportunity to explore the claims described above (most notably the ATP sensing mechanism) in the presence and absence of a promoter.

STUDIES OF REPLICATION

Very little is known about the mechanism of DNA replication in yeast mitochondria, with the vast majority of information being the result of inference from genetic evidence[50]. One thing that is known is the identity of the DNA polymerase at the center of the replicative process, Pol γ, which, like Rpo41, is homologous to those in the T-odd phage family and not those of the α-proteobacteria[91].

Circumstantial evidence has indicated that replication in yeast mitochondria is performed through a rolling circle mechanism[94]. The discovery of plasmid-like origins of replication that are also associated with transcription start sites suggest that Rpo41 primes replication[95], as does T7 RNAP prime replication in T7 phage[96]. This is further supported by the fact that strains with deficient expression of Rpo41 lose large portions of their mitochondrial DNA[69]. Furthermore, certain amino terminal deletion mutants of Rpo41 result in instability of the mitochondrial genome while retaining their transcriptional activity, indicating a duel-domain nature of the enzyme[65].

Rpo41 deletion strains do not, however, lose all of their DNA, indicating that, should Rpo41 be the replicative primase, it is not the only one. Furthermore, disruption of mitochondrial protein synthesis in general (which deficiency in Rpo41 certainly causes) also leads to genomic instability and propagation of fragments of the mitochondrial genome[66].

The phenomenon of propagation of small mtDNA fragments (instead of whole genomes) is not limited to mutant strains. Certain mtDNA fragments exhibit a replicative advantage, and can often replace wild-type mitochondrial genomes, despite the apparent selective disadvantage conferred to the strain due to a lack of a functional respiratory chain. This phenomenon has also been shown to be Rpo41 independent, i.e. replication of these "hypersuppressive" sequences is not primed by Rpo41.

Examination of transcription around mitochondrial origins of replication by single molecule methods may lead to insights into the mechanisms by which Rpo41 acts as a replicative primase, if it does at all. Further examination using templates containing sequences known to be hypersuppressive may also provide elucidation on the process of biased mitochondrial inheritance in yeast.

DETAILED COMPARISONS WITH PHAGE POLYMERASES

As described above, the fact that Rpo41 is more similar to Pol II than the T7 RNAP has significant implications regarding the origins of the eukaryotic cell. While T7 RNAP has been a model RNA polymerase for many years – it was first purified in 1984[97], and its crystal structure was first published in 1993[98] –single molecule studies have only been done at relatively low resolution[16-18]. Thus the pausing behavior of T7 RNAP has not been elucidated, a critical component in the comparison between it and any other characterized polymerase.

A benefit of the fact that T7 RNAP has been a model polymerase is that there are already established mechanisms for "rapid and simple" expression and purification of the enzyme[99] (the protein is even available commercially), as well as for in vitro biotinylation of the purified protein[100]. Thus establishment of high-resolution single molecule optical trapping studies of T7 RNAP should be relatively straightforward.

Future studies involving the study of Rpo41 initiation and termination, as well as its role in replication, may add further detail to the evolutionary past of the mitochondrial organelle.

Finally, though we have tried our best to expound on the potential implications single molecule data will have on the studies of eukaryotic evolution, we do not purport to be evolutionary biologists. Careful analysis of the kinetic properties of Rpo41 in the context of its evolutionary history and that of other RNAPs must be done in a systematic and rigorous manner.

MATERIALS AND METHODS

STANDARD TECHNIQUES AND STARTING MATERIALS.

Standard molecular biology techniques (restriction digestion, ligation, transformation, etc.) are as commonly described[101]. Plasmid minipreps from *E. coli* were performed using the Zymo Research (Orange, CA) Zyppy Plasmid Miniprep Kit. Gel extractions were performed using the Qiagen Gel Extraction Kit (Valencia, CA). Sequencing was performed as non-premix reactions by Sequetech, Inc. (Mountain View, CA). All enzymes were from New England Biolabs (Ipswich, MA). All chemicals were purchased from Sigma-Aldrich (St. Louis, MO), unless otherwise noted. All synthetic oligonucleotides, including those used as PCR primers, were synthesized by Integrated DNA Technologies, Inc. (IDT, San Diego, CA), and are given in list form in Table 3.

LB media (10g/L bacto-peptone, 5g/L yeast extract, 10g/L NaCl) containing appropriate antibiotics was used to culture *E. coli*. Solid media contained LB and 2% (w/v)

agar. Working (1×) concentrations of antibiotics were: 25 mg/L chloramphenicol (Cam) and 50mg/L ampicillin (Amp).

Table 3: Oligonucleotides used in this study

Name	Sequence (5' - 3')
Nar-Avi-Stu	P-CGCCGGCCTGAACGATATTTTTGAAGCGCAGAAA ATTGAATGGCATTGGAGG
Nar-Avi-Stu-rev	P-CCTCCAATGCCATTCAATTTTCTGCGCTTCAAAAA TATCGTTCAGGCCGG
Stu-Avi-Stu	P-CCTGGCCTGAACGATATTTTTGAAGCGCAGAAAA TTGAATGGCATTGGAGGCCT
Stu-Avi-Stu-rev	P-CCTCCAATGCCATTCAATTTTCTGCGCTTCAAAAA TATCGTTCAGGCCAGGCCT
NDS	P-CCGACGGTATAGGATACTTACAGCCGACGCCCGA GACAAGGCGAAAAGAGAACCCAAGCGACACCCAT
TDS	P-GGTGTCGCTTGGGTTCTCTTTTCGCCTTGTCTCGG GCGTCGGCTGTAAGTATCCTATACC
RNA9	GACGCCCGA
DraIII-phage-rev	AATATTCACCATGTGTTAGAAAACGATAACACCGTG
dig-phage-fwd	Dig-AAGCTGCATGTGCTGGAACTTCAC

"P" = 5' phosphorylation
"Dig" = 5' digoxygenin NHS ester

Cloning of a Biotinylation-Tagged Rpo41 Expression Vector

Synthetic oligonucleotides containing the 45 bp sequence encoding for the biotinylation tag ("avi-tag") (GGCCTGAACGATATTTTTGAAGCGCAGAAAATTGAATGGCATTGG) were flanked with restriction sites on both ends and synthesized by IDT. One oligonucleotide, Nar-Avi-Stu, contained a 5' NarI overhang and a 3' StuI blunt end. Another oligonucleotide, Stu-Avi-Stu contained StuI blunt ends at both termini. The reverse complements of both sequences were also generated, with the last two nucleotides of Nar-Avi-Stu left off in the reverse complement (in order to generate the NarI overhang). All sequences were synthesized with 5' phosphates. The lyophilized oligonucleotides were resuspended in TE (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8) to a final concentration of 10 μM. Complementary oligonucleotides were annealed by mixing the two solutions, placing the mixture on a 95°C heat block, and allowing the heat block to equilibrate to RT.

Plasmid pProExHtb-RPO41[77] was digested with either StuI or StuI and NarI by standard techniques. The digestion was terminated by incubation at 65°C for 20 min. The cut vector was then dephosphorylated using Antarctic phosphatase as per the manufacturer's instructions, and cleaned via gel extraction.

The cleaned products were ligated with either Nar-Avi-Stu (for the double-cut plasmid), or Stu-Avi-Nar, and transformed into *E. coli* strain DH5α. Positive transformants were grown in LB media containing 100 mg/L ampicillin and analyzed by restriction digestion

followed by sequencing. Colonies containing plasmids of confirmed sequence were re-grown and the plasmids purified via miniprep. Plasmids containing Nar-Avi-Stu were termed pProExHtb-Avi-RPO41, and those containing Stu-Avi-Stu in the same orientation as *RPO41* were termed pProExHtb-Avi2-RPO41. The cloning scheme is given graphically in Figure 1.

Expression and in vivo *Biotinylation of Rpo41*

The CodonPlus(RIL) plasmid was purified by miniprep of *E. coli* strain BL21 CodonPlus(RIL) and transformed into strain AVB100 (Avidity, LLC, Aurora, Co.) to yield strain AVB100 CodonPlus(RIL).

AVB100 CodonPlus(RIL) was transformed with plasmid pProExHtb-Avi2-Rpo41 using standard the chemical transformation protocol and selected on LB/Amp/Cam-agar plates. The whole transformation plate was scraped using sterile 50% glycerol, and stored in 750 µL aliquots at -80°C. One of these aliquots was used to inoculate 100 mL of LB/2×Amp/2×Cam, pre-warmed to 20°C, which was then incubated with shaking overnight at 20°C.

The saturated overnight culture was used to inoculate two 2 L flasks of LB/1×Amp/0.5×Cam, pre-warmed to 20°C. When the OD_{600} of the culture media reached 0.35, powdered biotin was added to a final concentration of 50 µM and overexpression of BirA was induced by addition of 40% filter-sterilized L-arabinose to a final concentration of 0.4%. When the culture reached an OD_{600} of 0.6, expression of the biotinylation-tagged Rpo41 protein was induced by addition of 1 M IPTG to a final concentration of 1 mM. After six hours of growth, the cells were rapidly cooled on ice (all steps heretofore were on ice or at 4°C, unless otherwise noted), and pelleted by centrifugation at 6000 RPM for 30 min. in a FIBERlite F8-4x1000Y rotor (FIBERlite Centrifuge, Santa Clara, CA). The cells were washed two times in 20 mM Tris-HCl pH 8 and snap-frozen in liquid nitrogen for storage. The wet weight of the cell pellet was 8 g.

Purification of Biotinylated Rpo41

All purification buffers contained 10% glycerol. All purification steps were performed at 4°C or on ice. The frozen cell pellets were resuspended in 35 mL TB300 (20 mM Tris-HCl, pH 7.9, 300 mM KCl, 10 mM $MgCl_2$, 10 mM DTT) containing two tablets of EDTA-free protease inhibitors (Roche Diagnostics GmbH, Mannheim, Germany). Phenylmethylsulfonyl fluoride (PMSF) was added immediately before sonication, to a final concentration of 1 mM. The solution was sonicated using 15 10 s pulses at 24 W, each followed by a one min. rest, using a Misonix 3000 sonicator (Farmingdale, NY) fitted with a microtip. The lysate was cleared by centrifugation at 3000 RPM for 30 min using a JS-13.1 rotor (Beckman Coulter, Inc., Brea, CA).

The cleared lysate was incubated with light shaking for 5 min. with 0.5 g DE52 anion exchange resin (Whatman, Inc., Piscataway, NJ) that had been equilibrated with TB300. The resin was then drawn through a polypropylene column using a 30 mL syringe, and the flowthrough collected.

The flowthrough was loaded at 6 ml/hour to a polypropylene column pre-packed with 1 mL of SoftLink Avidin Resin (Promega Corporation, Madison, WI) that had been prepared as per the manufacturer's instructions, and pre-equilibrated with TB300. The column was then washed with 10 mL TB300, and eluted with 30 mL TB300 containing 5 mM biotin.

The entire eluate from the avidin column was loaded onto a 1 mL HisTrap FF nickel column (GE Healthcare Life Sciences, Piscataway, NJ) that had been pre-equilibrated with TB300. The column was then washed with 10 mL TB300, and eluted in a step using TB300 containing 500 mM imidazole. Peak fractions were analyzed using polyacrylamide gel electrophoresis, pooled, dialyzed against TB300, aliquoted into 3 µL aliquots, snap-frozen in liquid nitrogen, and stored at -80°C.

Visualization of Biotinylated Protein

Polyacrylamide gel electrophoresis of purified proteins, chromatographic fractions, or cell lysates was performed as per common protocols. For cell cultures, 0.5 OD_{600} units were collected, pelleted by centrifugation at 4500×g for 1 min., washed with 1 mL of Tris pH 8, repelleted, and resuspended in 2×Laemmli buffer[102] to a final volume of 50 µL. These samples were incubated at 95°C for 15 min., chilled rapidly on ice, and centrifuged at 16,000×g for 15 min. on bench top centrifuge. The supernatant was removed and 5 µL loaded to an 8-16% gradient Tris-HCl polyacrylamide gel (Bio-Rad, Inc. Hercules, CA). Whole protein visualization was performed using the Invitrogen (Carlsbad, CA) SimplyBlue SafeStain, as per the manufacturer's instructions.

For specific visualization of biotinylated proteins, unstained gels were transferred to nitrocellulose membranes at 100 V for 1 hr., washed in TBS (50 mM Tris-HCl 150 mM DTT pH 7.4), blocked by incubation in TBS + 5% (w/v) BSA at RT for 1 hr., and incubated with a 1:10,000 dilution of streptavidin-conjugated horseradish peroxidase (Thermo Fisher Scientific, Rockford, IL) at RT for 1 hr. After three washes with TBS for 10 min. each, visualization was performed using Millipore (Billerica, MA) Immobilin western HRP chemiluminescent substrate as per the manufacturer's instructions.

Bulk In Vitro *Transcription Studies*

End-labeled RNA9 was prepared by combining 1 µL of 15 µM RNA9 with 1 µL of >7000 Ci/mmol, >100 mCi/mL γ-^{32}P-ATP (MP Biomedicals, Salon, OH), 20 U T4 Polynucleotide Kinase (PNK), and 2 µL 10×PNK buffer in 20 µL total reaction volume. After incubation for 1 hr. at 37°C, the reaction was stopped by incubation at 65°C for 20 min.

DNA/RNA hybrids were created by adding 1 µL 15 µM TDS to the end-labeled RNA9 reaction and incubating in a MJ Research (Bio-Rad, Hercules, CA) PTC-200 Thermo Cycler set to go from 45°C to 25°C over the course of 5 min. Association of Rpo41 (or Pol II) with the DNA/RNA hybrid was performed by addition of 0.8 µL of Rpo41 or 0.5 µL of Pol II to the DNA/RNA hybrid and subsequent incubation at 25°C for 10 min. Stable elongation complexes were constructed by adding 1 µL of 15 µL NDS to the DNA/RNA/RNAP reaction and incubation for 10 min. at 37°C.

The standard runoff transcription reaction consisted of 2.5 µL of (non-immobilized) elongation complexes 0.5 µM ea. NTP, and TB40 (20 mM Tris-HCl, pH 7.9, 40 mM KCl,

10 mM $MgCl_2$, 10 mM DTT) to a final reaction volume of 10 µL. For titration of NTP concentration, serial dilutions of a 100 µM stock were used.

For experiments in which multiple transcription products (besides the runoff) were to be analyzed (that shown in Figure 5), elongation complexes were immobilized on streptavidin agarose resin (Thermo Fisher Scientific), washed three times in TB40 and resuspended to 50 µL in TB40 before further steps. Prior to use, streptavidin-agarose resin was blocked by incubation with 1 mg/mL BSA for 10 min. at RT followed by three washes in TB40.

The streptavidin-agarose bound elongation complexes were then split into one 5 µL aliquot (the negative control) and one 20 µL aliquot. TB40 was added to double the volume of each reaction, and 4 µL 10 mM ea. ATP, CTP, and UTP were added to the 40 µL sample. After a 10 min. incubation at RT, the 40 µL sample was split into one 10 µL aliquot (the stalled sample), and one 30 µL aliquot. A mixture containing 10 mM of each NTP was then added to the 30 µL aliquot to a final concentration of 1 mM ea. NTP. This sample was again incubated at RT for 10 min., and split into three 10 µL aliquots. RNase H (5 U) was added to one aliquot, RNase T1 (1 U, Ambion, Austin TX) to the other, and nothing to the third (the chased sample).

All reactions (for both immobilized an non-immobilized samples) were stopped after 60 s by adding 10 µL stop buffer (8M urea, 89 mM Tris-Cl pH 8, 89 mM boric acid, 50 mM EDTA, 0.25% bromophenol blue 0.25% xylene cyanol), 1 µL 25:24:1 Phenol:Chloroform:Isoamyl Alcohol, and incubation at 65°C for 20 min. followed by immediate transfer to ice. Each reaction was then run on a 22×16.5 cm 20% polyacrylamide gel containing 8 M urea until the bottom dye front was 2/3 of the way to the bottom of the gel. Gels were then wrapped in plastic wrap and used to expose storage phosphor screens (GE Healthcare, Piscataway, NJ) overnight. Storage phosphor screens were imaged using a Typhoon Trio gel imager (GE Healthcare) as per the manufacturer's instructions.

Preparation of Stable Elongation Complexes for Optical Trapping Experiments

Digoxygenin labeled DNA handles were prepared by PCR using lambda DNA (NEB) as a template, DraIII-phage-rev as the reverse primer, dig-phage-fwd as the forward primer, and Taq DNA polymerase. The final concentrations were 0.4 µM (each primer), 0.1 ng/µL (lambda DNA), 0.1 U/µL (enzyme), and 0.2 mM (ea. dNTP) in 1×Thermopol buffer (NEB). Thermocycling conditions were 95°C for 5 min., then 30 cycles of 95°C for 30 s, 65°C for 30 s, and 72°C for 6 min. This PCR reaction was cleaned using the Qiagen PCR Purification Kit as per the manufacturer's instructions, and digested using 0.1 U/µL DraIII in 1×NEB Buffer 3 with 1×BSA by incubation at 37°C for one hour, followed by 65°C for 20 min. The final DNA concentration of the digestion reaction was 100 ng/µL. This reaction was cleaned again by using the Qiagen PCR purification kit, and concentrated to 200 ng/µL.

Protein G coated beads (0.2 µm Spherotech, Lake Forest, IL) were pelleted by centrifugation at 3000 RPM for 30 min. and resuspended Ab X buffer (100 mM Na_2HPO_4, 100 mM NaCl) to the original starting volume. Anti-digoxygenin antibodies (Roche) were dissolved in PBS (140 mM NaCl, 2.7 mM KCl, 80 mM K_2HPO_4, 20 mM KH_2PO_4, pH 7.4) to a final concentration of 1µg/µL; 65 µL of this solution was added to 1 mL of the beads in Ab X

buffer. Dimethyl pimelimidate (DMP) was dissolved in Ab X buffer to a final concentration of 50 mg/mL, and 30 µL of this solution was added to the bead/antibody mixture to initialize crosslinking. The crosslinking reaction was incubated with light shaking for 1 hr. at RT, pelleted, and resuspended by gentle shaking in 1 mL 1 M tris base for 2 hr. The beads were then pelleted again, and resuspended in TB40.

The DNA/RNA hybrid was annealed by mixing of 1 µL 0.75 µM TDS with 1 µL 1.5 µM RNA9, followed by incubation in a thermocycler set to bring the reaction temperature from 45°C to 25°C over 3 min. Rpo41 (2 µL) was added the hybrid and the reaction incubated at 25°C for 10 min., followed by addition of 1 µL of 15 µM NDS and incubation at 37°C for 10 min.

The elongation complex was then added to 28.6 µL ligation mastermix (0.5 mM ATP, 0.75 pmol DNA handle). After addition of 560 U (1.4 µL) T4 DNA ligase the reaction was incubated at 16°C for 1 hr. and stored on ice.

Single Molecule Studies

Both anti-digoxygenin and streptavidin-coated polystyrene (Spherotech) beads were blocked by incubation with 1 mg/mL BSA at RT for 5 min.

The ligated elongation complexes were incubated with 1 µL of the streptavidin beads in 5 µL final volume for 10 min. (the amount of ligation added was optimized for each preparation and differed during the day). ATP, CTP, and GTP were then added to a final concentration of 10 µM, and the reaction was incubated at RT for 5 min., followed by dilution of the reaction to a final volume of 1 mL using TB40. This bead preparation was introduced into the fluidics apparatus of the optical chamber, and single beads were trapped in one of the optical traps. Anti-digoxygenin polystyrene beads were diluted in the same manner, entered into the optical chamber, and trapped in the other trap.

The two beads were rubbed together by moving one of the optical traps towards the other. Tethers were identified by the increase in force after the beads were separated, and tethers of the correct length were determined by comparing the distances between the two traps with the predicted length of DNA as determined by the worm-like chain model of DNA elasticity[103]:

$$\frac{x}{L} = 1 - \frac{1}{2}\left(\frac{k_b T}{FP}\right)^{1/2} + \frac{F}{S} \tag{2}$$

where L is the contour length of the DNA (4975 bp), F is the force, P = 53 nm is the persistence length of dsDNA, and S = 1200 pN is its elastic stretch modulus.

After holding tethers of the correct length for 30 s for later calibration steps, chase buffer (TB40 with 1 mM ea. NTP 1 µM PP_i) was flown into the chamber until either transcription was detected by the concomitant shortening of the tether and increase in force between the beads, or there was no discernable activity after 30 s of NTP flow.

Analysis

Data was taken at 2000 Hz. Data from traces containing active transcription were averaged by decimation to 50 Hz and then smoothed using a second order Savitsky-Golay

filter with a time constant of 1 s. Detection of pauses is discussed in Chapter 3. All analysis was performed in Matlab (The Mathworks, Natick, MA). The instrument was controlled using LabView (National Instruments, Austin, TX).

Chapter 3: Nascent RNA Structure Determines the Transcriptional Dynamics of RNA Polymerases

Introduction

As was discussed in Chapter 2, transcription by both the main nuclear polymerase in yeast, Pol II, and the prokaryotic polymerase in *E. coli*, is not a continuous process, but rather is characterized times of active transcription with interspersed times of pausing. We have shown, for the first time, that Rpo41, a phage-like polymerase, shares this property. We describe here the current proposed model for the mechanism of pausing: backtracking. We then discuss the current status of this model, and the recent work that has been done to refine it.

The Backtracking Model of Transcriptional Pausing

As is shown in Figure 15, retrograde motion of the polymerase during pausing, termed backtracking, is clearly visible during some pauses for both enzymes. This has led to a model in which pauses are caused by this retrograde motion of the RNAP[104]. In backtracking, the polymerase moves backward with respect to the template DNA and nascent RNA. This results in a movement of the 3' end of the nascent RNA *downstream* of the catalytic center of the polymerase[105,106]. Note that once this occurs, a symmetry of the elongation complex is achieved – there is a disrupted RNA/DNA hybrid both upstream and downstream of the polymerase. Shifting of this system by one base pair backward does not change this geometry, to first order. Thus the backtracked polymerase occupies a periodic potential, with the periodicity being one base pair[104].

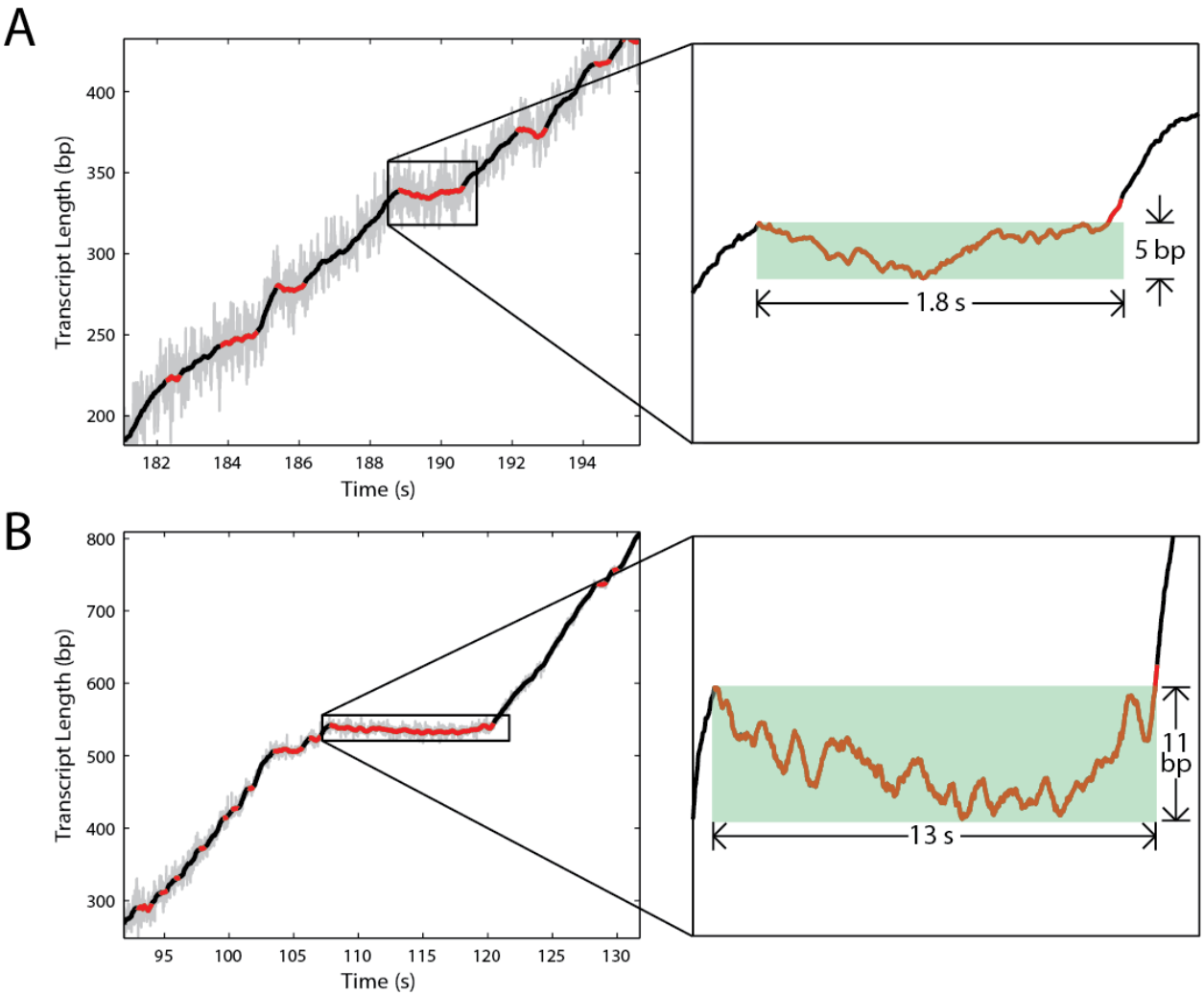

Figure 15: Transcriptional pauses containing backtracking. Traces of transcription by Rpo41 (**A**) and Pol II (**B**) show retrograde motion of the polymerases during pausing. The boxed regions of each trace are shown in greater detail on the right. The regions of backtracking are green, with the maximum backtrack distance and pause duration as labeled. Data averaged at 50 Hz is shown in grey, Savitsky-Golay filtered data (1 Hz) in black, and pauses in red.

A diagram of the backtracked polymerase occupying a periodic potential is given in Figure 16A. This potential is arbitrary, in that the amplitude of the potential is designated h, and the shape of the potential need not be as shown (only that it be periodic with period $2d$). The symmetry of the potential is broken, however, with the application of applied force. This shifts the potential landscape, and is shown as a slope in Figure 16A. We define a positive force as one that assists RNAP, i.e. acts in the direction of elongation. Thus the potential shown in Figure 16A is for an assisting force experiment; in an opposing force experiment, the slope of the potential would be upward.

Note that the backtracked polymerase occupies a fundamentally different kinetic space than does an actively transcribing polymerase. This is represented by the kinetic scheme in Figure 16B, where the backtracked states are off the main elongation pathway.

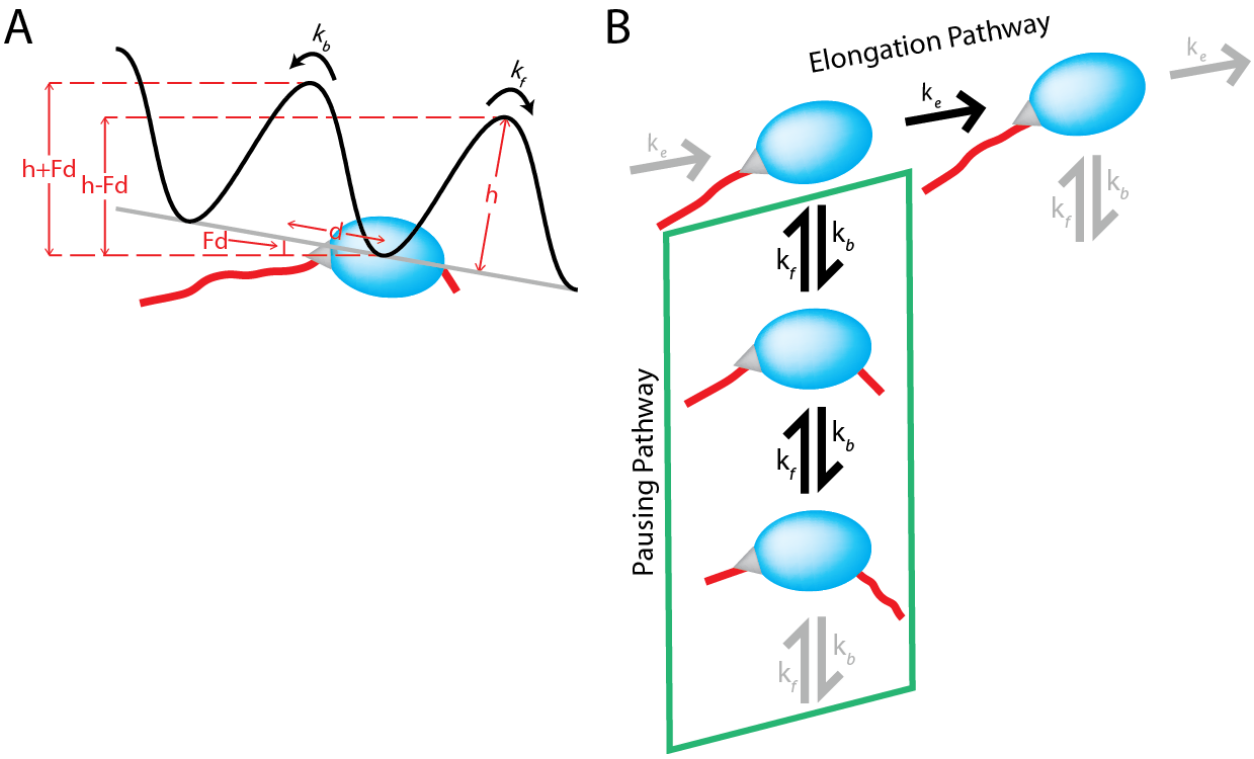

Figure 16: The backtracking model of transcriptional pausing. A. The potential landscape of a backtracked RNAP. The RNAP (blue) sits in a minimum of the potential (solid black line). The transcribed RNA (red) protrudes from both the exit and entry channel of the RNAP, signifying backtracking. The amplitude of the potential is h, and its period is $2d$ (one base pair). The forward and backward hopping rates between local minima are k_f and k_b, respectively. Application of an external force biases the potential, and is graphically represented by a slope of the potential (grey line). A positive (assisting) external force is shown here. **B.** The pausing pathway (boxed in green) is divergent from the main elongation pathway, and is governed by the rates of backward and forward hopping within the backtracked potential landscape as shown in A. Elongation is governed by the pause-free velocity, or elongation rate, k_e. Gray arrows indicate a continuation of their respective pathways.

The rate of hopping between local potential minima of the backtracked pathway of the potential shown in Figure 16A is exponentially dependent on the energy of the transition state[107]:

$$k_{f,b} = Ae^{(\pm Fd+h)/kT} = k_0 e^{\pm Fd/kT} \quad (3)$$

where k_f and k_b are the rates of forward and backward stepping of the polymerase during backtracking, respectfully, F is the applied force, k is the Boltzmann constant, and A is a

prefactor that is specific to the enzyme. The height of the potential, h is absorbed into this prefactor to yield k_0 – the intrinsic hopping rate of a particular enzyme within the potential under no force.

The trajectory of a backtracked polymerase, then, can be considered a biased random walk, where the probabilities of forward and backward steps are proportional to k_f and k_b, respectively. The return times of such a biased random walk, i.e. the time it takes for the walker to return to the same position, has been derived in the literature[14,82,83,104], and is given by the probability distribution

$$\psi(t) = \frac{1}{t}\sqrt{\frac{k_f}{k_b}}e^{-(k_f+k_b)t} I_1\left(2t\sqrt{k_f k_b}\right) \tag{4}$$

where t is the return time, $\psi(t)$ is the probability of such a return time, and I_1 is the modified Bessel function of the first kind. Because the duration of a pause is a return time – the time it takes for a polymerase to return to the state in which the 3' end of the RNA is in the active site – $\psi(t)$ is the probability of observing a pause of duration t.

The distribution of pause durations, $\psi(t)$, gives insight into one property of pausing – pause recovery. In other words, $\psi(t)$ characterizes a RNAP once it is in a pause – the boxed region in Figure 16B. Entry into pauses, however, is a somewhat different process. It occurs at the branch point between the pausing and main elongation pathways, and can be considered as a competition between the two:

$$P = \frac{k_b}{k_e + k_b} \tag{5}$$

where P is the pause density, or number of pauses per unit transcript length, and k_e is the elongation rate, or pause free velocity.

The Template Dependence of Pausing

Note that template dependence is not explicitly present in the above model of transcriptional pausing. However, sequence-specific pausing has been observed in a number of cases, in both bulk[108] and single molecule[109] assays. These assays have used specific pausing sequences, instead of (as we do below) forming a general model for transcriptional pausing based on nascent RNA structure.

A number of mechanisms of how sequence modulates transcriptional elongation have been investigated. This includes differences in base pair stacking of the 3' end of the nascent RNA with the template DNA[108], and a statistical analysis of the change in base stacking energies upon expansion and movement of the transcription bubble[110]. A computational study on transcriptional initiation also proposes that secondary structure of the nascent RNA aids in the transition from initiation to elongation[111].

Here we investigate further the mechanism by which sequence affects transcriptional pausing. We find that, in general, GC-rich sequences attenuate both the length and the number of pauses. We propose a modification of the current model of transcriptional

pausing to include an energy penalty for backwards movement of the RNAP during backtracking, and show that this model is consistent with our data. Finally, we propose that the sequence dependence of pausing is mediated through secondary structure of the nascent RNA as it exits the transcribing polymerase.

RESULTS AND DISCUSSION

As described in Chapter 2, we have established single-molecule assays to observe transcription by both the mitochondrial RNA polymerase, Rpo41, and the main nuclear polymerase, Pol II, in the yeast *Saccharomyces cerevisiae*. The details of both the experimental setup, as well as the investigation of the characteristics of transcription by Rpo41, an enzyme that had heretofore not been studied at the single molecule level, form the basis of that chapter. These experiments used a common DNA sequence as a template for transcription – a PCR product from the genomic DNA of the lambda bacteriophage. This DNA sequence was roughly random in its composition, with a mean GC content of 47%.

To elucidate the effect of base pair composition we performed single-molecule optical trapping experiments on three templates: the aforementioned "random" DNA originating from the lambda phage genome, an "AT-rich" template from the mitochondrial genome of *S. cerevisiae* (18% GC), and a "GC-rich" template from the genomic DNA of *Myxococcus xanthus* (70% GC). Sample traces of transcription by each enzyme on each template are shown in Figure 17.

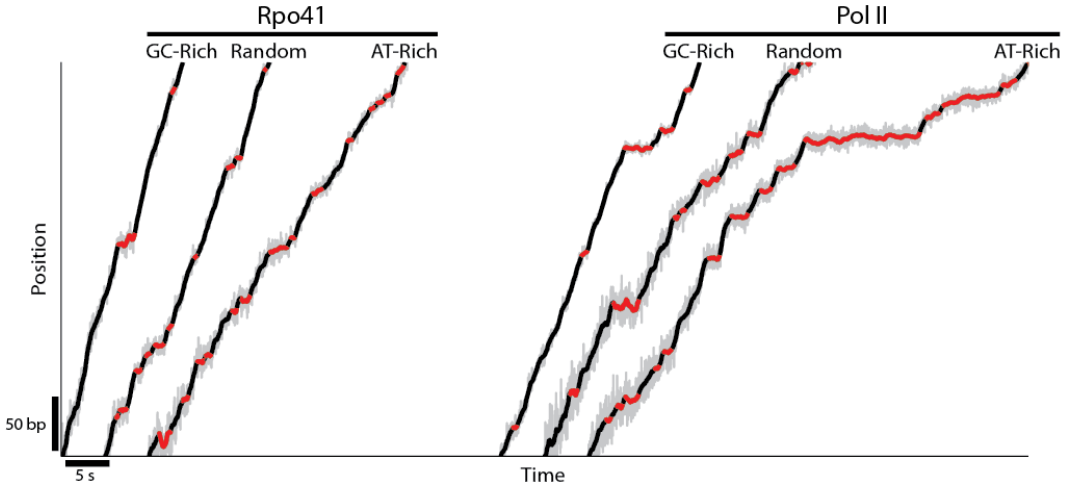

Figure 17: Template dependence of transcription. Traces of single-molecule transcription for Rpo41 and Pol II on GC-rich, random, and AT-rich templates as noted.

Inspection of the sample traces shown in Figure 17 indicates that, in general, transcription by Rpo41 is slightly faster than that by Pol II, and that Pol II also has longer pauses on a given template than Rpo41. The traces also indicate that, for a given enzyme, the number density of pauses on a given template is inversely proportional to the GC content of the template, as is the mean pause duration. Statistics of these characteristics of

transcription by the two polymerases on the three different templates are given in Table 4, and are described in more detail in the succeeding sections.

Table 4: Pausing characteristics for Rpo41 and Pol II on all templates in this study

	Rpo41			Pol II		
Template	AT-Rich	Random	GC-Rich	AT-Rich	Random	GC-Rich
% GC Transcribed (±SD)	23±2	35±3	67±2	24±2	37±2	68±2
Pause-Free Velocity (bp/s) (±SE)	22±3.5	24±3.5	24±5	18±2.3	19±2.7	17±2.5
Mean Pause Density (kB^{-1}) (± SE)	37±6	30±5	22±4	36±4	31±4	30±4
Mean Pause Duration (s) (±SE)	2.2±0.1	2.1±0.2	1.5±0.1	2.8±0.2	2.5±0.2	1.8±0.1
RNA Secondary Structure Energy Penalty, ΔG_{RNA} (kT)[a]	0.7±0.1	0.8±0.1	0.9±0.1	0.7±0.1	0.8±0.1	1.0±0.1
Intrinsic Hopping Rate, k_0 (s^{-1})[a]		5.4±1.8			2.9±0.3	

(a) Fitted parameter. See text for estimation of errors.

THE PAUSE-FREE VELOCITY IS TEMPLATE INDEPENDENT

As was discussed in the introduction to Chapter 2, the translocation step of elongation is really made up of a number of steps – the nucleotide addition cycle. This includes the thermal motion of the polymerase between the pre- and post-translocated steps, and stabilization of the post-translocated state upon NTP binding to the active site (the Brownian ratchet model)[112]. The oscillation of the polymerase between the pre- and post-translocated states, however, requires that the first few base pairs of the template downstream of the polymerase DNA be melted. The Brownian ratchet allows for thermal fluctuation of the downstream DNA between the single-stranded and double-stranded conformation, with the hopping of the RNAP into the post-translocated state being favored when the downstream template DNA is melted. Because the base stacking energy of GC base pairs is much higher than that of AT rich base pairs, it follows that this melted template conformation would be more favored in AT-rich DNA, and would thus result in faster rates of pause-free transcription.

Using the statistical algorithm discussed in Chapter 2 we have found the pause free velocity of Rpo41 to be 22±3.5 bp/s on AT-rich DNA, 25±3.5 bp/s on random DNA, and 24±5 bp/s on GC-rich DNA. The pause free velocities for Pol II are 18±2.5 bp/s on AT-rich DNA, 19±2.7 bp/s on random DNA, and 17±2.5 bp/s on GC-rich DNA (Figure 18). Thus, while the pause free velocity is consistently higher for Rpo41 than Pol II, there is no statistically significant difference between the pause free velocities of a given enzyme amongst templates.

The fact that pause-free velocities are template-independent strengthens the claim that RNAP translocation is not limited by the thermal fluctuations that drive the ratchet or open the downstream end of the transcription bubble. Instead, this finding suggests that the rate-limiting step comes after translocation; the slowest step during RNA synthesis could be the NTP hydrolysis, the new bond formation, or pyrophosphate release.

Note that the fluctuations between single-stranded and double-stranded template DNA also occur at the upstream end of the transcription bubble. In this case, the fluctuations

govern the closing of the bubble behind the transcribing polymerase, as opposed to the opening of it at the downstream end. Differences in base stacking interactions between the front end of the transcribing polymerase and the back end would then be reversed as the sequence at the front end of the polymerase becomes the sequence its back end. For example, the energy landscape of the backtracked polymerase is theoretically shifted to favor retrograde movement when a GC base pair is in front and AT base pair is behind the polymerase. However, the polymerase will eventually move forward, to the point that the GC base pair that was in front of the polymerase is now at its backside. Over the length of the transcript, these asymmetric base stacking interactions will nearly cancel. Thus, while there may be effects due to asymmetric base stacking interactions, they are well below the resolution of our experimental setup. This is discussed further below when considering different models of template modulation of pausing.

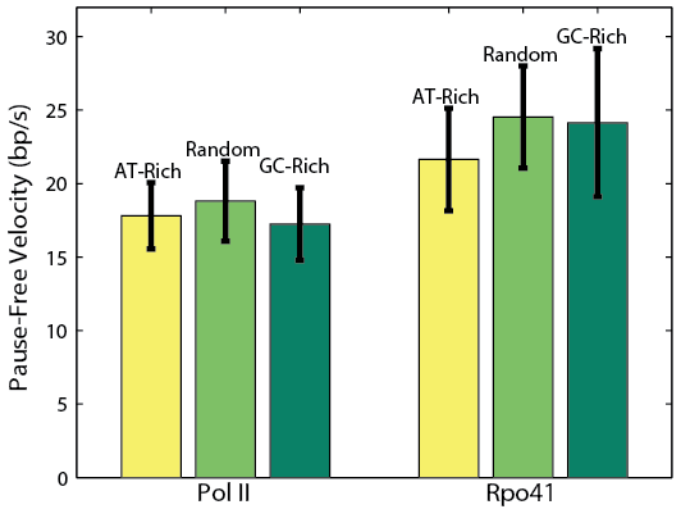

Figure 18: Pause free velocities are template independent. Pause free velocities were calculated by the fitting method described in Chapter 2. Errors are 95% confidence intervals of standard errors of bootstrapped means.

Determination of a Reliable Pause-Picking Algorithm

Analysis of pausing requires careful attention to the algorithm by which pauses are picked. We have used a method commonly used in the single-molecule transcriptional field[13,14,82,83], in which the following is done:

1. The raw position data (taken at 2 kHz) are averaged by decimation to 50 Hz.
2. The averaged data are filtered using a second order Savitsky-Golay filter with a time constant of 1 s.
3. The dwell times are measured from the filtered data by recording the time it takes the polymerase to advance by 3 bp.
4. The mean dwell time is calculated after the removal of outliers.

5. A pause threshold is calculated, which is a multiple of the mean dwell time. Regions of traces with dwell times longer than this pause threshold are considered pauses.

Optimization of this algorithm, most importantly the multiplier of the mean dwell time to determine the pause threshold, is a crucial component of pause analysis. We found that we simply could not pick all pauses accurately with a single multiplier. For example, for fast traces with little pausing – as is seen in Rpo41 transcription on GC-rich templates – the algorithm would find any slower region of the trace and deem it a pause, even if the velocity at those regions was in excess of 6 bp/s, the velocity resolution of our experimental setup. The above algorithm implicitly *requires* that some pauses be found, even if there are none.

Conversely, usage of the same multiplier used for traces that are slower and/or contain many pauses results in under-picking of the pauses, in which regions that are clearly pauses are deemed active transcription.

Because of the inability for a single computer algorithm to accurately pick *all* pauses for all types of traces in our data set, we chose only to consider pause durations above which we felt the computer algorithm was reasonably accurate at picking pauses and in which we could easily determine pauses by eye. This minimum pause duration is 0.5 s. All traces were visually inspected, and errors made by the pause picking algorithm were manually edited.

Note that, as determined by Equation (4), the fraction, F, of pauses above duration t is:

$$F(t) = \frac{\int_{t}^{\infty} \psi(t')dt'}{\int_{0}^{\infty} \psi(t')dt'} \tag{6}$$

For d = 0.5 bp, F = -5 pN, k_0 = 2 s^{-1}, and T = 25°C, only 0.09% of all pauses are expected to be above 30 s. This number increases to 0.7% for k_0 = 1 s^{-1}, and decreases to 5×10^{-4}% for k_0 = 5 s^{-1}. We therefore consider only pauses below 30 s, as those above 30 s can be considered extreme outliers.

We furthermore only considered data taken with an opposing force greater than 2 pN, above which the noise is low enough to reliably distinguish pauses longer than 0.5 s.

Note that the model presented in Equation (4) does predict a dependence of the pausing characteristics on force. However, because the distance to the transition state, d, is small, large forces are necessary to yield a significant bias to the forward and backward stepping rates. We therefore chose to analyze data under 10 pN, for at that force Fd = 1.8 pN-nm, which can still be considered small compared to kT (4.1 pN-nm).

In summary, all of the subsequent pause analysis performed uses pauses of duration between 0.5 and 30 s, and under forces between 2 and 10 pN. These limits on pause analysis are heretofore referred to as the "observation window."

Observation of transcription by single molecules of RNA polymerase in real time allows us to analyze the phases of the mechanochemical cycle of transcription independently. In contrast to the aforementioned template-independence of pause free velocity, we have seen that the properties of pausing are dependent on template composition.

The mean pause duration for Rpo41 was 2.2±0.1 s on AT-rich DNA, 2.1±0.2 s on random DNA, and 1.5±0.1 s on GC-rich DNA. The mean pause duration for Pol II was 2.8±0.2 s on AT-rich DNA, 2.5±0.2 s on random DNA, and 1.8±0.1 s on GC-rich DNA. Thus the mean pause durations are inversely correlated with template GC composition, as is shown in Figure 19A.

The mean pause density for Rpo41 (Figure 19AB) was 37±5 kB^{-1} for AT-rich DNA, 30±5 kB^{-1} for random DNA, and 22±4 kB^{-1} for GC-rich DNA, while for Pol II it was 36±kB^{-1} on AT-rich DNA, 31±4 kB^{-1} on random DNA, and 29±4 kB^{-1} on GC-rich DNA. Thus pause density has an inverse correlation with respect to template GC composition as well.

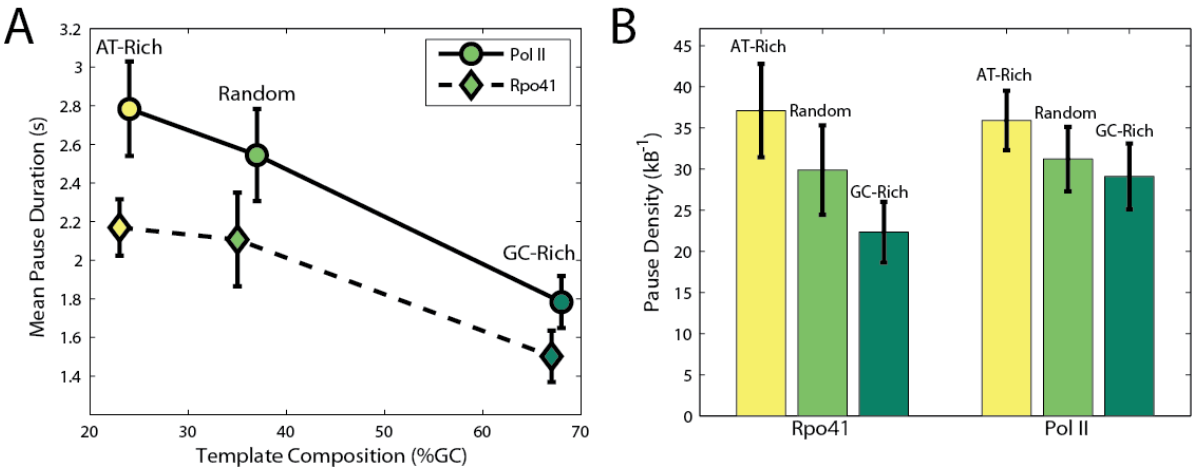

Figure 19: The mean pausing statistics are template dependent A. The average pause duration for Pol II (circles and solid line) and Rpo41 (diamonds and dashed line) on each template. The template composition is the run-length weighted mean GC content, and varies slightly for different enzymes on the same template due to differences in run lengths. **B.** The average pause density, or number of pauses per kilobase of template transcribed, for each enzyme and each template in this study. Errors are standard errors of the mean.

In addition to analysis of the mean values of the pause statistics, we analyzed the distribution of pause durations. We give the cumulative distribution of the pause durations for Rpo41 in Figure 20A and for Pol II in Figure 20B. We use cumulative distributions for two main reasons. First, it does not require the choice of a bin size, as the plotting of a normal distribution does. Binning can introduce artifacts or mask trends in the data. Second, the cumulative distribution is the basis of a statistical test by which we determine if two distributions are statistically different (discussed below).

Inspection of the cumulative distribution plots in Figure 20A&B indicate that, for both enzymes, increasing GC content shifts the distribution towards *shorter* pauses. This is indeed true to statistical significance, as is determined by performing a two-sample Kolmogorov-Smirnov (KS) test between the data sets[85]. This test utilizes the empirical

cumulative distributions (as plotted), and is nonparametric. Thus it makes no assumptions about the underlying distribution.

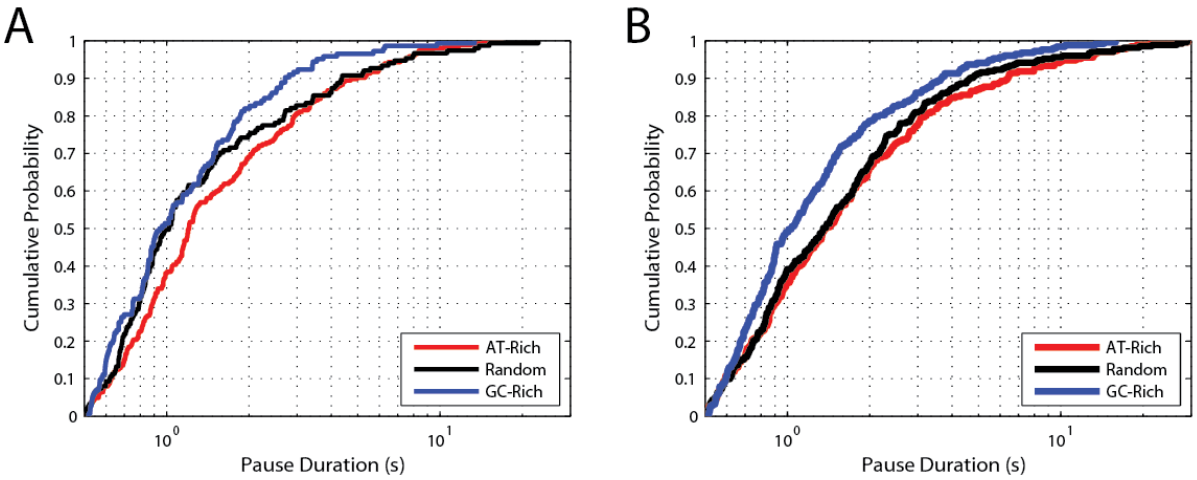

Figure 20: The distribution of pause durations are template dependent. A. Cumulative pause duration distribution for Rpo41 on all templates. **B.** Cumulative pause duration distribution for Pol II on all templates.

The *p*-values of the KS tests between distributions are given in Table 5. They indicate the pause durations of AT-rich and GC-rich arise from differing underlying distributions to greater than 95% confidence for both enzymes. Furthermore, the random data set differs significantly from the GC-rich data set in Rpo41, and AT-rich for Pol II. The disparity between the two enzymes in this regard may indicate differences in the sensitivity of the polymerases to template composition. However, the fact that two distributions do not pass the KS test does not indicate that they stem from the same distribution. Rather, it indicates that one cannot say with confidence that they differ. Thus it is entirely possible that the distributions do differ, but that these differences cannot be discerned statistically.

Table 5: *p*-values from KS tests between pause duration distributions[a]

		Rpo41			**Pol II**		
		AT-Rich	Random	GC-Rich	AT-Rich	Random	GC-Rich
Rpo41	AT-Rich	1	0.00878	0.010991	0.133765	0.260198	0.015989
	Random		1	0.320349	0.002101	0.005782	0.89092
	GC-Rich			1	0.002118	0.006328	0.7747
Pol II	AT-Rich				1	0.863263	0.000662
	Random					1	0.002263
	GC-Rich						1

(a) Values in red indicate 95% confidence (or greater) that the underlying distributions differ ($p \leq 0.05$).

MODELS FOR THE TEMPLATE DEPENDENCE OF TRANSCRIPTIONAL PAUSING

As described above, all of the pausing characteristics we have measured indicate that GC-rich content attenuates pause entry and shortens pause durations. We explored two possible mechanisms by which this could occur: changes in the backtracking potential due to base stacking effects on the moving transcription bubble, and from the effect of nascent RNA secondary structure.

A discussion of the effect of base stacking energies on the backtracked transcription bubble is aided by reference to Figure 21A. This figure represents a transcription bubble in the backtracked state. Note that forward movement by the bubble requires the breaking of one DNA/DNA bond and the creation of one DNA/RNA bond at the downstream end of the bubble. At the upstream end the opposite is true – one DNA/DNA bond is formed as the template DNA closes behind the RNAP, and one RNA/DNA bond is broken. Thus we expect a net zero change in free energy due to template base stacking energies upon movement of the backtracked transcription bubble.

The cancellation of base stacking interactions upon movement of the transcription bubble is broken, however, when one considers that the GC composition at the downstream end of the transcription bubble may be different than at the upstream end. This would impose an energetic bias for movement of the backtracked transcription bubble – a backward bias if the base pairs at the downstream end of the bubble are GC and the base pairs at the upstream end are AT, and a forward bias in the converse. However, these biases, on average, will occur equally as forward and reverse biases, and thus the average energy bias due to base stacking interactions over the entire template is expected to be small. We calculated these energy biases using published values for the base stacking energies of DNA/RNA[113] and DNA/DNA[114] duplexes. A histogram of these calculated energy biases on the templates used in this study is shown in Figure 21B. Note that the histogram is symmetric around zero, as expected. Indeed, the average energy penalties over each template were on the order of 10^{-4} kT for all templates. This is strong evidence that these base stacking energy penalties are not the main cause of the observed template dependence of pausing.

A second model for the mechanism of the template dependence of pausing is shown graphically in Figure 21C. Because the transcribed RNA is single stranded, it follows that it would be able to form secondary structures with itself upstream of the RNAP. Should such RNA structure be present immediately upstream of the transcribing polymerase, it would serve as a barrier to backtracking, and thus would result in a decrease in pause entry. Structure further upstream would serve as a physical barrier to maximal backtracking, and thus would shorten pause durations. Again, because the base stacking energies of GC base pairs are greater than those of AT (or AU) base pairs, one would expect that transcripts resulting from GC-rich DNA would have more secondary structure and thus pause entry and pause durations would be proportionately disfavored in such a template.

We calculated the energies of formation of secondary structure within our templates by calculating the lowest energy of the secondary structure of the 30 bp of RNA trailing the transcribing polymerase[115,116]. A histogram of those energies, denoted ΔG_{SS}, is shown in Figure 21D. It shows a clear difference between the distributions of energies of formation amongst templates. Indeed, the average energy of formation for the AT-rich template was

2.0±1.8 kT, whereas it was 3.4±2.9 kT on random DNA and 8.1±3.0 kT on the GC-rich template.

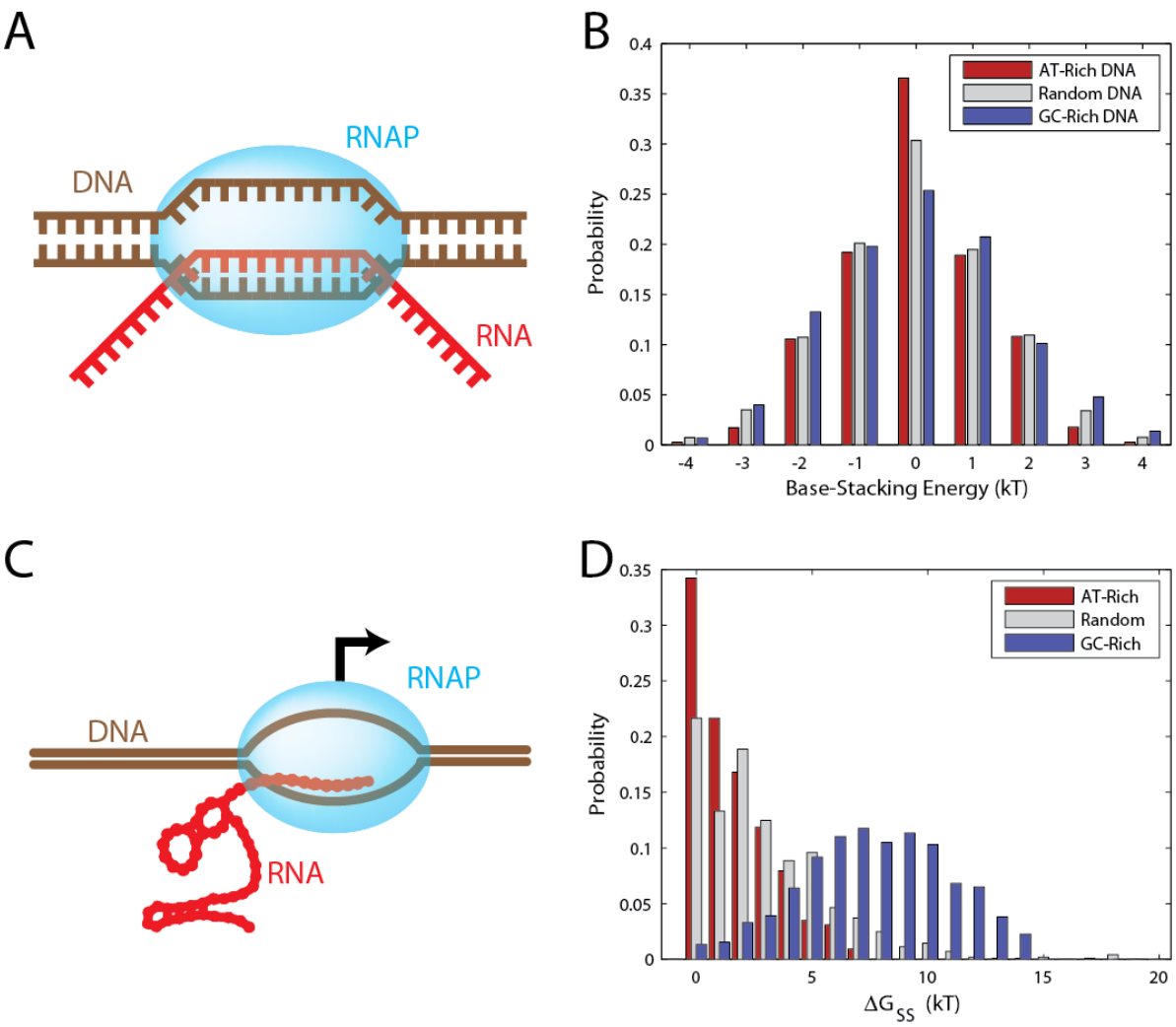

Figure 21: Models for template dependence of transcriptional pausing. A. A model of the backtracked polymerase and transcription bubble. Note that, due to symmetry, a net zero number of DNA/RNA and DNA/DNA bonds are formed during translocation of the bubble. **B.** Histogram of energy penalties that occur when the change in base-stacking energies at the front of the bubble are different than those at the back of the bubble, due to differences in sequence. The change in energies were calculated over the entire template. See text for details. **C.** A model of the effect of nascent RNA structure on transcription. The RNA in this figure was created by L. Bintu, and is used with permission. **D.** Histogram of the energies of formation of secondary structure (ΔG_{SS}) within the templates used in this study. Energies were calculated using a 30 bp sliding window along each template. See text for details.

The contention that RNA secondary structure modulates the pausing behavior of RNA polymerases requires a modification of the existing model of transcriptional pausing by backtracking. No longer can the backtracked potential be considered symmetric. Rather this model adds an increase to the energy barrier of backwards hopping, as is depicted in Figure 22. This added energy is not the same as ΔG_{SS}, which is an average of the energies of formation of 30 bp secondary structures of the transcribed RNA. Thus we define a new energy penalty, ΔG_{RNA}, which is defined as the average energy added to the potential landscape upstream of the RNAP due to RNA effects.

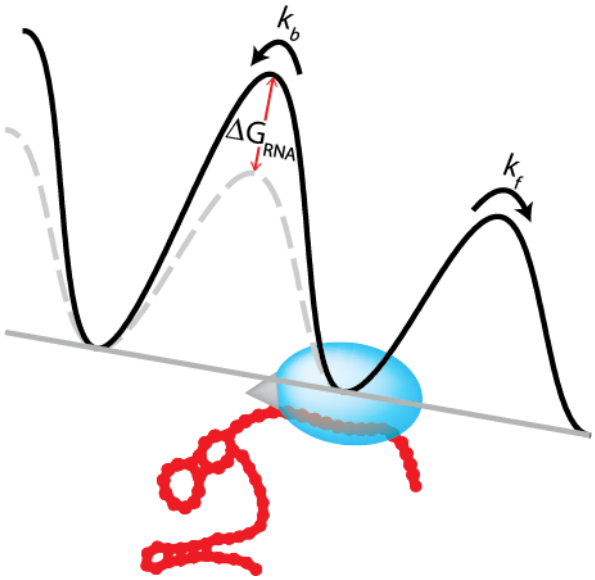

Figure 22: A modified backtracking potential. The proposed model incorporates and energy barrier, ΔG_{RNA}, to the backwards translocation during backtracking. This energy barrier is due to the energy needed to disrupt base-stacking interactions of secondary structure of the nascent RNA upstream of the transcription bubble. Other annotations are as described in Figure 16.

This introduction of ΔG_{RNA} into the current model of pausing by backtracking is simple. We introduce it solely as an added term in the exponential of the backward hopping rate of the polymerase within the backtracked potential,

$$k_b = k_0 e^{-(Fd + \Delta G_{RNA})/kT} \tag{7}$$

This simple modification of the existing pausing model has significant consequences in that it allows us to elucidate the enzyme-specific determinants of pausing (k_0) with those that arise from sequence affects (ΔG_{RNA}). Indeed, this modification allows us to fit, for the first time, the pausing behavior of a RNAP transcribing on different templates with a single value for k_0.

A Method for Fitting the Template-Dependent Pausing Model with Observed Data

Here we describe the process of fitting our model of transcriptional pausing to the observed data. This task is nontrivial, as the distribution of pause durations, $\psi(t)$, is a nonlinear function, containing both an exponential and a Bessel function.

The addition of a second parameter, ΔG_{RNA}, into the pausing model turns the parameter space from one dimension to two. The pausing characteristics, then, can be considered three dimensional surfaces over that parameter space. Figure 23A shows one such surface: the theoretical pause density (brown) of a polymerase within the observation window, as a function of ΔG_{RNA} and k_0. The functional form of this surface is simply Equation (5), corrected to account for the missed pauses resulting from the use of a finite observation window[81],

$$\begin{aligned} P_{obs} &= P\int_{t_1}^{t_2}\psi(t)dt \\ &= \frac{k_b(k_0,\Delta G_{RNA})}{k_e + k_b(k_0,\Delta G_{RNA})}\int_{t_1}^{t_2}\psi(t,k_0,\Delta G_{RNA})dt \end{aligned} \tag{8}$$

where t_1 and t_2 define the bounds of the observation window (in our studies, 0.5 s and 30 s, respectfully).

Overlaid on the surface plot of the pause density in Figure 23A are the values for the bounds of the observed pause density for a particular enzyme on a particular template, i.e. the mean plus or minus the standard error (the example shown in Figure 23 uses Rpo41 on AT-rich DNA). Because they are of a single value, they are represented as planes over the parameter space. The intersections of these planes with the theoretical surface define the bounds within which the pausing model would predict the pause duration to be within the error of the observed value. These intersections are curves in the parameter space, and are heretofore referred to as the "contours."

The mean pause density can also be plotted as a surface over the parameter space, and is shown in Figure 23B. The functional form of the mean pause density within the observation window is

$$\left\langle t_{obs}\right\rangle = \frac{\int_{t_1}^{t_2} t\cdot\psi(t,k_0,\Delta G_{RNA})dt}{\int_{t_1}^{t_2}\psi(t,k_0,\Delta G_{RNA})dt} \tag{9}$$

The intersections between the mean pause density surface and the planes defined by the observed mean pause durations define a second set of contours. The two sets of contours – one from pause density overlaps, and one from mean pause duration overlaps – intersect each other as shown in Figure 23C. The region defined by the intersections of these curves, referred to here as the contour overlap, is the region within which the pausing model predicts the mean values of pause density and pause duration for that enzyme on that template.

The contour overlaps for transcription by Rpo41 on all templates are shown in Figure 23D, while those of Pol II are shown in Figure 23D. The constraint that the value of k_0 is enzyme specific (and template independent) is shown graphically as a shaded region in each figure, spanning the constrained values of k_0. This is the region in which the pausing model predicts the observed mean pause characteristics for all templates using a single k_0.

The combined contour overlaps for both polymerases on all templates is shown in Figure 23E and displays a number of properties of transcription discussed above. First, note that in general the possible values for ΔG_{RNA} are similar between the enzymes on the same template. Likewise, the values of k_0 are similar between templates for the same enzyme. This serves to support the notion that the addition of ΔG_{RNA} *to* the backward hopping rate in the pausing model does indeed separate the enzyme specific determinants of pausing from those of the template.

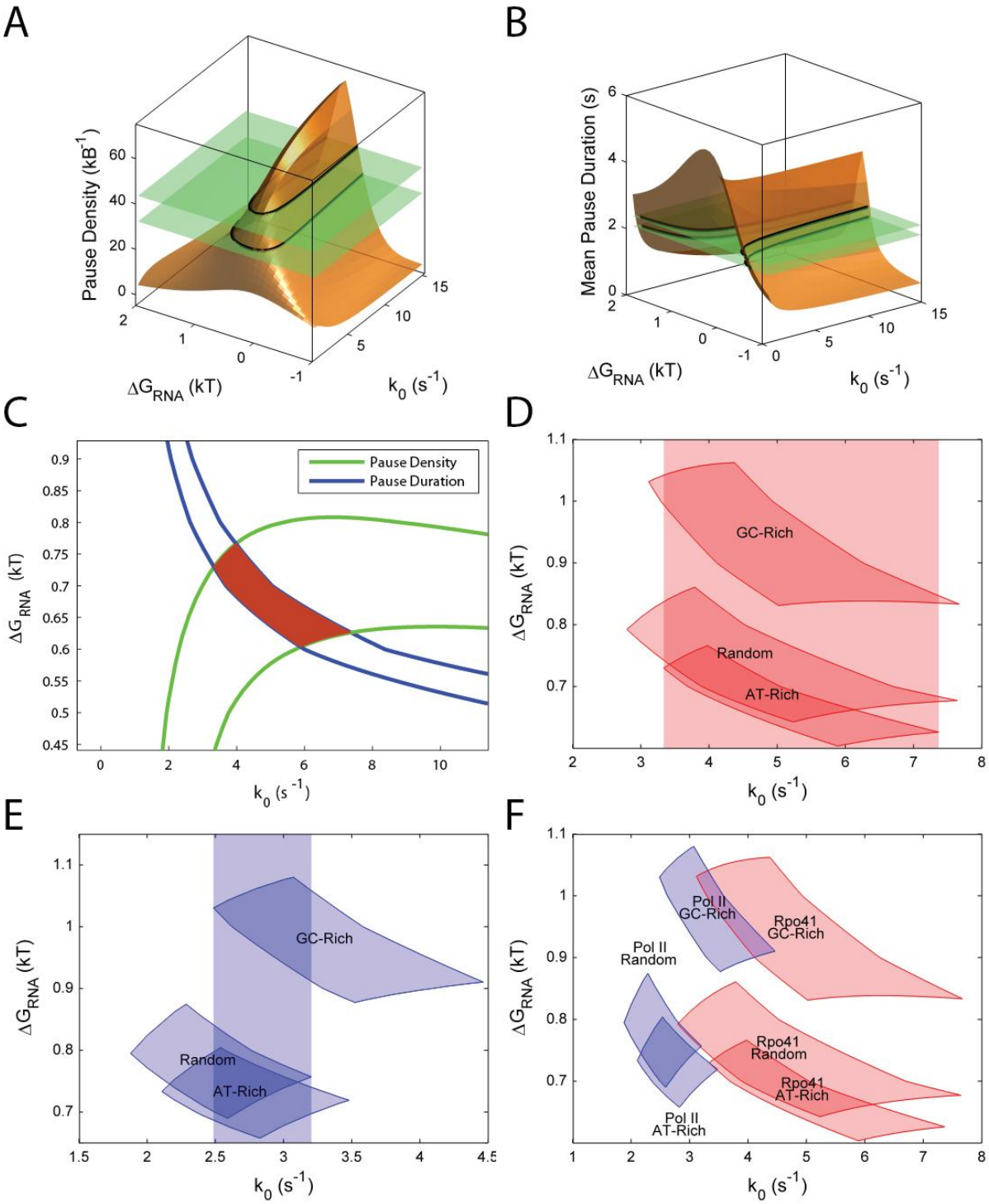

Figure 23: Fitting of the model of transcriptional pausing to mean values of observed pausing parameters. A. The range of observed pause densities (green planes) overlaid with the theoretical distribution of pause densities (brown). **B.** The observed mean pause duration (green planes) overlaid with the theoretical distribution of pause durations (brown). In both A&B, the upper plane is the upper limit of the observed parameter (mean

plus standard error of the mean), the lower plane is the lower limit (mean minus standard error of the mean), and the theoretical surface has been corrected to account for the finite observation window of the observed values. The intersections between surfaces are shown as black curves ("contours"). **C.** The combination of the contours from pause density (green curves) and pause duration (blue curves) define a space (a "contour overlap", brown) within which values of the pausing parameters (k_0 and ΔG_{RNA}) correspond to those observed. **D.** The contour overlaps for transcription by Rpo41 on all templates. **E.** The contour overlaps for transcription by Pol II on all templates. In both D&E the boxed region defines the area in which the mean pause density and mean pause duration for all templates can be fit by a single, enzyme-specific value for k_0. F. The contour overlaps for Rpo41 (red) and Pol II (blue) on all templates.

We used the contour overlaps to define the bounds for the fitting of the pause duration distribution. Specifically, we performed the KS test over all of the values for k_0 and ΔG_{RNA} contained in the regions defined by the fitting of the mean values as described above. This further defined k_0 and ΔG_{RNA}. The regions in which the values of k_0 and ΔG_{RNA} fit both the mean values and the pause duration distributions is shown in Figure 24A. These fits result in the values of k_0 to be determined to be 5.4±1.8 s^{-1} for Rpo41 and 2.9±0.3 s^{-1} for Pol II. The values for ΔG_{RNA} for Rpo41 are 0.7±0.1 kT for AT-rich DNA, 0.8±0.1 kT for random DNA and 0.9±0.1 for GC-rich DNA. For Pol II these values are 0.7±0.1 kT for AT-rich DNA, 0.8±0.1 kT for random DNA, and 1.0±0.1 for GC-rich DNA. Constraining the values of ΔG_{RNA} further by requiring them to be the same between enzymes for a given template results in enzyme-independent values of ΔG_{RNA} to be 0.7±0.1 kT for AT-rich DNA, 0.8±0.1 kT for random DNA, and 1.0±0.1 kT for GC-rich DNA.

In Figure 24A&B we plot the theoretical cumulative distributions of pause durations along with the empirical distributions. They show that the fitting algorithm described above passes the ultimate test of its validity – the model fits the data by eye.

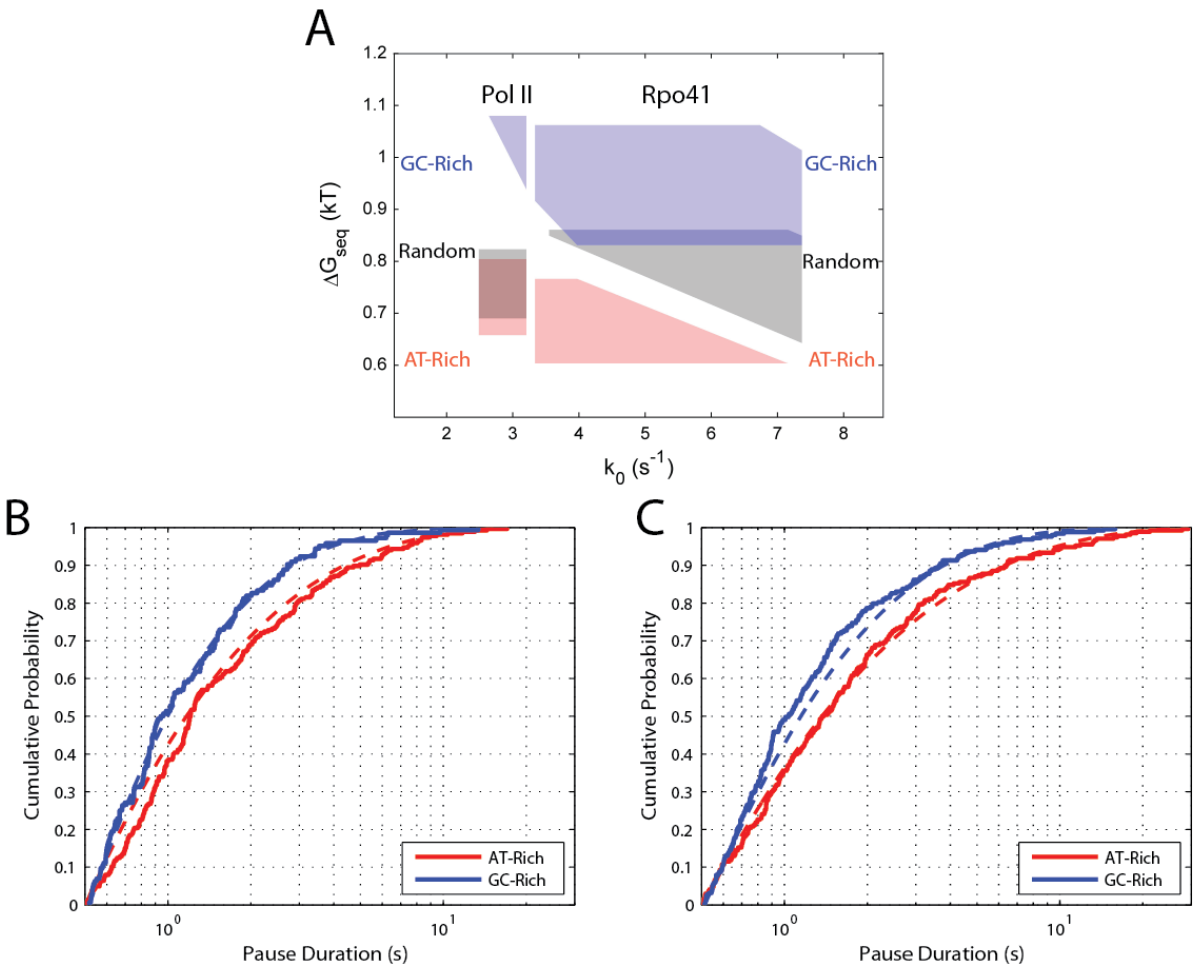

Figure 24: Fitting of the model of transcriptional pausing to the distribution of pause durations. A. Values of the pausing parameters that resulted in statistically-significant similarity between the pausing parameters of the model described in the text (k_0 and ΔG) for each polymerase and template. The bounds of this fitting were constrained to the contour overlaps discussed in the text and shown in Figure 23D. **B.** The fitted (dashed lines) and observed (solid lines) cumulative distributions of pause durations for Rpo41 on AT-Rich and GC-Rich DNA. **C.** The fitted and observed cumulative distributions of pause densities for Pol II on AT-Rich and GC-Rich DNA. The theory fits the data for Random DNA as well, but is excluded in both B and C for clarity.

Disruption of RNA Structure Abolishes Template-Dependent Pausing

We have shown in the above sections that the introduction of an energy penalty term to the backward hopping rate of backtracking allows for a fitting of both the mean pause characteristics and the distribution of pause durations using a single enzyme-specific parameter and a single template-specific parameter. Our introduction of this term was motivated by the theory that backward movement of the backtracked polymerase is disfavored due to secondary structure of the nascent RNA chain. We have not, however, *proven* that the RNA is the cause of this energy penalty. In order to do so, we have

performed the same single-molecule transcription assays as has been described in this chapter and the last, with Pol II on the same templates. However, in these experiments we have added RNase A to the reaction buffer. RNase A specifically hydrolyses after C and U residues and, in our buffer conditions, does so without specificity between single stranded and double stranded forms of the RNA[117]. The introduction of RNase A into the reaction buffer, then, should attenuate the effects of RNA secondary structure on pausing.

Plots of the mean pause duration and pause density for Pol II with and without the presence of RNase A in the buffer are shown in Figure 25A&B respectfully. Note that the presence of RNase abolishes the template dependence of these properties. Furthermore, in the presence of RNase these values tend towards those on AT-rich DNA without RNase.

The cumulative distributions of pause durations for Pol II on all templates, with and without RNase in the buffer, are shown in Figure 25C. Those same distributions are shown in Figure 25D, only the random data sets have been removed for clarity. Once again transcription on any template in the presence of RNase is similar to that on AT-rich without RNase. This confirmed by KS tests between the data sets, the *p*-values of which are shown in Table 6.

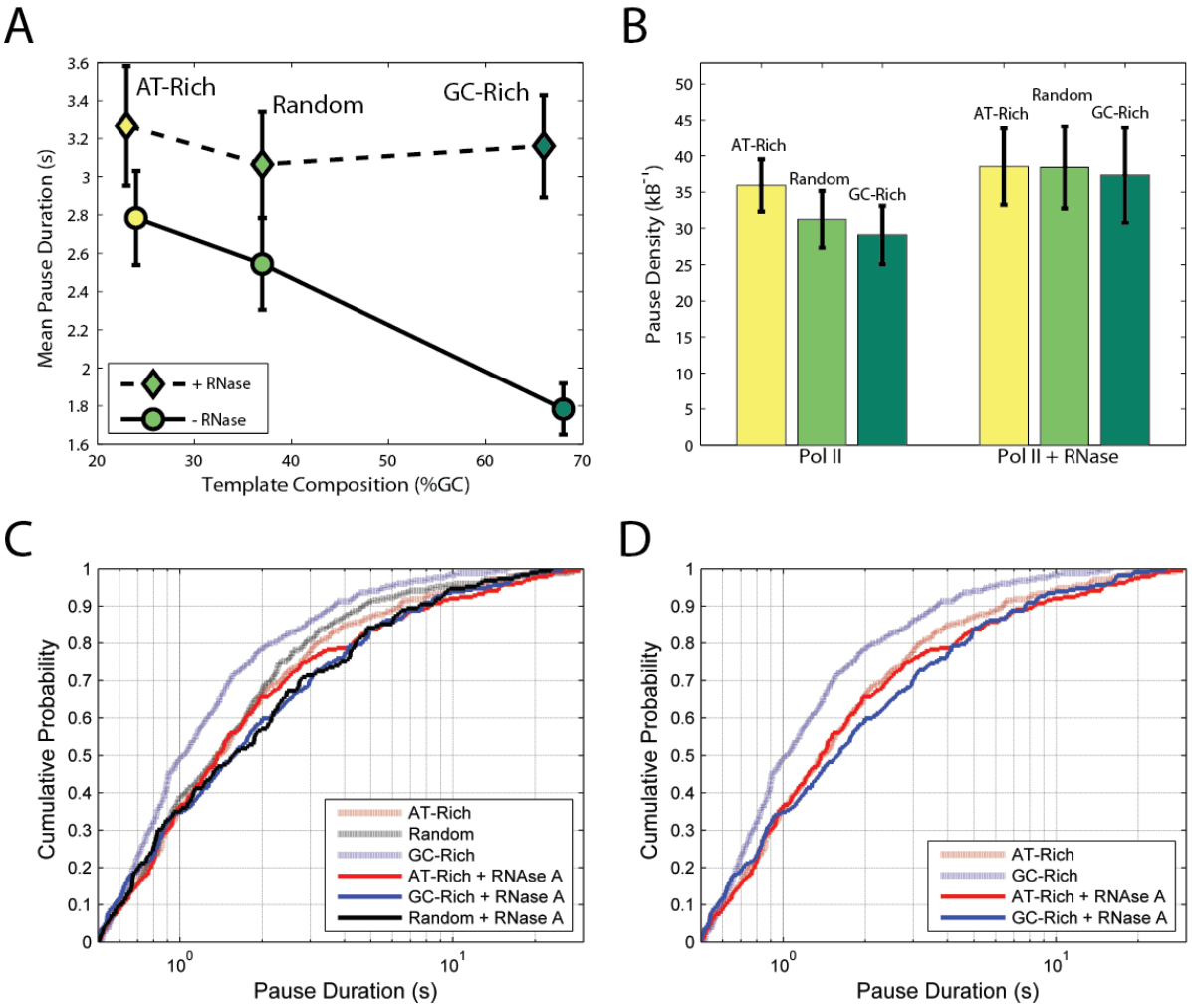

Figure 25: The use of RNase abolishes the sequence dependence of pausing. A. The mean pause durations for transcription by Pol II on all templates, with and without RNase A in the reaction buffer (solid and dashed lines, respectfully). **B.** The mean pause densities for transcription by Pol II on all templates, with and without RNase A in the reaction buffer. **C.** The cumulative distribution of pauses Pol II on all templates, with and without RNase A in the reaction buffer (dark and light curves, respectfully). **D.** The same cumulative distributions as in A, but with the random data sets removed for clarity.

Table 6: *p*-values from KS tests between pause duration distributions for Pol II on all templates, with and without RNase[a]

		+ RNase		
		AT-Rich	Random	GC-Rich
- RNase	AT-Rich	0.63887	0.151868	0.257674
	Random	0.21986	0.058761	0.032762
	GC-Rich	0.00257	1×10^{-5}	1×10^{-5}

(a) Values in red indicate 95% confidence (or greater) that the underlying distributions differ.

The fact that the characteristics of pausing in the presence of RNase are similar to that on AT-rich templates, as opposed to GC-rich templates, necessitates further discussion. Before we had taken any data, we concluded that we could not predict *a priori* which direction the presence of RNase would take the pausing data. Indeed, if RNA degradation was complete and fast compared to elongation, one would expect that RNase would cleave the entire nascent RNA as it exits the polymerase, and therefore there should be no backtracking and no pausing. This, of course, is much more similar to transcription on a GC-rich template than on an AT-rich one.

However, if the kinetics of RNA degradation by RNase A are too slow to fully degrade the nascent RNA as it exits the RNAP, or if steric interactions between the two enzymes prevent such degradation, some amount of RNA is left upstream of the transcription bubble. Incomplete degradation may therefore relax the secondary structure of the nascent RNA, but still leave enough of it for backtracking to occur. This would result in transcription similar to that on AT-rich DNA than on GC-rich DNA.

In hindsight, it is much more plausible that transcription in the presence of RNase would be similar to that on AT-rich DNA than on GC-rich, since theoretically the presence of even one extra RNA base pair upstream of the polymerase would be enough for pause entry. Likewise, the vast majority of pauses backtrack only a few base pairs, and therefore contribute the most to the pause duration distribution and mean. Histograms of the observed backtrack distances above 4 bp are shown in Figure 26A for Pol II on AT-rich DNA. These data fit an exponential with a decay constant of 0.381 (R^2 = 0.9947), which implies that 75% of all backtracks are 4 bp or less, and 95% of all backtracks are 8 bp or less.

Figure 26B shows the distribution of backtracking distances for Pol II on GC-rich DNA with and without RNase. Note that the data set that includes RNase has more backtracks than the dataset without RNase for backtracks less than 12 bp, supporting our model in which the secondary structure of a short region exiting the polymerase is relaxed in the presence of RNase.

Furthermore, the dataset containing RNase does not have any data points above the 12 bp bin. Indeed, the maximum backtrack distance observed for the data set in which RNase was in the buffer was at the center of that bin – 12 bp. For data taken without RNase in the buffer, the maximum backtrack distance was 52 bp. The average maximum backtrack distance amongst all templates in the presence of RNase was 17±2 bp. This is consistent with bulk data, which have shown that 17-19 RNA base pairs are protected from degradation during transcription[118]. Thus it is likely that steric interactions between the RNAP and the RNase limit the amount of degradation of the transcribed RNA, resulting in pausing that is similar to that on an the AT-rich template and in the absence of RNase.

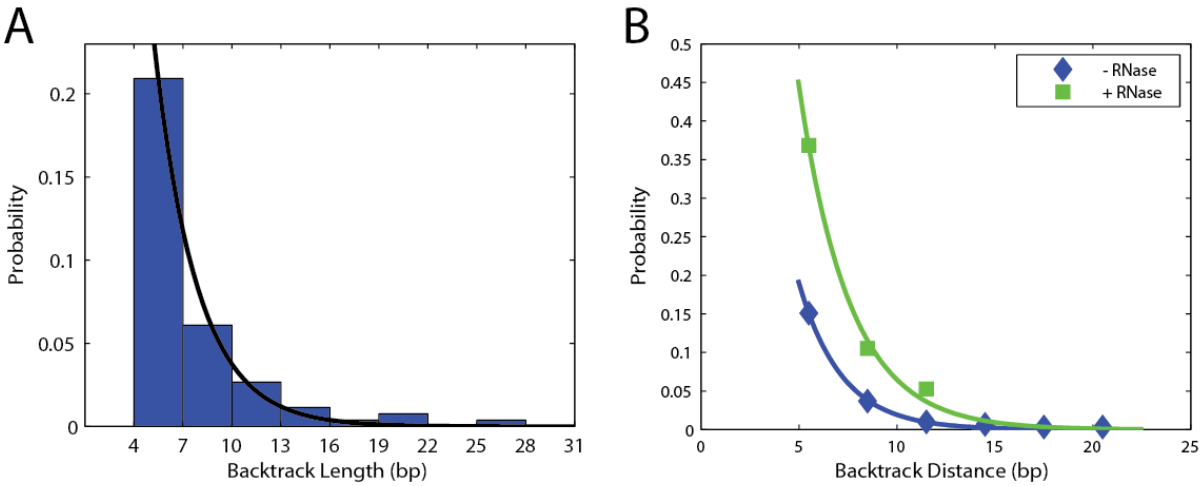

Figure 26: Histograms of backtrack distances. A. Data (blue) for transcription by Pol II on AT-rich DNA, with a fit of the data to an exponential (solid line). **B.** Comparison of histograms for transcription by Pol II on GC-rich DNA with (green squares) and without (blue diamonds) RNase A in the reaction buffer. A fit of each data set to an exponential is shown as a solid line in the same color as the data. In both A and B, the probabilities given are the probability of observing a backtrack with a length as given on the x-axis. Only backtracks of length greater than 4 bp were considered; the probabilities are normalized as such.

Exclusion of RNA/RNAP Interactions as the Mechanism for Template Dependence

We believe that we have proven, to a high degree of certainty, that the secondary structure of the nascent RNA modulates transcriptional pausing. However, note that we have not formally excluded the possibility that interactions with the RNA and the RNAP protein itself may also pose an energy barrier to backward movement during backtracking, and contribute to the effects we have discussed in this chapter. However, addition of oligonucleotides complimentary to regions of the nascent RNA has been shown to attenuate backtracking and stabilize active transcription[119]. This is the opposite of what one would expect if the mechanism of pause attenuation is mediated through RNA/protein interactions, as the hybridization of the introduced oligonucleotides would presumably compete with the RNA/protein interactions. The more likely explanation is that the introduced of oligonucleotides form RNA/DNA hybrids that are analogous to the secondary structure we propose is formed by the RNA during transcription. These base-pairing interactions, be they DNA/RNA or RNA/RNA, attenuate transcription by adding an energy barrier to backward movement of the backtracked RNAP.

Also note that we have shown that the phenomenon of modulation of pausing by template GC composition occurs in both Pol II and Rpo41. These enzymes are evolutionarily distinct: Rpo41 has homology to RNAPs of the T-odd bacteriophage family and is a single-subunit enzyme, while Pol II is a multi-subunit, eukaryotic enzyme[19,91]. There is no discernable homology between the two enzymes. Thus it is more likely that the mechanism of pause modulation between the enzymes is due to the things that were the

same in these experiments – the sequence – than the things that are drastically different – the polymerases.

Future Directions

Incorporation and Validation of The Model With Other Experiments

A number of studies have been reported in the literature on the backtracking nature of transcription, and others are underway. Collaborative efforts should be established to determine if those other studies can confirm this newest model of template dependence by RNA secondary structure. These experiments include the work of transcription by Pol II through the nucleosome, which has shown sequence-specific pausing[81] (in progress). Other studies on sequence-specific pausing should be incorporated as well[109].

Base-Pair Resolution Studies

The work we have cited here has focused on aggregate data of pause characteristics, be it distributions of the pause durations, or the mean values of pause durations and pause densities. This is because we cannot align our traces to base pair, or even near base pair accuracy. Recent analysis of transcription by Pol II through the nucleosome used the transcript runoff as a second data point (in addition to the start site) to align traces, with significant improvement in the positional accuracy of their data[83]. Unfortunately, studies using opposing force, such as this one, do not have runoffs. Thus either an alternative method for calibration, or a switch to an assisting force geometry may help to improve the positional accuracy of experiments studying template dependent pausing phenomena.

Further advances in optical tweezers technology have resulted in base pair resolution[71,72]. Transfer of these studies to one of these new instruments, in addition to the aforementioned implementation of a second calibration point, may lead to further elucidation of the effect of nascent RNA secondary structure on transcription at the base-pair scale.

A Priori *Prediction of Pausing Sites*

In the above study we give a very simple prediction of the effect of template sequence on the properties of pausing: we calculated the energies of formation of RNA secondary structure in 30 bp windows along the template. This, we believe, gives a reasonable qualitative estimate of the differences in energy biases backtracked polymerases encounter during transcriptional pausing amongst templates. We make no claims that these calculated energies are the energies needed for pausing by the polymerase on a given template, just that, in aggregate, different templates impose differing average barriers to backtracking.

We are currently developing computer algorithms to predict the actual energy barriers encountered by the RNAP at a given site on the template. This includes consideration of the kinetics of both folding and transcriptional elongation, and allows for cotranscriptional folding of the RNA as is it synthesized[120]. Such simulations would greatly strengthen analysis of data taken with base pair accuracy, and may also strengthen any arguments to

distinguish the effects of RNA secondary structure from RNA/protein interactions as described above.

Materials and Methods

All of the single-molecule assays described in this chapter were identical to those described in Chapter 2. We refer you to the Materials and Methods section of that chapter. Here we describe only the preparation of the two new templates: GC-rich and AT-rich.

The AT-rich DNA was prepared by PCR using mitochondrial DNA from *Saccharomyces cerevisiae* strain BY4743. Whole cell genomic preparations of this strain were prepared as per common menthods[121]. The PCR reaction consisted of 1×Thermolpol Buffer (New England Biolabs, Ipswitch, MA), 0.2 μM primer mito-BstAPI-fwd (5'-CAGATTGTACTGAGAGTGCACCATATGCATGTTATTTTCAGGATTCAGCAACACC-3'), 0.2 μM primer dig-cox2-16 (5'-Dig-CCCATGGGGTCCCTCACTCCTTACG-3'), 200 μM each dNTP, 1 μL genomic DNA/100 μL final reaction volume, and 0.025 U/μL Taq DNA polymerase (NEB). The thermocycling conditions consisted of an initial denaturation step of 95°C for 5 min., followed by 30 cycles of 95°C for 30 s and 60°C for 7 min., followed by a final elongation step of 60°C for 10 min. This reaction was cleaned using the Qiagen PCR purification kit (Valencia, CA) and concentrated. The DNA was then cut at a concentration of 100 ng/μL using 0.1 U/μL BstAPI (NEB) in 1×NEB buffer 4, and 0.1 μg/μL BSA. The digestion reaction was performed at 60°C for 4 hours, followed by heat inactivation by incubation at 80°C for 20 min.

The GC-rich DNA was a PCR product using a whole-cell genomic preparation of *Myxococcus xanthus*, (courtesy of D. Zusman) as a template. It consisted of 1×MasterAmp PCR Premix J (Epicentre Biotechnologies, Madison, WI), 0.5 μM primer dig-ZproR (5'-Dig-GAATTCTTGTCCAACTCGTCCACCA-3'), 0.5 μM primer xanthus-draIII-fwd (5'-AATATTCACCATGTGCGGGATCCAGCGTCGGGTCCTCATCGTCG-3'), 0.1 U/μL Platinum Taq (Invitrogen, Carlsbad, CA), and 1 μL genomic DNA/100 μL final reaction volume. The thermocycling conditions consisted of an initial denaturation step of 95°C for 5 min., followed by 30 cycles of 95°C for 30 s, 57°C for 1 min., and 72°C for 5 min., followed by a final elongation step of 72°C for 10 min. This reaction was cleaned using the Qiagen PCR purification kit and concentrated. The DNA was then cut at a concentration of 100 ng/μL using 0.1 U/μL DraIII (NEB) in 1×NEB buffer 3, and 0.1 μg/μL BSA. The digestion reaction was performed at 37°C for one hour, followed by heat inactivation by incubation at 65°C for 20 min. All templates were cleaned using the PCR purification kit after digestion.

Chapter 4: Creation of a Transcriptionally Independent Mitochondrion

Introduction

The Need for a Minimal Organism

As the field of synthetic biology matures, as more complicated and complex recombinant pathways are being generated in host organisms, as fluxes of metabolites are being modulated to high precision, one big problem looms: the "background issue." Metabolic engineers often grapple with the innate metabolic processes that are going on in the host organism. They struggle with fluxes, as alternative pathways can divert intermediate metabolites, decreasing yields, as well as toxicity issues from those same intermediate (or ultimate) metabolites. These problems have taken many a researcher on long and winding paths, relying on a host of techniques such as elimination of competing pathways[122,123] or substrate sequestration[124] as a means to increase yields. The number of hours and dollars lost to these background effects is likely underestimated by the literature, since those most challenging projects – in which yields oftentimes never reach a substantial amount – are often not published, and are relegated to the lab notebooks of history.

The field of synthetic biology, then, needs a minimal cell. It needs an organism – a so called "chassis" – with defined metabolic pathways. In this way, future metabolic engineers will know of competing pathways or toxic intermediates while designing their experiments, rather than finding out about them *post facto*.

The creation of a minimal organism in turn begs a different question: what are the minimal genes necessary for life? This question, though less practical and more academic than the one of the creation of a well-defined chassis, has been debated for many years now, and explored through many methods.

Indeed, biology is now in the midst of a full-scale effort to define the minimal genes essential for life. These studies either utilize mutagenesis of existing free-living organisms[125-127], comparative genomics and network analysis[128,129], or *de novo* reconstruction of biological processes[130,131]. Each of these methods has its benefits and detractions. Minimizing the genome of a free-living organism, for example, has resulted in data that varies depending on the organism studied. Using transposon mutagenesis of the human pathogen *Mycoplasma genitalium*, the organism with the smallest genome that can still be cultured, researchers at the J. Craig Venter institute identified 382 protein-coding genes that are "essential"[125]. More recently, a team at the University of Cambridge used the same technique in *Staphylococcus aureus*, and found 351 genes that are important for growth in culture[127].

The calculations involved in this reductionist approach are made more difficult due to such biological phenomena as redundant and coordinated metabolic pathways, as well as an inability to score a slow growing culture as viable. This is further exemplified by the fact

that the results of comparative genomics and network analysis do not agree with those of the reductionist approach.

In contrast to the reductionist approach lies the *de novo* approach, i.e. one in which a minimal cell is created from the fundamental chemical parts that make up all biological systems. Much progress has been made in this regard in recent years, most using lipid vesicles as a "bag" within which biological processes could be mimicked. Both transcription[132] and translation[133] have been achieved within lipid vesicles, as have more complicated processes, such as self-encoded gene replication using a replicase[134].

A cell constructed purely *de novo* would be a milestone of science, if not humanity. Unfortunately, creation of such an organism would require a simultaneous construction of all of the processes necessary for said organism to function. Furthermore, these processes must be highly coordinated; it is not sufficient to merely place every necessary element of a living system into one compartment. Rather, each element must be inserted in the correct order, at the appropriate time. In that regard, even those who have been attempting the *de novo* or lipid-vesicle approach admit that the full scale reconstruction of a cell using this approach is nowhere near possible today:

> *The "bottom-up" preparation of an artificial cell – defined here as a system with all the known characteristic properties of a contemporary biological cell… – by assembling individual molecular components is not possible today. The same is true for the "bottom-up preparation" of a minimal cell - again this minimal cell being defined in the sense of a cell as we know it today, equipped with a minimal size genome*[135]*.*

We have taken an intermediate approach. Instead of attempting to create a bag, a genome, and all of the accessory factors that are necessary for their propagation, we start with an evolutionarily derived bag and genome – a mitochondrion – that also contains, at any instant, all of the accessory factors necessary for its continued existence. However, in contrast to an already viable organism, the mitochondrion is not a free-living entity; it is not "alive." Instead, it imports the vast majority of the accessory factors necessary for its propagation from a "biological incubator" we call the host-cell cytoplasm. Indeed, the mitochondrion is a true intracellular symbiont, taking from its host proteins, small molecules, etc., and giving to it, for example, cofactors, iron sulfur clusters, reducing equivalents, and most famously energy in the form of ATP[136].

Because the mitochondrial genome is incapable of propagating the mitochondria alone, one can consider modifying this genome as an intermediate approach to the creation of a minimal organism. It is more top-down than the *de novo* approach, since it begins with an evolutionarily derived and very highly coordinated system, yet more bottom-up than the reductionist approach, since an isolated mitochondrion is incapable of self-propagation and is therefore not considered "alive." This middle-ground approach, then, can be summed as follows:

> *We aim to define the minimal genes necessary for life, and in the course of doing so, establish a minimal organism within which a new generation of*

metabolic engineering and synthetic biology experiments can be performed, by the stepwise addition of essential genes to the mitochondrial genome, and the subsequent silencing or deletion of that gene's nuclear counterpart, until we have created a self-propagating, free-living organism, at least within a maximally supportive medium.

The definition of a maximally supportive medium, of course, requires some discussion. Indeed, a mitochondrion could be considered to already be living in a maximally supportive media – the cytoplasm, in which case the proposed project is trivial. We do not accept this definition, but do acknowledge that the definition of a maximally supportive medium is somewhat subjective and vague. Nonetheless, we define it as one that supplies all of the nucleic acids, amino acids, salts, and carbon sources for our minimal organism to live, but not whole proteins, cofactors, etc. Thus we require our minimal organism to be capable of the myriad of metabolic and catabolic processes within the field of biochemistry, most notably replication, transcription, translation, and division.

Where to Begin

As Lao-tzu purportedly stated ca. 2500 years ago, "a journey of a thousand miles begins with a single step." So, too, does the journey to create a synthetic cell out of a mitochondrion. To embark on this journey, then, one must ask: what is the first process we would like to introduce into the mitochondria, or, more specifically, what is the first pathway we would like to encode within the mitochondrial genome? Even this question is hopelessly vague, for a pathway is made up of many reactions, each often catalyzed by a different enzyme that is encoded by a separate gene. The question then becomes which should be the first gene to be placed into the mitochondrial genome, and where should it be placed? To some extent, the answer to such a question is arbitrary; since the creation of a minimal organism, by definition, requires addition of all the genes essential for life, it does not matter, in theory, which pathway to choose first. However, there are some practical considerations.

Some processes have already been inserted into the mitochondrial genome. One step in the arginine biosynthetic pathway – the gene *ARG8* – has been inserted into the mitochondrial genome of *Saccharomyces cerevisiae*[58], as has the gene encoding for the Rieske iron sulfur protein, Rip1[137], a subunit of the cytochrome c1 complex that is involved in electron transport from ubiquinol to cytochrome c1[138]. Thus it is possible to insert genes into the mitochondrial genome and have them expressed to the phenotypic level. Neither of these genes, however, are necessary for the survival of the mitochondrion *per se*. Instead they are essential for the wild-type function of the organism as a whole. In the case of *ARG8*, the entire cell – cytoplasm, nucleus, mitochondria, etc. – cannot survive without the amino acid arginine in the culture medium. In the case of *RIP1*, the whole cell cannot survive on a medium lacking fermentable carbon (i.e. a medium that requires the cell to derive its ATP from oxidative phosphorylation, which requires active mitochondria, as opposed to fermentation, which occurs in the cytoplasm). In both cases the deletion of each gene would not prevent the mitochondria of the cells from continuing to support all of the

processes necessary for its propagation; rather it would disrupt processes that are needed for survival of the cell on a particular medium.

It makes sense, then, to start with a core process that is necessary for the mitochondria to propagate. The obvious processes to consider, then, are those at the center of the central dogma of molecular biology: transcription, translation, and replication.

Creation of a mitochondrion that is independent of any of the core processes of the central dogma is an undoubtedly ambitious goal. Each process, by its very nature, is highly coordinated, and tightly regulated. The cell is highly resistant to modifications of any of those pathways. But there are levels of complexity even amongst these three processes.

Translation, for example, relies on the ribosome as its central player. This massive complex is composed multiple RNA and ribosomal proteins, and is dependent on a myriad of other proteins for initiation, termination, and elongation of translation[139]. When considered in this light, even such a complex process like transcription seems simpler.

Replication, as well, is much simpler than translation. In the mitochondria, however, replication and transcription are tightly linked in that the mitochondrial polymerase performs both transcription and primes replication[64]. Thus any investigation into the replicative independence of the mitochondrion must start with transcription.

Transcription, then, is a logical first choice as the initial pathway to create within mitochondria. It is complicated and interesting enough to merit further investigation, necessary for the establishment of a minimal organism, and simple enough in terms of components to be a valid proof of principle for the greater project.

CHOOSING AN ORGANISM

In addition to deciding on the initial process to insert into the mitochondrial genome, one must also consider which organism to work in. One must take into account the varied complexity amongst the taxa, as well as the availability of basic biological techniques within a given organism, previous work in the mitochondrial field for that organism, as well as potential applications for the research, i.e. its potential impact.

Mammals

Doing our work in a mammalian organism, especially in humans, would certainly have the most potential impact, for the establishment of mitochondrial technologies and the elucidation of mitochondrial processes in humans can have a direct effect on global health and wellbeing. Indeed, more than 130 pathogenic mtDNA mutations have been discovered[140,141], including those having been shown to have a direct correlation with diseases such as MELAS[142], NARP[143], LHON[144] (see the table of abbreviations for the full names of these diseases), type 2 diabetes[145], and Leigh's syndrome[146]. Mitochondrial diseases affecting the oxidative phosphorylation pathway alone are symptomatic in approximately 1 in 7600 people[147].

However, the impact of a particular vein of research is not the only criterion for pursuing it. Its feasibility must also be considered. Transgenic mitochondrial genetics in

mammalian systems suffers from a number of experimental barriers to its implementation, as well as drawbacks from other systems.

The doubling time of mammalian cells is on the order of days (as opposed to hours in *S. cerevisiae*), and because the time dependence of culture densities depends exponentially on the inverse of the doubling time, any experiment using mammalian cells takes orders of magnitude longer as compared to those in yeast.

There are fewer tools available for nuclear genetic manipulations of mammalian systems, including the creation of genetic knockouts – just one technique crucial to this project. Even the tools that do exist are much more complicated to implement than in lower eukaryotes.

The very nature of the mammalian mitochondrial genome offers challenges to its manipulation. Transcription of the genome occurs in only two primary transcripts – one in each direction – covering almost the entire mitochondrial genome (16,295 bp in mouse[148] and 16,569 bp in humans[149]). The genome is also incredibly compact, with many overlapping reading frames, and only two major non-coding regions[149]. Modification of the mitochondrial genome without disrupting any endogenous mitochondrial genes can only be performed in these two regions.

One noncoding region is located in a tRNA-rich section of the genome, is within the primary transcript, and contains the origin of replication for the "light" strand of the genome (due to differences in base-pair composition, the two strands have different sedimentation coefficients in $CsCl_2$ gradients, hence the designation "light" strand and "heavy" strand[150]). Thus modification of this region may adversely affect both expression of endogenous mitochondrial genes as well as coordination of replication of the mitochondrial genome.

The other untranscribed region of the mammalian mitochondrial genome occurs between the two promoters, and is just upstream of the origin of replication for the "heavy" strand. It is there that we have designed and cloned constructs which maintain the structure and sequence of the wild-type mammalian mitochondrial genome, but add inward-facing promoters between them, as depicted in Figure 27.

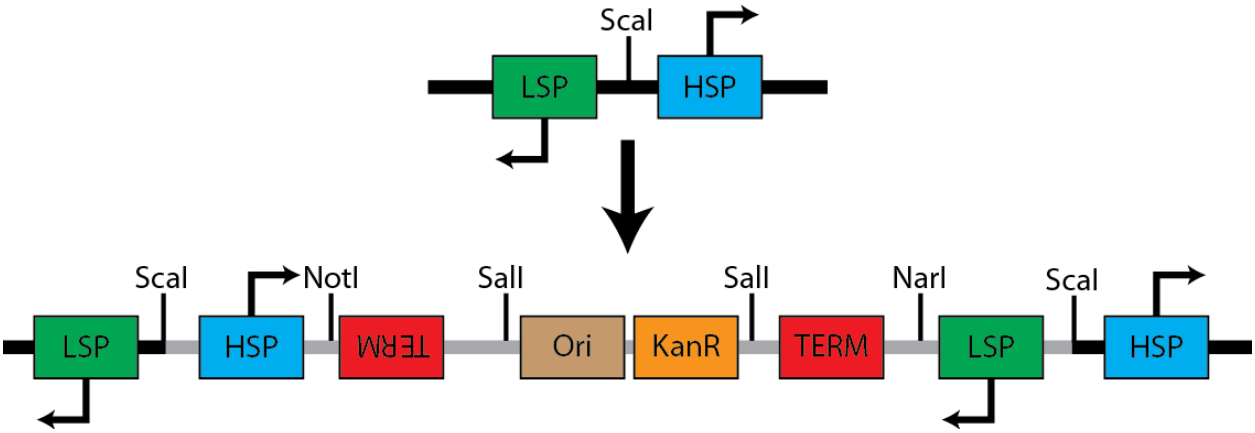

Figure 27: Creation of expression constructs for the mammalian mitochondrial genome. The ScaI restriction site lies in a unique noncoding region of the mammalian mitochondrial genome (top), between the heavy-strand promoter (HSP, blue), and the

light-strand promoter (LSP, green). The direction of the arrows extending from a particular promoter indicates the direction of transcription from that promoter. Insertion the region designated by the grey line into the ScaI site of the mouse mitochondrial genome creates a modified mitochondrial genome with two inward-facing promoters (bottom). Heterologous genes can then be inserted in the NotI or NarI restriction sites between the introduced promoters and the transcriptional terminators (TERM, red). Terminators prevent transcription past the modified region. An *E. coli* origin of replication (Ori, brown) and kanamycin resistance gene (KanR, orange), allow for the propagation of the entire mitochondrial genome in *E. coli*[151], and can be excised by restriction digestion with SalI and subsequent ligation.

We have created the construct depicted in Figure 27, as well as derivations of it and its insertion into versions of the mouse mitochondrial genome subcloned in *E. coli*[151]. This was done in parallel to other projects in the lab, which aimed at creating methods for transformation of isolated and/or *in vivo* mammalian mitochondria.

Efficient transformation of mammalian mitochondria has only been reported a few times in the literature, and in no case independently reproduced outside of the reporting lab. We have tried many of these techniques – including the use of protein transduction domains (peptides that can purportedly permeable carry cargo across membranes)[152-156], high-velocity microprojectile bombardment[157], and lipid mediated vesicles[158-160] – with no success. This is in addition to numerous other labs that, due to the implications such a technique would have on the medical and foundational research fields, have spent countless hours attempting to transform mammalian mitochondria. This fact, that *there currently exists no efficient and reproducible method for transforming mammalian mitochondria*, is ultimately the main disadvantage for the selection of mammalian cells as our system.

Yeast

In contrast to the mammalian system, the use of the baker's yeast *Saccharomyces cerevisiae* has a number of distinct advantages that make it the logical choice for transgenic mitochondrial studies. The main advantage is that yeast mitochondria have been transformed *in vivo* using high velocity microprojectile bombardment (described in detail below), and that this technique has been accomplished repeatedly, by a number of labs, with clear protocols available[58,137,161-166].

S. cerevisiae is a facultative anaerobe, meaning it can produce energy through respiration via the oxidative phosphorylation pathway present in the mitochondria, or through fermentation in the cytoplasm. This provides a distinct advantage when manipulating the mitochondria of yeast instead of other obligate aerobes such as mammalian systems: strains in which the mitochondria are nonfunctional can be propagated in facultative anaerobic systems, while they cannot in obligate aerobes. Furthermore, the status of respiration (and therefore mitochondrial function), can be easily scored in yeast. Those with defective mitochondria ("respiratory deficient") cannot utilize carbon sources that require respiration (glycerol, ethanol, etc.) and can only grow if the culture media contains fermentable carbon (e.g. glucose).

There can be few systems better than *S. cerevisiae* upon which to perform conventional genetic manipulations, as it is has been a model system in the genetic and molecular biology field for decades. Crosses, selections, transformations and other techniques are all well established. The creation of genetic knockout strains and mutants is relatively simple, and a collection of all viable knockouts is commercially available[167] (as well as conditional knockouts of those that are not viable).

Mitochondrial studies also have a long history in yeast. Indeed, the discovery of slow growing colonies (termed "petites") was first discovered in *S. cerevisiae* in 1949[168], and led to the first studies of respiration and mitochondrial biogenesis[169]. Thus studies of mitochondrial genetics in yeast is more developed than in most other organisms. Consequently, there are a number of mitochondrial mutants available, as well as techniques for conventional manipulation of mitochondrial genomes[170].

A graphic map of the yeast mitochondrial genome is given as Figure 28. At over 85 kB[53], the yeast mitochondrial genome is much different than that in mammals. While polycistronic transcripts are ubiquitous, there are still a number of primary transcripts (as opposed to just two in mammals). Furthermore, the majority of the yeast mitochondrial genome contains non-coding DNA. Thus there is ample space for the insertion of heterologous genes.

Finally, the doubling time of wild-type haploid yeast is approximately 100 minutes[171], allowing for growth, modification, recovery, and re-growth of a given strain in a few days.

The combination of highly developed techniques for genetic and genomic manipulation, an established *in vivo* mitochondrial transformation procedure, the ability to grow nonrespiring strains, and the space for genetic modifications of the mitochondrial genome, make baker's yeast an unavoidable choice for a system in which to create a minimal cell from a mitochondrion.

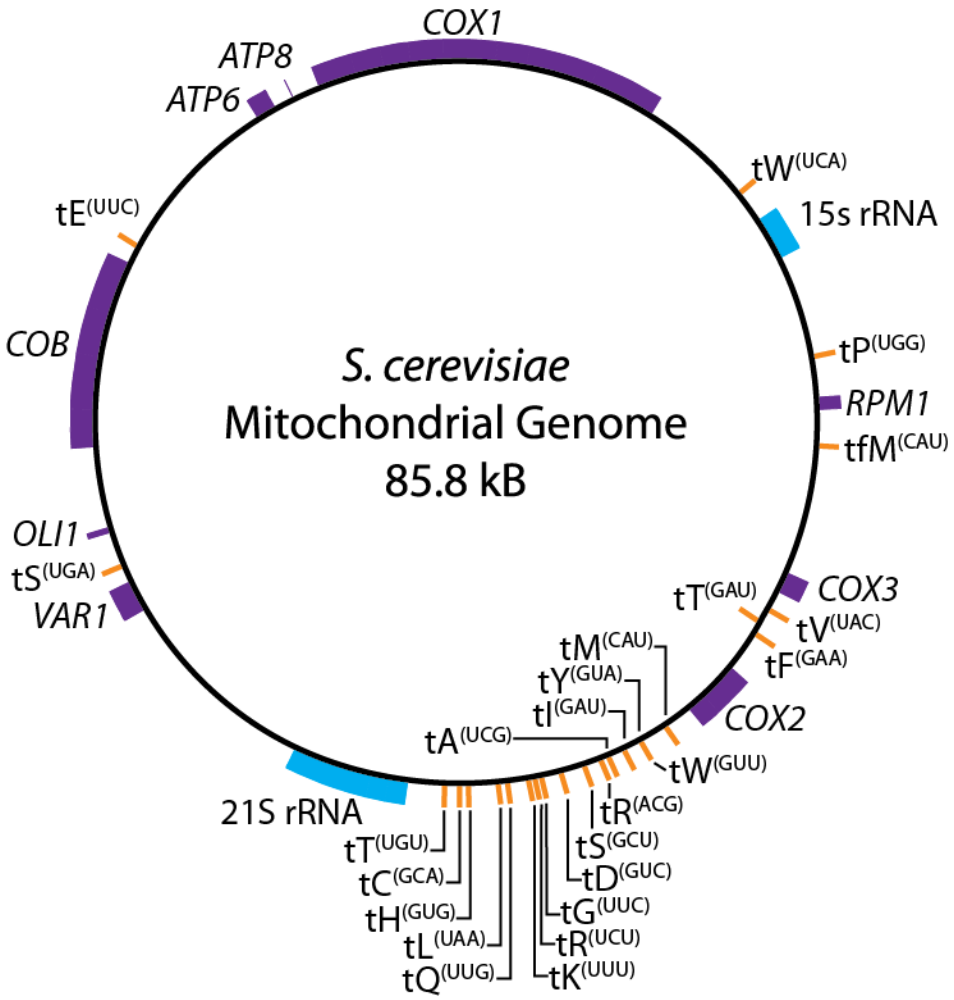

Figure 28: The mitochondrial genome of *Saccharomyces cerevisiae*. Protein-encoding genes are shown in purple, ribosomal RNAs in blue, and transfer RNAs in orange. The transfer RNAs are denoted by a lowercase t, followed by the single-letter code for the amino acid they attach to, and the codon that they base pair to as a superscript.

Gene Expression In Yeast Mitochondria

The creation of a transcriptionally independent mitochondrion in yeast requires an intimate understanding of the organization of the mitochondrial genome, as well as the central molecular players and mechanisms involved in gene expression.

A detailed discussion of transcription in yeast mitochondria, including its role in gene regulation, is given in the introduction of Chapter 2. Below we restate, in bullet form, those properties of gene expression in yeast mitochondria that are most important for expression of Rpo41 in the yeast mitochondrion.

- Initiation of transcription occurs at specific promoters within the mitochondrial genome[54]. The mitochondrial transcription factor, Mtf1, is responsible for promoter melting and open complex stabilization, but dissociates from the transcription complex during the transition from initiation to elongation[25,44,45].

- Rpo41 alone performs transcriptional elongation[45].
- Regulation of the transcription occurs via a multitude of mechanisms, including proximal and distal cis-acting DNA elements[51-53], and trans-acting factors[56,57]. These can mediate nucleo-mitochondrial interactions, as well as other external signals (e.g. glucose repression).
- The majority of primary transcripts in the mitochondria are polycistronic. They are cleaved post-transcriptionally[54].
- There is a lack of correlation between transcription rates and protein levels, indicating that transcription is not the rate limiting step of gene expression in the mitochondria[58].
- Gene expression is regulated through the RNA processing step, including through modulation of the mitochondrial degradosome[50,59,60].
- Gene expression is also regulated at translation. This is mediated through nuclearly-encoded translational activators, which tether specific mRNAs to the inner side of the mitochondrial membrane, where they can associate with the mitochondrial ribosome[62]. This allows for co-translational insertion of proteins into the membrane[61].
- Rpo41 also serves as a primer for replication in mitochondria[64-66]. Rpo41-deficient strains are incapable of maintaining wild-type mitochondrial genomes[50,67-69]. However, certain mitochondrial DNA fragments are stable in Rpo41-deficient strains, a property called hypersuppressiveness[67].

A Note on Nomenclature

Geneticists in a multitude of organisms have developed nomenclature systems, which aim to distinguish when one is talking about a particular gene or gene variant, its protein product, and the phenotypic consequences of its modification. In this chapter we make extensive use of the standard nomenclature utilized by yeast geneticists. Those unfamiliar with said nomenclature, or who become confused in the succeeding text, should refer to this section periodically.

Genes, in general, are given in italic font. Wild-type genes are given in uppercase with no suffix, while alleles (versions of the gene with sequences that differ from the wild type) are denoted with a dash and the enumerated allele number. For example, the wild-type gene encoding for the second subunit of cytochrome oxidase is denoted *COX2*, whereas one allele (in this example defined as a specific deletion in the 5' region of the gene) is denoted *cox2-62*. The fact that the *cox2-62* allele is recessive (*i.e.* would not be phenotypically manifest when present with a wild-type allele in a diploid) is denoted by the writing the allele name in lowercase.

Oftentimes the particular identity of a nonfunctional allele is not important, and the allele enumeration is left off. For example, one could write the *cox2-62* allele as simply *cox2*, in which case the author would be leaving out the specific identity of the allele, but nonetheless would be conveying the fact that a particular strain is recessive. Oftentimes it is presumed that the recessive allele is nonfunctional.

It is common practice in yeast genetics to remove all or part of a gene from the genome. This is denoted with an uppercase Greek letter *delta*. For example, the deletion of *RPO41*, the yeast mitochondrial RNA polymerase, would be denoted *rpo41Δ*. Deletions are often performed by replacing the given locus with another gene (oftentimes a selectable marker) by homologous recombination. In this case, the disrupting gene is often denoted by placing its name after the deleted gene's name, separated by a double colon symbol. For example, the deletion of *RPO41* with the *KanMX4* cassette, which confers kanamycin resistance, is denoted *rpo41Δ::KanMX4*.

Table 7: Phenotypes used in this study

Description	Wild-Type	Mutant	Mutant phenotype
Resistance to the antibiotic G418	$G418^S$	$G418^R$	$G418^s$ strains cannot grown on media containing G418.
Argenine auxotrophy	Arg^+	arg^-	arg^- strains cannot grow on media lacking argenine.
Uracil auxotrophy	Ura^+	ura^-	ura^- strains cannot grow on media lacking argenine.
Respiratory proficiency	Mit^+	mit^-	mit^- strains have nonfunctional mitochondria, and cannot grow on media containing only nonfermentable carbon as a carbon source.

Fused genes are denoted by a dash symbol, e.g. *ARG8-GFP* would denote a genetic fusion between the *ARG8* gene and the *GFP* gene, both being wild-type in sequence.

Protein names are not italicized, and are denoted in title case (*i.e.* only first letter capitalized). For example, the protein product of the *RPO41* gene is Rpo41. Others within the yeast genetics community put a "p" at the end of a protein name, e.g. Rpo41p. We do not use this convention.

Phenotypes are not italicized. Auxotrophies are in title case, with a superscript "+" at the end of the word, whereas prototrophies are lowercase, with a superscript "-" at the end of the word. For example, arginine prototrophs would be denoted Arg^+, whereas arginine auxotrophs would be denoted arg^-. Specific phenotypic designations used in this text are given in Table 7.

The status of the mitochondrial genome is designated by a superscript over the Greek letter *rho*. Strains with intact or nearly intact mitochondrial genomes are designated ρ^+, strains with large deletions in the mitochondrial genome are designated ρ^-, and strains in which the entire mitochondrial DNA has been removed are designated ρ^0. Note that the demarcation of a strain as being ρ^+ says nothing about its mitochondrial function, whereas ρ^- and ρ^0 strains implicitly have no or greatly reduced mitochondrial function. In the case of ρ^+ strains, then, one must resort to the Mit^+/mit^- phenotypic designation as described in Table 7.

There is one nomenclature rule specific to the very small number of scientists who aim to insert genes into the mitochondrial genome. Because the mitochondrial genetic code differs from the standard code (such as that in the nucleus), yeast mitochondrial geneticists denote genes that have been modified for mitochondrial expression with the letter "m," often in superscript, as a suffix. A particularly important example in this work is the difference between *RPO41* and $RPO41^{m}$. Both (theoretically) encode for the exact same protein when expressed in the nucleus and the mitochondria, respectively. The two sequences, however, are drastically different, and neither would result in functional protein if expressed in each other's compartment.

Mitochondrial Transformation

With the choice of initial gene (*RPO41*) and the organism (*S. cerevisiae*) justified, we now discuss the technique of transformation of yeast mitochondria by high-velocity microprojectile bombardment.

The hardware involved in the physical transformation procedure is shown in Figure 29A&B. In this procedure, cargo DNA is precipitated onto 600 nm tungsten particles ("microcarriers"), and shot into the bombardment strain using high pressure helium. More specifically, the microcarriers are deposited on plastic discs ("macrocarriers"), and these discs are placed into a holder within the evacuated chamber of the gene gun system. Above the mounted macrocarrier is a gas tube sealed with a plastic seal (a "rupture disc"). When helium gas is flown into the gas tube, the pressure rises until the pressure within the tube exceeds the failure pressure of the rupture disc, upon which time the disc bursts, and a helium shockwave is sent out from the gas tube into the evacuated chamber (Figure 29C). This shock wave meets the macrocarrier in its holder, and rapidly accelerates it toward an agar plate containing a thickly spread lawn of bombardment cells. On the way to the agar plate, the macrocarrier reaches a stopping screen, which stops the macrocarrier, but does not stop the microcarriers deposited on its surface. The microcarriers then travel through the evacuated chamber to the plate, impinging upon the cells spread on its surface, and transforming some of them in the process.

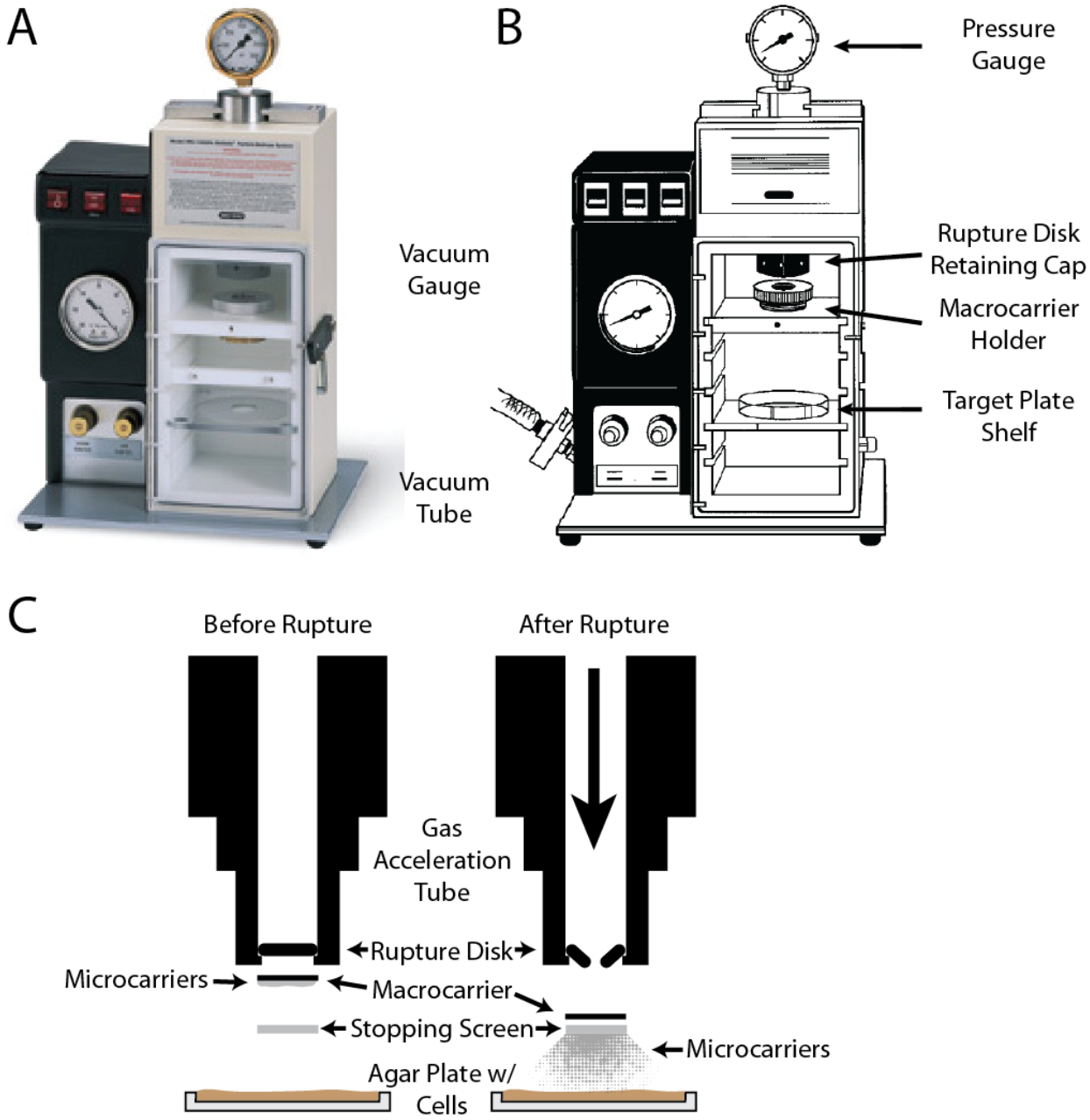

Figure 29: Transformation by high velocity microprojectile bombardment. A. A color picture of the gene gun used in this study. **B.** A diagram of gene gun, with the component parts as labeled. **C.** The mechanism by which the gene gun delivers DNA to cells (spread on the agar plate). See text for a description of the transformation procedure. Figures A&B are from by Bio-Rad, Inc. (Hercules, CA), while Figure C is a modification of one of their images. Used with permission.

The subsequent steps involve confirmation of transformation, insertion of the transformed DNA into the mitochondrial genome, and scoring for gene expression. While this technique is general, and has been used to express a number of genes, the example we

use here is the one upon which we based our experimental design[172]. This experiment aimed at creating a selectable marker for mitochondrial transcription. This selectable marker was Barstar, an inhibitor of the RNase Barnase. Without mitochondrial expression of Barstar, mitochondrially-localized Barnase would degrade the mitochondrial RNAs and confer a respiratory deficient phenotype. Mitochondrial expression of Barstar, however, would protect the mitochondrial transcriptome from such degradation. Because utilization of nonfermentable carbon requires respiration (as opposed to fermentation), growth on such a media indicates a respiratory competent strain. Thus selection for growth on nonfermentable carbon is akin to selection for a complete respiratory pathway, including mitochondrial gene expression.

Before transformation, the gene encoding for Barstar had to be modified. This is because there are a number of differences between the standard genetic code and the mitochondrial code, and expression using the unmodified gene coding for Barstar (or any heterologous gene) would result in a number of nonsense or missense mutations. This mitochochondrially-encoded version of the Barstar gene was called *BARSTM*.

Subsequent cloning steps resulted in the vector pHM102, shown in Figure 30A. For reasons that will become clear below, this vector contains the *BARSTM* ORF flanked by the 5' and 3' untranslated leader regions (UTRs) of C*OX2*, a mitochondrially-encoded gene involved in the electron transport chain. A full-length copy of *COX2* was also placed downstream of this cassette. In the experiments below and as is shown in Figure 30B&C, we replaced *BARSTM* with a mitochondrially-encoded version of *RPO41*, and performed virtually identical experiments to those using pHM102, with the exceptions as noted in the text.

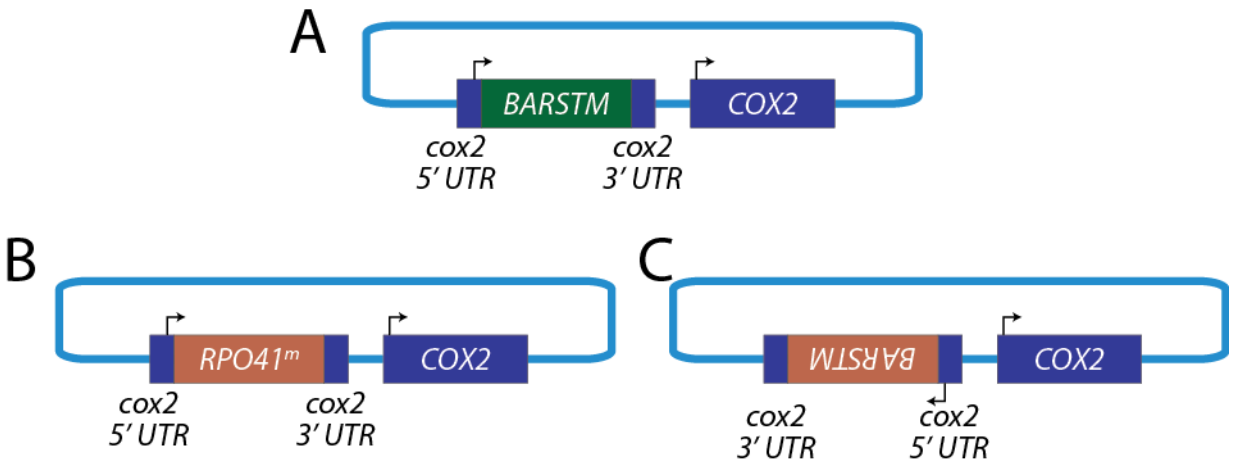

Figure 30: Mitochondrial transformation plasmids. A. pHM102, the plasmid used for expression of mitochondrially-encoded Barstar (*BARSTM*). **B.** pRPO41^{m}, the plasmid used in this study to express mitochondrially-encoded Rpo41 (*RPO41^{m}*). **C.** pRPO41^{m}-rev, a second plasmid used in this study for *RPO41^{m}* expression. Arrows indicate promoters and point in the direction of transcription. The vector backbone is in light blue, the *BARSTM* ORF in green, the *RPO41^{m}* ORF in brown, and the *COX2* ORF in dark blue. The *cox2* 5' and 3' UTRs are in dark blue, and flank *BARSTM* and *RPO41^{m}*.

For reasons that remain unclear, it has been shown that mitochondrial transformation is most efficient in a ρ^{0} strain, *i.e.* a strain that does not contain any mitochondrial DNA[173]. Strain background has also been shown to affect transformation efficiency. For example, strains derived from DBY947 have relatively high transformation efficiencies, whereas W303 and D273-10B, both common laboratory strains, are more difficult to transform[173]. Transformation of *BARSTM* used strain MCC123ρ^{0}, a D273-10B derivative that has had its mitochondrial DNA removed by the standard technique of treatment with ethidium bromide[174]. This strain was co-transformed with pHM102 and pRS316, a plasmid that expresses a nuclearly-encoded *URA3* gene, and is therefore selectable on media lacking uracil. Selection for nuclear transformants first allows for discrimination against cells that have not been transformed at all, or have been severely damaged by the transformation procedure, and gives time for any cells that have been transformed with the mitochondrial plasmid to replicate it and increase its copy number.

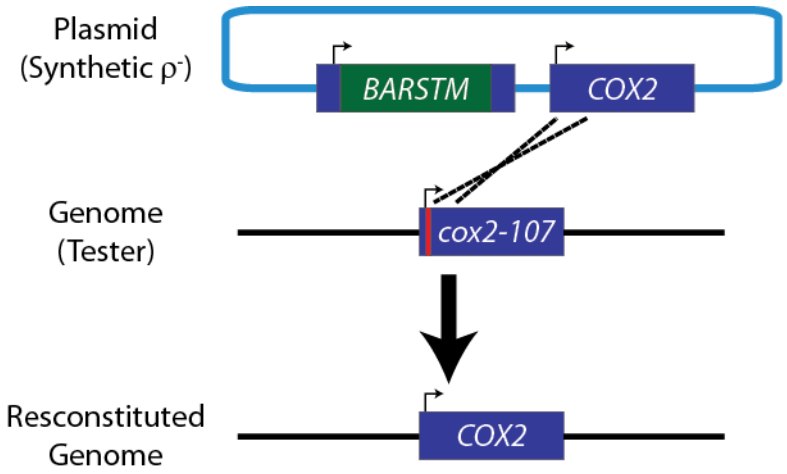

Figure 31: The marker rescue procedure. Mating between a strain transformed with a mitochondrial transformation plasmid (synthetic ρ^-, top) and a strain "tester" strain (middle) containing the *cox2-107* allele, results in homologous recombination (dashed lines) between the mitochondrial genomes of each strain, and reconstitution of a wild-type mitochondrial genome (bottom). Thus the presence of the mitochondrial transformation plasmid in the synthetic ρ^- complements the respiratory deficient phenotype that arises from the point mutations (red) that constitute the *cox2-107* allele in the tester strain. Genomic DNA is in black. All other annotations are as in Figure 30.

To isolate the mitochondrial transformants, one must perform a "marker rescue" procedure, as shown in Figure 31. This entails mating all of the transformants to a tester strain – a strain that has a mitochondrial genotype that can be rescued by the mitochondrial expression vector. In the above example, this tester strain is HMD7, which contains a fully wild-type mitochondrial genome except it has a *cox2-107* allele. This allele has seven point mutations in the cox2 5' UTR, which makes it unable to respire. Replica-plating a transformation plate to a lawn of the tester strain spread on nonselective, fermentable media allows for mating, homologous recombination of the mitochondrial genomes, reconstitution of the genes encoding for all of the mitochondrially-encoded components of the respiratory chain, and their expression to the phenotypic level. Oftentimes the mitochondrial transformants are directly distinguishable on this mating plate, as their respiratory proficiency results in a faster growing phenotype (*i.e.* the colonies are often bigger after a few days on fermentable media). However, strict selection is performed by replica plating this mating plate to a plate containing nonfermentable media. Those cells with complete mitochondrial genomes (and therefore respire) will be the only ones that grow.

One can then go back to the original transformation plate and pick those cells that were complemented by the tester strain. Because the transformation plate is crowded with both singly transformed and co-transformed colonies, these potential transformants are re-streaked and the marker rescue procedure repeated a number of times. Note that these mitochondrial transformants are likely mitochondrially polyploid, with some mitochondrial genomes containing the mitochondrial expression vector, and others remaining ρ^0. However, the phenomenon of supressiveness dictates that the strains will

tend to homoplasmy rather quickly[68]. Thus repeating this marker rescue procedure a number of times, in addition to distinguishing the mitochondrially-transformed cells from those near them on the plate, serves as an enrichment procedure for the desired strain; a strain that is a result of many marker rescue procedures will have a relatively pure mitochondrial network, containing many molecules of the mitochondrial expression vector. All of these stably transformed strains could be considered ρ^- strains, as they contain a small amount of DNA in their mitochondria (the mitochondrial expression plasmid). Consequently, there are referred to as stable synthetic ρ^- strains.

BARSTM was then inserted into the mitochondrial genome of a different strain, NB40-16D, as shown in Figure 32. This "final mater" strain, contains the *cox2-62* allele, which has a large deletion of the *cox2* 5' region, including the UTR and parts of the ORF. Thus reconstitution of the mitochondrial genome required insertion of *BARSTM*.

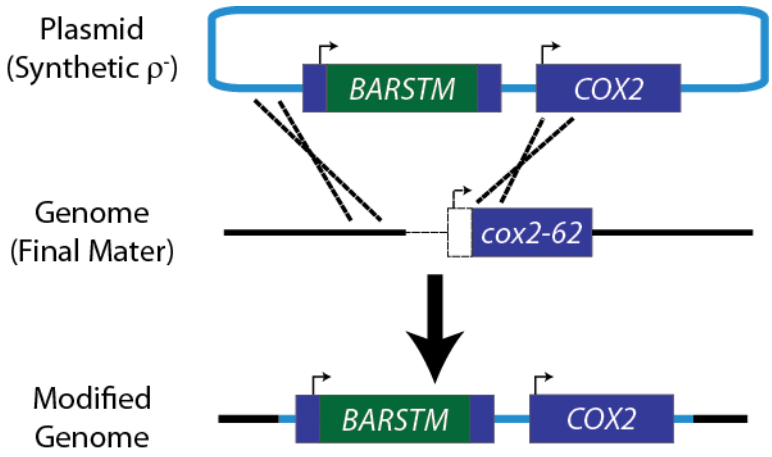

Figure 32: Insertion of *BARSTM* into the mitochondrial genome. When the mitochondrial transformation plasmid (top) is present in a strain containing the *cox2-62* allele, homologous recombination (thick dashed lines) can insert the *RPO41*^{m} ORF, along with its *cox2* UTR flanks, upstream of the *cox2* locus. This reconstitutes a full *COX2* allele from the deletion that defines the *cox2-62* allele (light dashed lines), and results in respiratory proficiency. See Figure 30 and Figure 31 for a description of all other annotations.

The insertion of *BARSTM* into the mitochondrial genome of NB40-16D did not occur through mating. Instead, it was inserted by cytoduction – the fusion of cells without the fusion of their nuclei[175]. This occurs between the mating of a *kar1-1* strain (MCC123ρ^0), which has defects in its karyogamy pathway[176], and a *KAR1* strain (NB40-16D). The offspring of this procedure are mostly haploids, with mitochondrial networks that are mixtures of the two parent strains. In addition to eliminating the time-consuming and labor-intensive process of sporulation and tetrad dissection of the diploids that result from conventional mating, cytoduction results in haploids ("cytoductants") that are identical in their nuclear genotype to their individual parents – Mendelian segregation of alleles does not occur.

Expression of *BARSTM* to the phenotypic level was confirmed by transforming a strain with *BARSTM* integrated in its mitochondria with a nuclear plasmid that expresses mitochondrially-localized Barnase. Here the example experiment and ours diverge, as will be discussed in the Results and Discussion section below.

Overview of the Experimental Steps

We now give a broad theoretical overview of the experimental procedure we describe in the Results in Discussion section below. The steps, then, to expressing a *RPO41* in the yeast mitochondria are as follows:

1. Creation of a mitochondrially-encoded version of the *RPO41* gene.
2. Cloning of said gene into a pHM102.
3. Deletion of the nuclear *RPO41* gene in a pre-established yeast transformation strain MCC123ρ^{0}.
4. Transformation of the modified transformation strain with the transformation vector via high-velocity microprojectile bombardment.
5. Mating of transformants with a "final mating" strains to establish a complete mitochondrial genome with inserted, mitochondrially encoded *RPO41*.
6. Sporulation of the mated strain and isolation of respiring strains containing only mitochondrially-encoded *RPO41*.

Results and Discussion

Creation of a mitochondrial transformation/expression vector

In order to transform a yeast mitochondrion with the *RPO41*^{m} gene, we first needed to create an *RPO41*^{m} expression vector. Merely cloning the nuclear version of the gene would not work, as not only is there a strong codon bias towards AT rich codons in the mitochondrial genome, there are a number of differences between the mitochondrial codon table and the "universal" one. Consequently, expression of heterologous genes in the mitochondria would result in a number of nonsense and missense mutations.

Gene synthesis at the time that we were preparing our optimized sequence was less sophisticated than it is now, and the sequence analysis software utilized by the synthesis companies we investigated was incapable of utilizing a different codon table (only a different codon *frequency* table). We therefore had to manually optimize our sequence by using the codon table and codon frequencies specific to the yeast mitochondrion, available in the Codon Usage Database[a]. To do this, we reverse-translated the amino acid sequence of *RPO41*, using only the most frequent codons corresponding to a given amino acid. We were constrained, however, by a number of considerations. We required that the entire sequence have no EcoRI sites, as this was the enzyme we intended to use for our cloning. In order to prevent any deleterious effects of cryptic translation and/or transcription, we also were careful to make sure the final sequence did not contain the consensus mitochondrial

[a] http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4932.mitochondrion

promoter, ATATAAGTARY[51] (R=G or A, Y = C or T). These constraints made it impossible to use the most frequent codon for each amino acid, and in the cases where we could not use the most frequent one, we used the second most frequent one, et cetera. This optimized sequence was called *RPO41^m*. The 5' and 3' *cox2* UTR flanks present in pHM102 were added to the *RPO41^m* ORF, and the sequence was submitted to GenScript for secondary structure analysis, which further changed the *RPO41^m* ORF from that of the maximally optimized sequence. The *cox2* UTRs were not changed.

The average codon frequency (the fraction a particular codon is used for a given amino acid) of our optimized sequence (ORF only) was 0.72±0.22 (error is the standard deviation). This means that each amino acid, on average, is encoded by a codon that is used 72% of the time in wild-type mitochondria. The maximum possible codon frequency (if every amino acid was encoded by the most common codon) is 0.74±0.20. A histogram of codon frequencies for our optimized sequence and the maximally optimized sequence is given as Figure 33.

GenScript synthesized the final optimized sequence, cloned into the EcoRI site of pUC57, yielding pUC57-Rpo41^m.

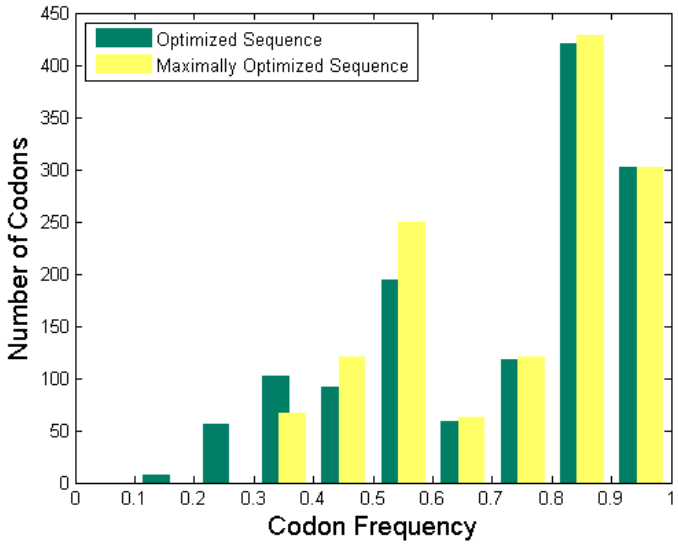

Figure 33: Histogram of codon frequencies for all codons within the synthetic ORF *RPO41^m*. Codon frequency is defined as the fraction a particular codon is used for a given amino acid within the coding sequence of the mitochondrial genome. The maximally optimized sequence, in which all codons have the maximum frequency for a particular amino acid, is given in yellow. The sequence used as *RPO41^m* in this study differed slightly in codon usage due to secondary structure and other considerations as described in the text, and is given in green.

The region of pUC57-Rpo41^m containing *RPO41^m* flanked by *cox2* 5' and 3' UTRs was inserted into the EcoRI sites of pHM102, yielding pRPO41^m and pRPO41^m-rev, as shown in Figure 30B&C. Thus pRPO41^m is effectively a vector identical to pHM102, with *RPO41^m* in

the place of *BARSTM*, while pRPO41^{m}-rev is the same as pRPO41^{m}, but with the orientation of *RPO41^{m}* in the opposite direction to that of *COX2*.

Cloning of the *cox2* 5'-*RPO41^{m}*-*cox2* 3' fragment into pHM102 allows for recombination, insertion of the *RPO41^{m}* gene 5' of the *cox2* locus, and phenotypic selection for such an insertion. Figure 34 shows the recombination event described above.

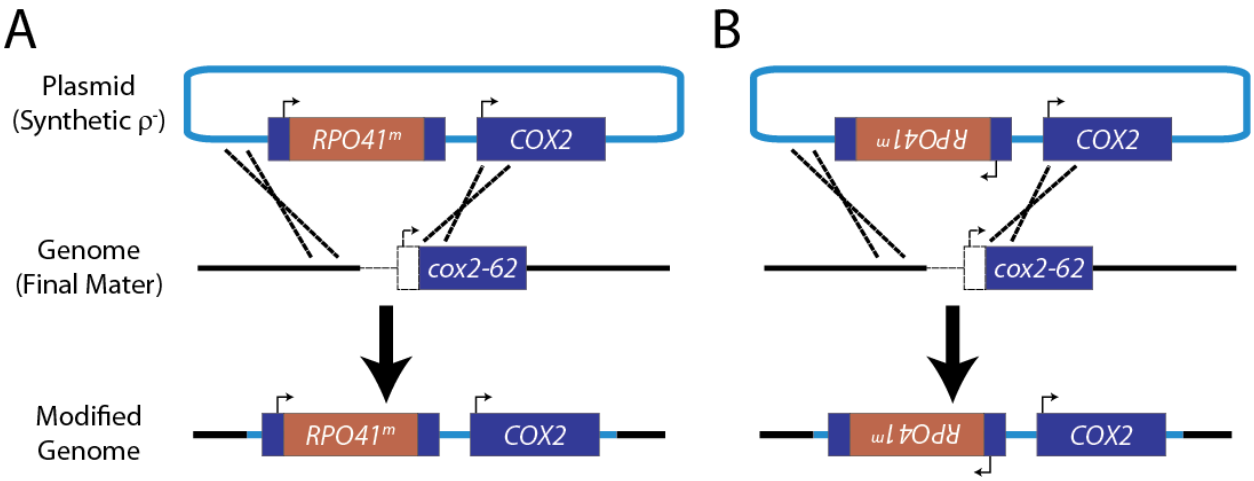

Figure 34: Insertion and selection procedure for pRPO41^{m} A. Insertion of *RPO41^{m}* into the mitochondrial genome in the same way as was described for *BARSTM* in **Figure 32**. The top is plasmid pRPO41^{m}, the middle is the *cox2* region of the final mating strain NB40-16D, and the bottom is the *cox2* region of the diploids (or cytoductants) that result from the merging of their mitochondrial genomes. **B.** The same mechanism for insertion of *RPO41^{m}* in the reverse orientation with respect to *COX2*, using plasmid pRPO41^{m}-rev (top) instead of pRPO41^{m}.

Creation of bombardment and selection strains for RPO41^{M} *expression*

As was discussed in the introduction, there are three strains involved in any mitochondrial transformation and selection experiment: the bombardment, or transformation strain; the tester strain used to check for mitochondrial transformants; and the final mater strain in which the integration and expression of the heterologous mitochondrial gene occurs. All three are haploid strains. Though a system for transformation and expression of genes in mitochondria has been developed[172], as was described in the introduction, modifications of such strains were necessary for our experiments. The properties of these three strains and their modifications for this study are described below. A flowchart of the strain modifications used in this chapter is given in Figure 35, and a list of all strains used is given in Table 8 (in the Materials and Methods section).

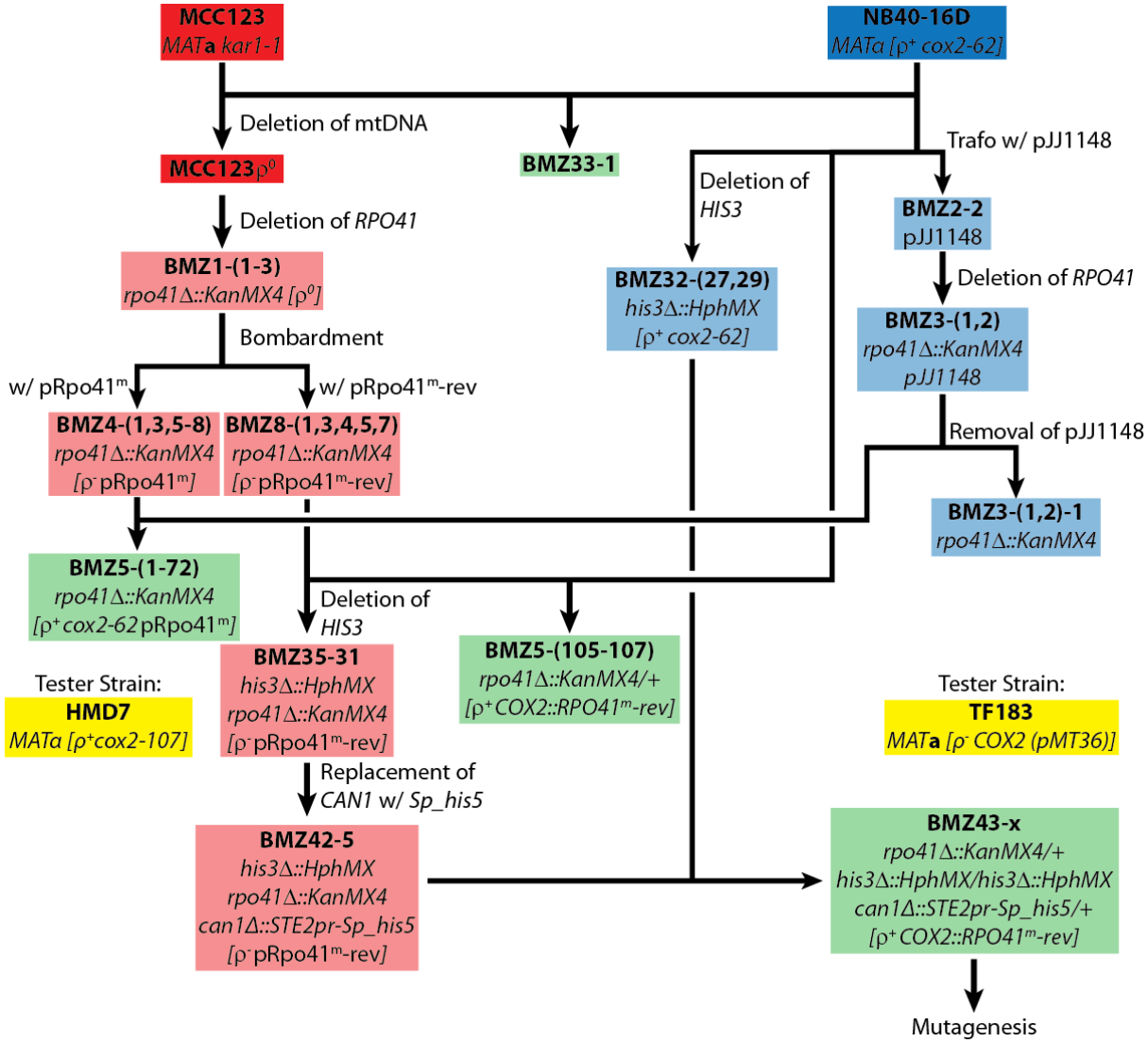

Figure 35: Flowchart of strain manipulations used in this study. Haploid strains of mating type MAT**a** are boxed in red, while MATα strains are in blue. Diploids are in green. Strains created in this study are in lighter shading. Tester strains, which are used to determine the status of the mitochondrial genome, are boxed in yellow, and are on the side of the lineage they are used to test. See text for details. Mitochondrial genotypes are within square brackets.

Modification of the transformation (bombardment) strain

As previously discussed, mitochondrial expression of *BARSTM* was verified via the resistance against Barnase, a cytotoxic RNase[172]. Our selection plan was quite different, and consequently required modification of both the bombardment strain and the final mating strain (discussed below). Because the intent of the project was to express the mitochondrial RNA polymerase within the mitochondrial matrix, our end goal was to produce a strain that did not contain any nuclear version of *RPO41*, yet was still capable of

respiration (i.e. had a fully functional mitochondrial network, including mitochondrial transcription performed by the new mitochondrially-encoded RNA polymerase). Therefore, one of the strains involved in the final cytoduction event, when the *RPO41^{m}* gene was to be inserted into the mitochondrial genome, needed to have a deletion of the nuclear *RPO41*.

As stated above, strains containing insufficient amounts of functional Rpo41 are prone to modifications of the mitochondrial genome, including massive deletions and rearrangements[64]. Deleting *rpo41* from the final mating strain, then, would turn it from a strain containing a mostly wild-type mitochondrial genome (save the aforementioned *cox2-62* allele), to a *ρ*$^{-}$ strain with much of its mitochondrial genome absent. Subsequent insertion of *RPO41^{m}* into the genome of the final mater strain would then be unlikely, and recuperation of the respiratory proficient phenotype even less so. However, *rpo41* strains are capable of maintaining *ρ*$^{-}$ genomes[69], including those consisting of a transformed plasmid[165]. Therefore, deletion of *rpo41* had to be accomplished in MCC123*ρ*^{0}, the initial bombardment strain.

In order to delete the *RPO41* gene in MCC123ρ^{0}, we PCR amplified around the *rpo41* region in BY4743 YFL036W, a diploid strain from the Stanford Deletion Project[167] in which the *rpo41* gene has been replaced with the *KanMX4* gene (it is homozygous at that locus). This "RPO41 deletion cassette" was transformed into MCC123ρ^{0}. Strains that had the *KanMX4* gene inserted into their genome were selected using G418 in the culture media (the gene product of the *KanMX* gene confers resistance to the antibiotic G418).

Deletion of *rpo41* was confirmed by analytical PCRs on genomic DNA from the transformants as well as for wild-type and *rpo41*-deleted controls, using a forward primer upstream of the insertion site and a reverse primer either inside the *RPO41* ORF or the *KanMX4* ORF. We found a number of transformants that amplified only when the primer inside the *KanMX4* ORF was present, indicating proper insertion of the *KanMX4* gene in the place of the *RPO41* ORF (shown in Figure 36). These strains were labeled BMZ1-1, BMZ1-2, and BMZ1-3.

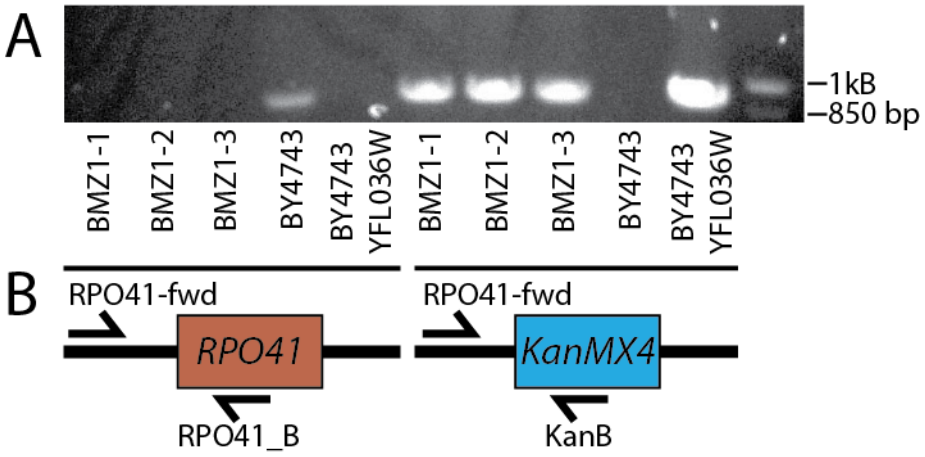

Figure 36: Creation of *rpo41*-deleted transformation strains. A. Results of analytical PCR on deletion mutants. Lanes 4-8: PCR reactions specific for the *rpo41Δ::KanMX4* deletion cassette using RPO41-fwd and RPO41-B. Lanes 1-4: PCR amplifications specific to the wild-type version of *RPO41* using the primer pair RPO41-fwd and KanB. BMZ1-1,

BMZ1-2, and BMZ1-3 are transformants of the bombardment strain MCC123ρ^0 with the *KanMX4* cassette. These strains amplify in an identical manner to BY4743 YLF036W, a strain identical to BY4743 (negative control), but with *RPO41* replaced with the *KanMX4* cassette, **B.** Schematic representation of the primers localization on *RPO41m and rpo41Δ::KanMX4*

Modification of the final mating strain

The final mating strain has to harbor a ρ^+ mitochondrial genome for insertion of the synthetic *RPO41^m* construct. This implies that the mtDNA has to be replicated with the help of a functional Rpo41. Cytoduction allows for the mitochondrial genome to be kept intact (in NB40-16D) until it interacts with pRPO41^m during cytoplasmic (and mitochondrial) fusion. Because the nuclei do not fuse during cytoduction, half of the cytoductants will be *rpo41Δ::KanMX4*. However, the initial transcription of *RPO41^m* in the *rpo41Δ::KanMX4* cytoductant must come from the leftover Rpo41 protein from the fusion event. Because the rate of expression of heterologous genes within the mitochondria is not well understood, and because the amount of Rpo41 protein accumulation necessary to manifest itself phenotypically is not known, we were unsure if the leftover Rpo41 protein was present in enough quantity and for a long enough time to ensure a seamless transition from a nuclear source of Rpo41 protein to a mitochondrial one. We therefore decided to prepare strains for counterselectable expression of nuclear *RPO41*. Diploids, then, could be created and allowed to grow for as long as necessary, nuclear expression of *RPO41* turned by removal of the plasmid via conterselection, and the resultant phenotype scored.

To do this, we transformed NB40-16D with pJJ1148[26], a plasmid containing the *RPO41* gene under its natural promoter and a *URA3* gene for selection/counterselection as described below. This yielded strain BMZ2-2. After confirming transformation of the plasmid, we deleted *RPO41* from the genome of BMZ2-2, using the procedure as described for the creation of BMZ1-1, yielding strains BMZ3-1 and BMZ3-2.

PCR reactions using RPO41_fwd and either RPO41_B or KanB (Figure 37) resulted in amplification in all reactions (save the negative control). This indicates that both the deletion cassette and the RPO41 gene were present in strains BMZ3-1 and BMZ3-2 but does not indicate the location of either. It is possible, then, that the *KanMX4* cassette was inserted into the plasmid-borne copy of *RPO41*.

To determine the location of the inserted *KanMX4* cassette, we removed pJJ1148 by counterselection using 5-fluoroorotic acid (5FOA). 5FOA is converted to fluorodeoxyuridine, a thymidylate synthetase inhibitor via the Ura pathway in yeast[121]. Strains with functional Ura pathways, then, cannot grow on media containing 5FOA. Because NB40-16D (and its derivatives) is *ura3*, the loss of pJJ1148 (and its *URA3* markers) allows for growth on media containing 5FOA.

PCRs products of genomic DNA from strains of BMZ3-1 and BMZ3-2 that had pJJ1148 removed via counterselection (named BMZ3-1-1 and BMZ3-1-2, respectively) only resulted in products when KanB was present, indicating that the *KanMX4* cassette is present at the nuclear locus of *RPO41*, and not the plasmid-borne copy (see Figure 37).

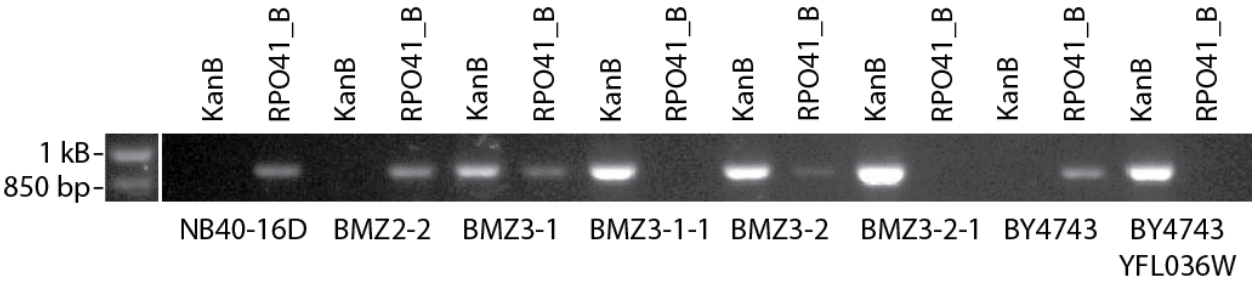

Figure 37: Creation of conditional-expression *RPO41* final mating strains. Final mating strain NB40-16D was transformed with *RPO41*-expression vector pJJ1148 to yield strain BMZ2-2. Subsequent deletion of *RPO41* yielded strains BMZ3-1 and BMZ3-2. PCR amplification of BMZ3-1 and BMZ3-2 indicates the presence of both deleted and wild-type *RPO41* alleles (compare with wild-type control BY4743 and deletion control BY4743 YFL036W). Removal of the pJJ1148 plasmid resulted in BMZ3-1-1 and BMZ3-1-2. The *RPO41*-specific band disappears in these strains, indicating that the deletions in BMZ3-1 and BMZ3-2 are on the genome as opposed to the plasmid. Primers locations are as described in Figure 36B, with the reverse primers as indicated above the gel image in this figure.

We performed marker rescue procedures to determine if the mitochondrial genomes of BMZ3-1 and BMZ3-2 were intact. In this case the tester strain TF183, which contains only *COX2* in its mitochondria (an extreme version of a ρ^-, as it is missing all of its mitochondrial genome save *COX2*), was mated with both BMZ3-1 and BMZ3-2. This mating plate was then replica-plated to YPEG, resulting in a lawn of respiring diploids. This confirmed that the mitochondrial genome of strains BMZ3-1 and BMZ3-2 were nonrespiring only because of mutations at the *cox2* locus.

TRANSFORMATION OF YEAST MITOCHONDRIA WITH A PLASMID CONTAINING RPO41^M

In our experiment, we transformed BMZ1-1 with the *URA3*-containing plasmid pRS316 and either pRPO41^m or pRPO41^m-rev, and selected for Ura$^+$ transformants. Because MCC123ρ^0 (and therefore BMZ1-1), is *ura3,* untransformed strains were unable to synthesize their own uracil (i.e. they had an Ura$^-$ phenotype), whereas strains successfully transformed with pRS316, became Ura$^+$. This selection was done on SD-Ura+G418 (synthetic media containing the antibiotic G418, the fermentable carbon dextrose, and all essential amino acids except uracil), with 1M sorbitol added to protect transformed cells from osmotic shock.

After much optimization, we found that the most important variables for efficient nuclear transformation via high velocity microprojectile bombardment (in order from most to least important), were: chamber pressure (the maximum vacuum, 29.5 in. Hg. below atmospheric pressure, was the best), colony density on the plate (1×10^8 cells per plate was best), and rupture pressure (1100 PSI rupture disks worked marginally better than 1350 PSI rupture disks). At our best optimization, we were achieving thousands of nuclear transformants per plate (see Figure 38).

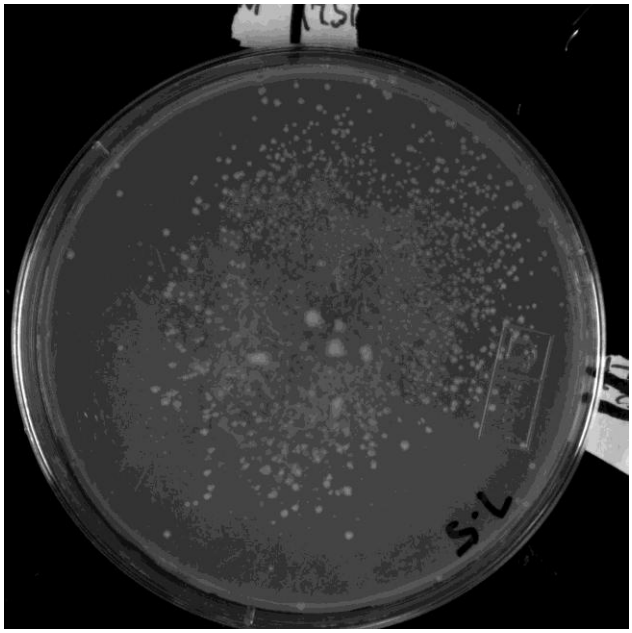

Figure 38: Nuclear transformants from the high velocity microprojectile bombardment procedure. BMZ1-1 cells were bombarded with pRS316 and pRPO41^{m} precipitated on 600nm tungsten particles (Materials and Methods). Plate media is synthetic media containing G418, sorbitol, and lacking uracil. Cells transformed with pRS316 become Ura^{+}, and are seen here.

We performed our transformations in sets of six plates using 4.25 µg nuclear plasmid. We obtained approximately 3000 nuclear transformants per plate. This corresponds to a transformation efficiency of ~4200 transformants/µg DNA.

SELECTION OF MITOCHONDRIAL TRANSFORMANTS

Of the thousands of nuclear transformants per bombardment plate, only a fraction of them are the desired double transformants, having been co-transformed with both the nuclear and mitochondrial expression vectors. Furthermore, the desired strain must have the mitochondrial expression vector delivered to its proper place: the mitochondrial matrix. As was described in the introduction, mitochondrial transformants are identified by a marker rescue procedure, in which potential transformants are mated with a strain that would distinguish mitochondrial transformants by complementation. We performed the marker rescue procedure of our transformation plates using the tester strain HMD7, a strain with a full mitochondrial genome (ρ^{+}), except for seven point mutations in the *cox2* 5' UTR that prevents respiration (mit^{-}). This allele, *cox2-107*, can be easily rescued by complementation with the *cox2* regions of pRPO41^{m} or pRPO41^{m}-rev (see Figure 39).

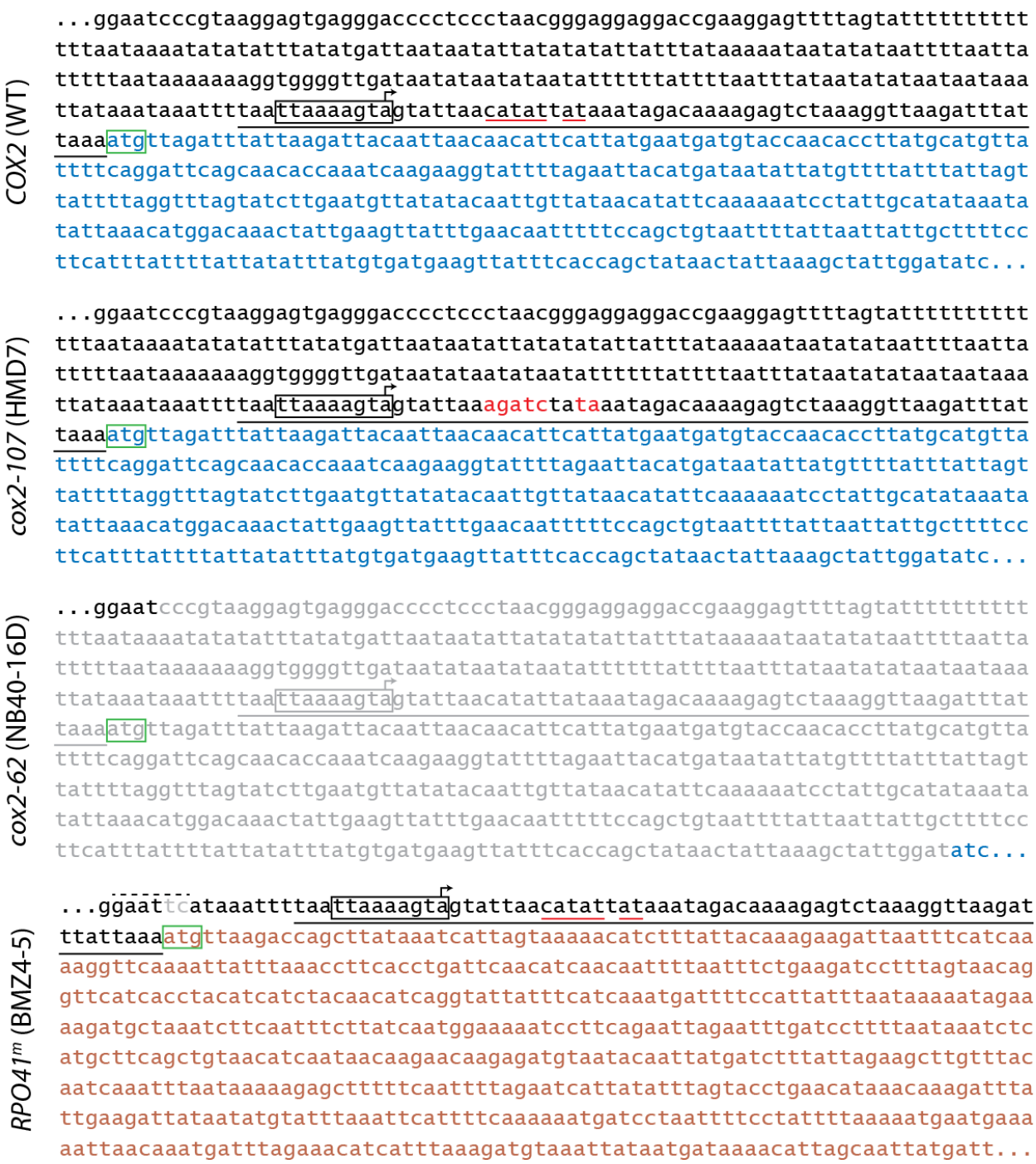

Figure 39: 5' regions of COX2 alleles used in this study. The *COX2* ORF is in blue typeface, while the promoter is boxed in black, with an arrow over the first transcribed nucleotide, and the initiation codon boxed in green. The regulatory 5' untranslated region (UTR) is underlined in black. Point-mutations present in the *cox2-107* are unlined in red in the wild-type sequence (top) and the synthetic *RPO41*^{m} construct (bottom), and are in red typeface in the *cox2-107* allele (second from top). These mutations abolish respiration. Thus strain HMD7 can be used to test if other strains have a wild-type *COX2* 5' region. The

deleted region of the *cox2-62* allele (third from top) is in gray. Homologous recombination of the upstream region of this allele with the same region in *RPO41^m* (bottom) results in insertion of the *RPO41^m* ORF (brown). The EcoRI restriction site used for cloning the *RPO41^m* construct is indicated by a dashed line over the sequence, with nucleotides of differing sequence from the wild-type *COX2* sequence in gray.

In the case of transformation of BMZ1-1 with pRS316 and pRPO41^m, the initial marker rescue of approximately 9000 nuclear transformants (6 plates of approximately 3000 transformants each), resulted in 12 mitochondrial transformants. This corresponds to a mitochondrial transformation efficiency of 13 mitochondrial transformants per thousand nuclear transformants. These transformants were subjected to the marker rescue procedure three times, and eight colonies were isolated and labeled BMZ4-1 through BMZ4-8. Genomic preparations from these colonies were subjected to analytical PCR to confirm the presence of *RPO41^m*. All but one (BMZ4-4) did (see Figure 40A).

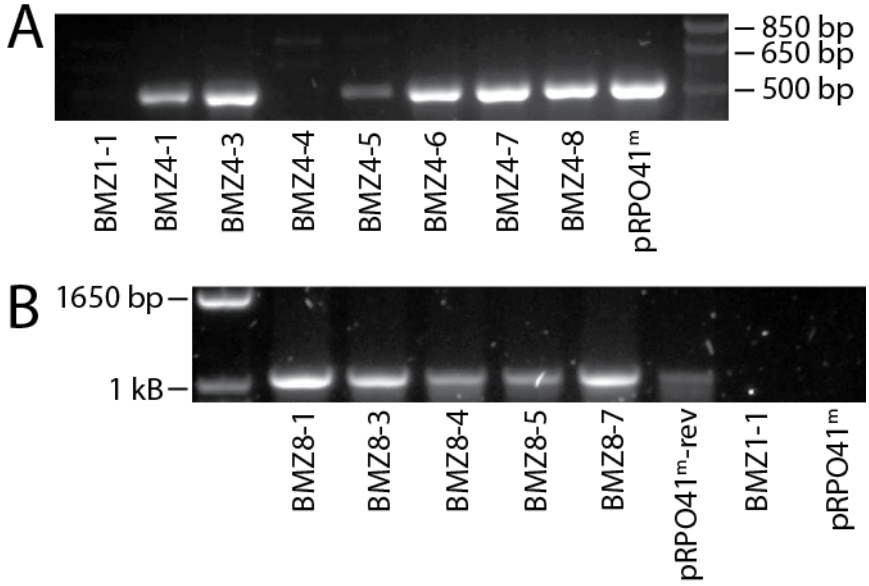

Figure 40: Conformation of mitochondrial transformation of BMZ1-1 *RPO41^m* plasmids. A. Confirmation of the transformation of BMZ1-1 with plasmid pRPO41^m. **B.** Confirmation of the transformation of BMZ1-1 with plasmid pRPO41^m-rev. Genomic preparations of the strains as labeled were subjected to PCRs specific to the transformation plasmid. BMZ1-1, the untransformed strain, showed no amplification in each case (negative control), while amplification using the transformation plasmid as a template resulted in products identical to those in the transformants.

Transformation of BMZ1-1 with pRS316 and pRPO41^m–rev was done in a more systematic manner. Twenty-three plates were bombarded, resulting in approximately 14,000 nuclear transformants (600 per plate). Only nine of those nuclear transformants marker rescued, indicating a mitochondrial transformation efficiency of 0.6 mitochondrial transformants per thousand nuclear transformants. Because this transformation was done many months after the transformation pRPO41^m, the decrease in mitochondrial

transformation efficiency may be due to degradation of the spermidine used in the precipitation reaction, or other systematic changes in the transformation procedure.

Of the nine successful transformants, five were selected after multiple marker rescues and confirmation of the insertion via PCR (see Figure 40B). These strains were named BMZ8-1, BMZ8-3, BMZ8-4, BMZ8-5, and BMZ8-7.

INTEGRATION OF RPO41^{m} *INTO THE MITOCHONDRIAL GENOME*

Once a stable synthetic ρ^- strain is isolated, the final cytoduction procedure can be performed. This essentially consists of incubating the stable synthetic ρ^- with the final mater strain for a limited amount of time, and plating the resultant mixture onto media that allows for identification of haploids (using screening of auxotrophic markers), deletion of the nuclear *RPO41* (by selection on media containing G418), and respiratory proficiency (by selection on media containing only nonfermentable carbon).

We ran into many challenges in attempting to insert *RPO41^{m}* into the mitochondrial genome. We summarize those challenges below, and conclude with limited progress: we have succeeded in placing *RPO41^{m}*, only in the reverse orientation, into the mitochondrial genome of a diploid heterozygous for *RPO41* at the nuclear locus.

Cytoduction results only in respiring diploids

We attempted cytoduction using both BMZ4-5 and BMZ8-7 strains with the final mater strain NB40-16D. We recovered thousands of respiring colonies, 100% of which were diploids. Diploids occur in low frequency during cytoduction. This indicated that expression of *RPO41^{m}* alone to the phenotypic level was not occurring. We reasoned that this could have been the result of insufficient time for expression of *RPO41^{m}* to occur. Therefore, we moved on to conditional expression of *RPO41* in a diploid.

Conditional expression of nuclear *RPO41* in plagued by the inability to counterselect against nuclear expression.

As was described above, we created strains BMZ3-1 and BMZ3-2, which are isogenic to strain NB40-16D, except (a) they have been transformed with pJJ1148, an *RPO41* expression vector, and (b) they have had their genomic version of *RPO41* deleted. Because pJJ1148 has the selectable/counterselectable *URA3* marker, nuclear expression of *RPO41* can, in principle, be turned off by removal of pJJ1148.

We therefore mated strain BMZ4-5 with either BMZ3-1 or BMZ3-2 on YPD (yeast extract, peptone and dextrose – a rich medium with a fermentative carbon source). This mating mixture was then transferred to SD+5FOA (synthetic medium containing dextrose) to select against pJJ1148. After three days of growth, this plate was replica plated to SEG+5FOA (synthetic medium with ethanol and glycerol as non-fermentative carbon sources). Full lawns were recovered each time, with controls (BMZ1-1 mated to BMZ3-1 and BMZ3-2, as well as all strains unmated) yielding blank plates. The lawns were streaked on SEG+5FOA and SD-Ura. Strains grew on SEG+5FOA and did not grow on SD-Ura, confirming the ura$^-$ phenotype and the respiratory proficiency. We labeled the strains that were the result of matings between BMZ4-5 and BMZ3-1 as BMZ5-x, where “x” is a number

of the colony and ranges from one to over one hundred. Matings between BMZ4-5 and BMZ3-2 were labeled as BMZ6-x. Both strains BMZ5-x and BMZ6-x will heretofore be commonly referred to as "the diploids."

We tried many different protocols do isolate and select for the diploids, including different orders of selection and waiting times between selections. PCR analysis of over eighty of these strains indicated that the plasmid-borne copy of *RPO41* was indeed present, despite the lack of *ura3* activity (see Figure 41). This can be explained by mutations within the *URA3* gene on the plasmid, rendering it nonfunctional (J. Rine, personal communication). This would result in the ura$^-$, 5FOA^R phenotype, but the continued presence of the plasmid would provide the cell with a continued supply of nuclear (plasmid-born) Rpo41 protein.

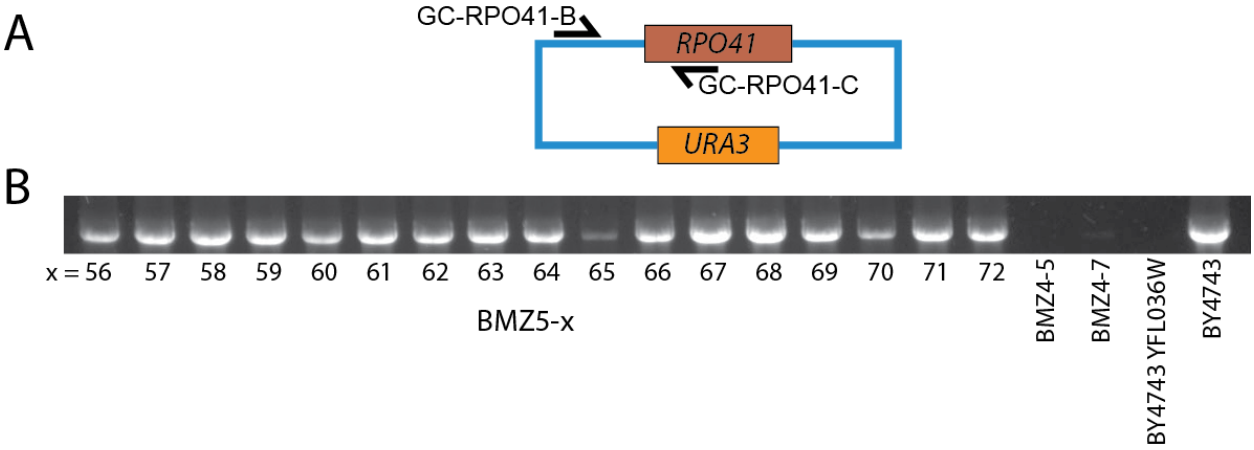

Figure 41: The plasmid-born version of *RPO41* remains after counterselection. A. Schematic of the confirmation PCR, using primers internal to the plasmid-borne copy of *RPO41*, and the vector backbone. **B.** BMZ5-x strains contain mitochondrial-inserted *RPO41^m*, and have been counterselected for removal of the plasmid-born copy. Analytical PCR shows that these strains retain the plasmid-born copy of *RPO41*, even while surviving counterselection against *URA3* expression using 5FOA (see text). BY4743 is a common wild-type strain (positive control). Negative controls include the initial mitochondrial transformants (BMZ4-5 and BMZ4-7), and BY4743 YFL036W, a strain identical to BY4743 but with the *RPO41* removed.

The fact that none of the dozens of colonies that we scored for loss of the shuffle plasmid lost it after counterselection suggests that the nuclear version of *RPO41* – in this case the plasmid-borne version – is essential for viability of the strains under selection for respiration. In other words, though not conclusive, a plausible explanation for the inability for us to isolate a respiring strain in which the only source of Rpo41 came from expression of *RPO41^m* is that expression of *RPO41^m* is either insufficient or deleterious to achieving the respiratory phenotype.

Intramolecular recombination of pRPO41^m rescues the *cox2-62* allele, and obviates the requirement for insertion of *RPO41^m*.

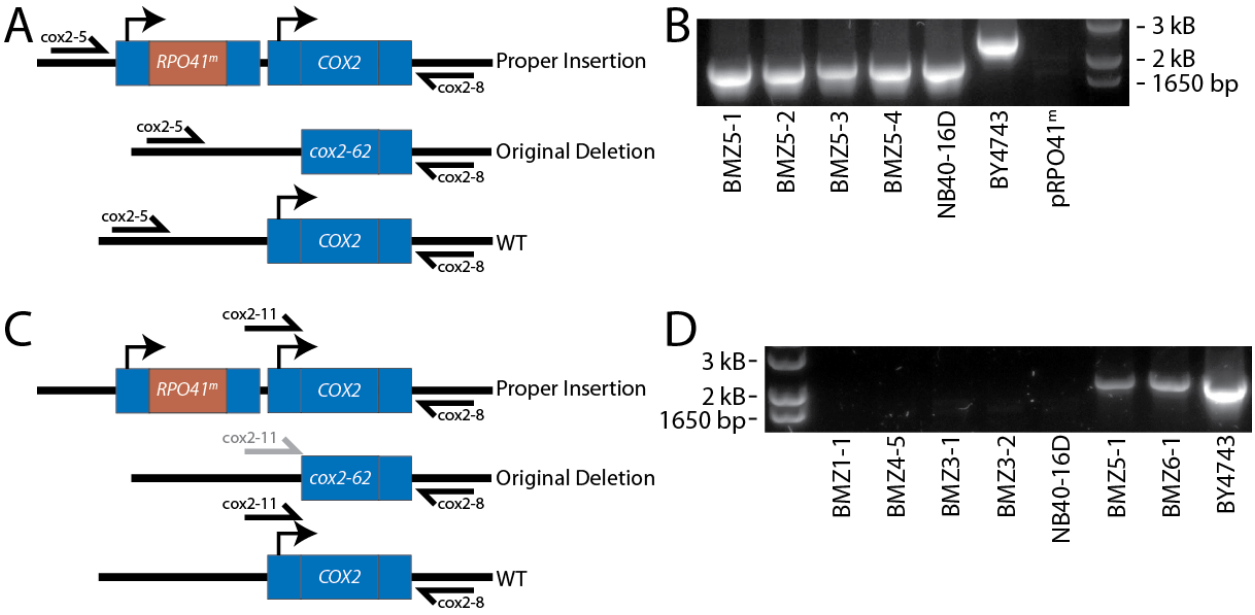

Figure 42: Diploids contain both the *cox2-62* and the *COX2* alleles. A. Schematic representation of the analytical PCR used to determine the nature of the *COX2* allele present at the *COX2* locus. Primers *cox2-5* and *cox2-8* were designed to amplify around the *COX2* locus, and allow for determination of the allele by the size of the PCR product. **B.** Agarose gel of the PCR reactions depicted in (A). All diploid strains (BMZ5-x) contain the same *COX2* allele as that present in NB40-16D – *cox2-62* – and do not have the full-length *COX2* allele present in BY4743 at that locus. The original transformation plasmid (pRPO41^{m}) did not amplify (negative control). **C.** Schematic representation of the analytical PCR used to determine the presence of full-length *COX2*. Primer cox2-11 sits in the region deleted in the *cox2-62* allele, allowing for specific amplification of full-length *COX2* alleles. **D.** Agarose gel of the PCR reactions depicted in (A). The diploid strains (BMZ5-1 and BMZ6-1) contain a full-length *COX2* allele in addition to the cox2-62 allele. Negative controls include the original transformation strain (BMZ1-1), a transformant (BMZ4-5), and final mating strains containing the *cox2-62* allele (BMZ3-1, BMZ3-2, and NB40-16D).

Note that there are two requirements for BMZ5-x strains to respire:

1. There must be an ample supply of Rpo41 within the mitochondrial compartment. As described above, this requirement was met not by insertion of *RPO41^{m}* in the mitochondrial genome and its subsequent expression, but from the plasmid-born copy which we are incapable of selecting against.
2. There must be a reconstitution of the full-length *COX2* allele from its partially deleted *cox2-62* form in the final mater (as in Figure 34).

We therefore performed PCR reactions around the deleted region of the *cox2* locus as shown in Figure 42A. Proper insertion of *RPO41^{m}* would result in a PCR product much longer than the wild-type *COX2* allele, which in turn is longer than the *cox2-62* allele. A gel of the PCR products is given as Figure 42B. To our surprise, the lengths of the PCR products

in the BMZ5-x strains corresponded exactly to those of NB40-16D, and were smaller than that of the wild type strain. This indicates that not only was there an apparent lack of insertion of *RPO41*^{m} upstream of the *cox2* locus, but also that the *cox2-62* allele was not reverted to wild-type.

The *cox2-62* simply could not be the only version of the *COX2* gene present in our strains, for the *cox2-62* allele has a mit^{-} phenotype, and our diploids respire. We therefore designed a PCR reaction as shown in Figure 42C. This reaction used a primer that sat in the deleted region of the *cox2-62* allele and therefore would only amplify a full-length *COX2* gene. As the gel in Figure 42D indicates, the diploid strains do contain a full-length version of *COX2*.

To determine the location of the full length version of *COX2* in the mtDNA, we performed an inverse PCR procedure[177]. Inverse PCR is a method to determine the flanking sequences around a region of known sequence, and is diagrammed schematically in Figure 43. In this method, one digests the template with a restriction enzyme with a recognition sequence that has a high probability of being present in the flanking regions, but does not occur in the known region. In our case, we used the enzyme ApaI, which cuts the mitochondrial genome 44 times, but does not cut *COX2*.

The digested DNA is then subjected to a ligation reaction, which results in intermolecular and intramolecular ligations, the latter of which form DNA circles of varying size. These DNA circles can then be subjected to a PCR reaction with divergent primers that are designed from the region of known sequence. This PCR amplifies the unknown sequence bordered by the primer pair (Figure 43). These PCR products can be gel purified and commercially sequenced to determine the sequence of the regions flanking the region of known sequence.

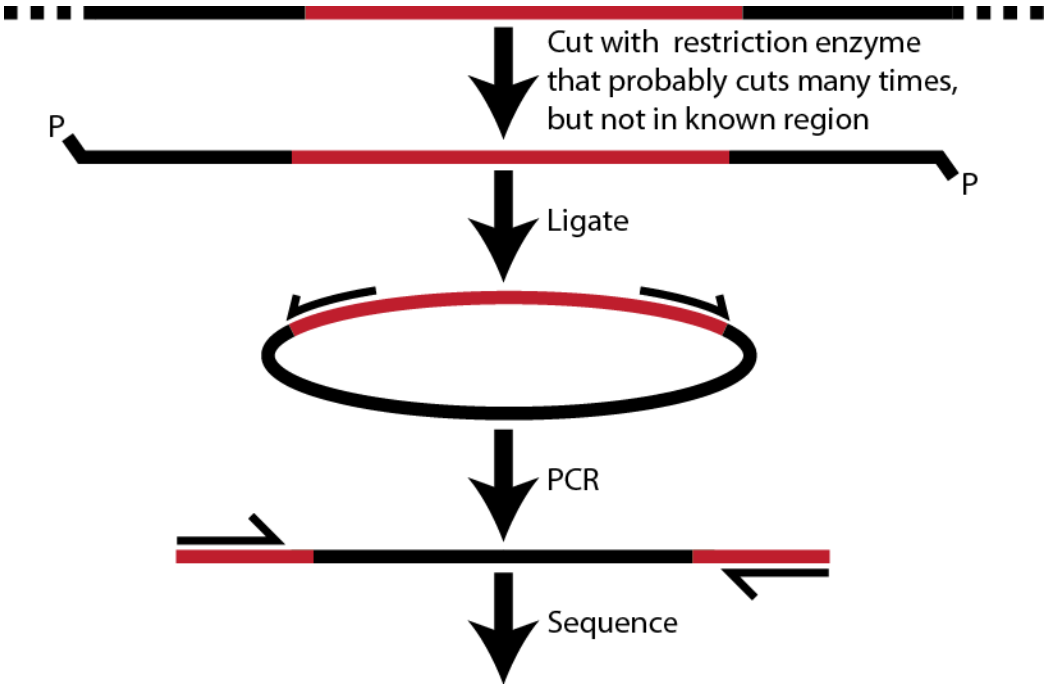

Figure 43: Schematic diagram of the inverse PCR method of determining the location of a region of known sequence. Sample DNA is first digested with a restriction enzyme that has a high probability of cutting the DNA, but does not cut the region of known sequence (red). The fragments created by restriction digestion are suitable for ligation, as represented by the 5' phosphates ("P") on each end. Ligation results in circular DNA molecules that can then be PCR amplified by using primers that face away from each other, and lie within the region of known sequence. These PCR products can then be sequenced, giving the sequence information of the region surrounding the region of known sequence (black).

We performed this inverse PCR reaction with two primer sets: cox2-6 with cox2-11, and cox2-6 with cox2-8. Cox2-6 and cox2-11 both anneal to the deleted region of the *cox2-62* allele, and point in opposite directions. This allowed for specific determination of the location of the full-length *COX2* allele, as opposed to that present at the *cox2* locus, which we already knew contained a *cox2-62* allele. Cox2-8 sits far downstream of *COX2*, and in conjunction with cox2-6, served to determine if there was a large amount of *COX2* 3' sequence downstream of the undetermined full-length *COX2* allele, and would give an amplification product even if the DNA circle that resulted from the inverse PCR was too large to amplify. Both PCRs amplified, as shown in Figure 44.

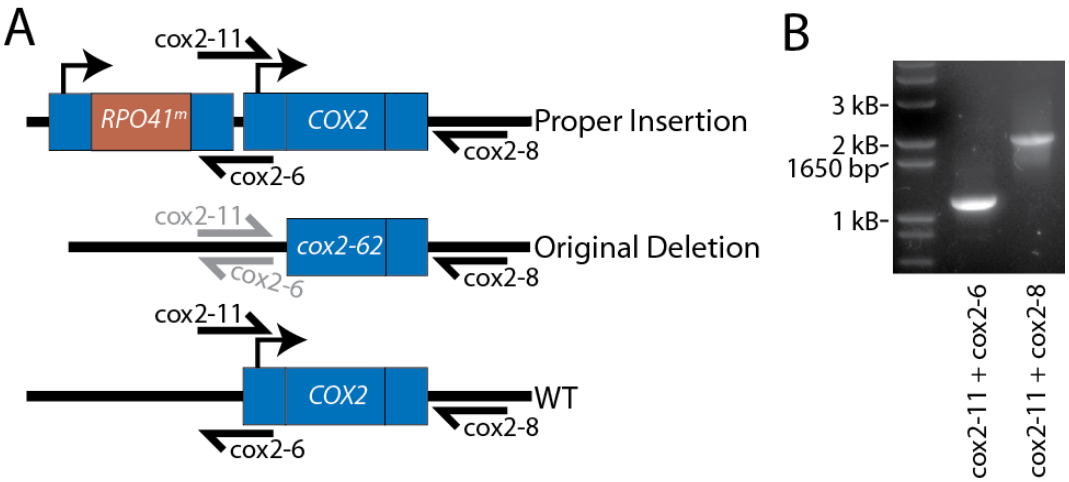

Figure 44: PCR shows the presence of full-length COX2 alleles in BMZ5-1. A. Schematic for the inverse PCR reaction used in this study. Primers cox2-11 and cox2-6 sit within the deleted region of the *cox2-62* allele of the final mating strain NB40-16D, while primer cox2-8 sits downstream of the *COX2* locus. **B.** Agarose gels of the PCR reactions depicted in (A), with genomic DNA from strain BMZ5-1 as a template. No ligase or restriction enzyme was used in these reactions. Ligase-independent amplification using cox2-11 and cox2-6 indicates that the amplicon exists as a pre-existing miniplasmid in the template DNA, while amplification using cox2-11 and cox2-8 indicates that the full-length *COX2* gene is present within BMZ5-1.

As a negative control, we performed the inverse PCRs without the addition of restriction enzyme and ligase (in separate reactions). Surprisingly, we found that amplification of the product as shown in Figure 44 was independent of both ligation and restriction digestion. This indicates that formation of DNA circles was not necessary for amplification, because the DNA circle was already formed. Sequencing of the genomic DNA confirmed this: the full length version of *COX2* is present alone as an extrachomosomal miniplasmid within the mitochondrial matrix. This is possible through intramolecular recombination between the two corresponding UTRs of *COX2* on pRPO41^{m} – one flanking *RPO41^{m}*, and one flanking the *COX2* ORF. Both the 5' and 3' UTRs can participate in this recombination, leading to two species, as is shown in Figure 45. It is these two species that result in the two bands shown in Figure 44.

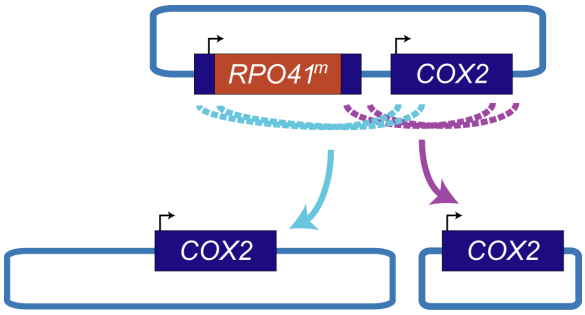

Figure 45: Intramolecular recombination pRPO41^{m} excises *RPO41^{m}*, while keeping full-length *COX2*. Intramolecular homologous recombination between the two 5' ends

(light blue) or the two 3' ends (purple) in pRPO41^{m} results in plasmids that contain a full-length copy of *COX2*, but have excised *RPO41^{m}*.

Note that if the original transformation plasmid was pRPO41^{m}-rev instead of pRPO41^{m}, the corresponding homologous recombination event would not result in a reconstitution of a miniplasmid containing full length *COX2*. As shown in Figure 46, the resultant recombination products would either lack both *COX2* and *RPO41^{m}*, and therefore would not survive selection on YPEG, or would result in a plasmid with the vector backbone excised, but with both *RPO41^{m}* and *COX2* present.

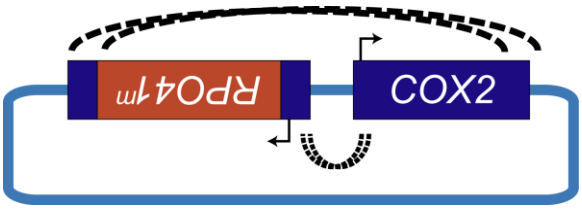

Figure 46: Intramolecular recombination on pRPO41^{m}-rev cannot excise *RPO41^{m}*. The resultant plasmids that would result from an intramolecular recombination of the plasmid pRPO41^{m}-rev would either lack any *COX2*, or contain both *RPO41^{m}* and *COX2*. This plasmid, when transformed using the high velocity microprojectile bombardment procedure described in the text, resulted in integration of RPO41m into the mitochondrial genome.

Use of pRPO41^{m}-rev results in stable homozygous diploids with integrated *RPO41^{m}*.

The above sections discuss two major challenges to getting integration of *RPO41^{m}* into the mitochondrial genome using the transformation plasmid pRPO41^{m}:

1. Mutations in the *URA3* marker on the shuffle plasmid give the same phenotype as its loss, resulting in our inability to use counterselection as a means to "turn off" nuclear *RPO41* expression.
2. Intramolecular homologous recombination within the pRPO41^{m} plasmid allowed for the reconstitution of wild-type *COX2* (on a miniplasmid) while obviating the need for integration of *RPO41^{m}* into the mitochondrial genome.

To overcome those issues, we applied two strategies:

1. Use Mendelian segregation in a diploid heterozygous at the nuclear *RPO41* locus as a means to provide transient nuclear *RPO41* expression. This is in contrast to counterselection of a shuffle plasmid.
2. Use pRPO41^{m}-rev as the mitochondrial transformation plasmid. This eliminates the problem of intramolecular recombination of pRPO41^{m}.

To this end, BMZ8-7, a stable synthetic ρ^- that had been transformed with pRPO41^m-rev, was mated with NB40-16D, the final mating strain (containing the *cox2-62* allele) that has been successfully used in previous studies[172]. The result of this mating were diploids BMZ5-(105-107), which were heterozygous for nuclear *RPO41*, *i.e. rpo41Δ::KanMX4/RPO41*. Proper *RPO41* expression was confirmed by growth on YPEG, while the presence of the deletion (and therefore expression from the *KanMX4* gene) was confirmed by growth on YPDA+G418 (rich medium containing glucose and G418, supplemented with adenine. See), and is show in Figure 47. The negative controls of each strain unmated yielded blank plates.

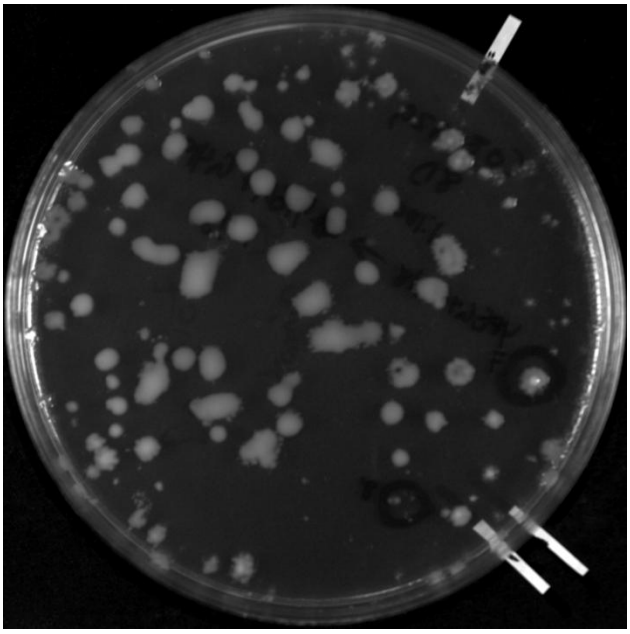

Figure 47: Mating of BMZ8-7 and NB40-16D yields diploids that are heterozygous at the RPO41 locus. Diploids that result from the mating of BMZ8-7 and NB40-16D can respire (and therefore grow on YPEG) due to the wild-type copy of *RPO41* brought from NB40-16D. They are G418^R due to the deleted copy of *RPO41* brought from BMZ8-7, which was deleted using the *KanMX4* marker that confers G418 resistance. Control plates of BMZ8-7 alone, NB40-16D alone, and BMZ1-1 mated with NB40-16D resulted in blank plates (data not shown). The colonies on the YPEG+G418 plate are smeared due to replica plating and regrowth before the image was taken.

Another negative control in the mating screen, the mating of BMZ1-1 and NB40-16D, resulted in no growth on YPEG+G418 plates. Note that BMZ1-1 is isogenic to BMZ8-7, *except* that BMZ1-1 is ρ^0, and BMZ8-7 is the synthetic ρ^- containing pRPO41^m-rev. The mating of the two did not grow on media containing nonfermentable carbon as the sole carbon source because the *cox2-62* allele in NB40-16D was not complemented. Conversely, diploids BMZ5-(105-107) were respiratory proficient because there they each contained a full-length version of *COX2*, presumably through homologous recombination between pRPO41^m-rev at the *cox2* locus of NB40-16D.

To confirm insertion, we again performed an analytical PCR across the expected 5' homologous recombination site. The PCR schemes, as well as the agarose gel confirming insertion, are given as Figure 48.

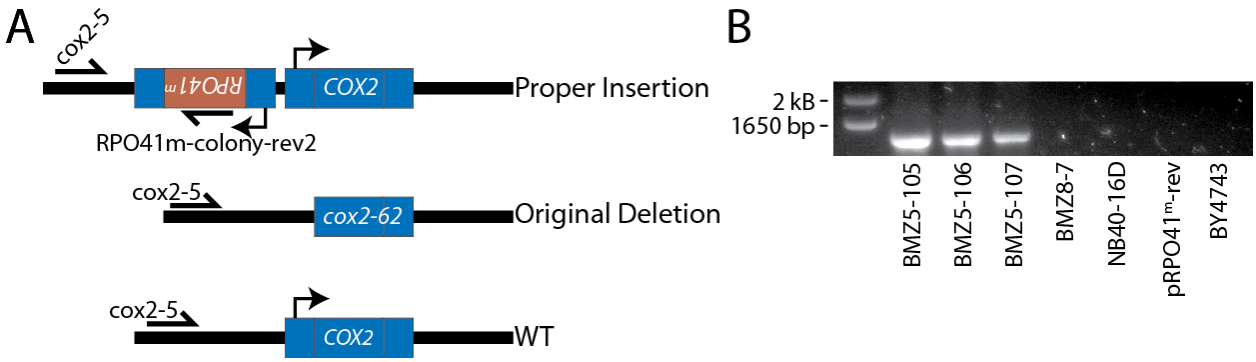

Figure 48: Insertion of *RPO41m* into the mitochondrial genome. A. A schematic of the PCR reaction used to determine insertion of *RPO41m* upstream of the *COX2* locus. **B.** An agarose gel showing the results of the PCR reaction shown in (A). Diploids that were the result of strain BMZ8-7 (the synthetic ρ^- containing pRPO41m-rev) and NB40-16D (the final mater containing the *cox2-62* allele) contain *RPO41m* inserted into their mitochondrial genome, whereas the synthetic ρ^- and the final mater strains do not. Additional negative controls were the bombardment plasmid (pRPO41m-rev), and a common wild-type strain (BY4743).

Expression of RPO41m *alone is unable to perform transcription to phenotypic levels*

Once the insertion *RPO41m* into the mitochondrial genome of NB40-16D to yield the BMZ5-(105-107) strains was confirmed, we attempted to see if its gene product was being expressed to phenotypic levels. To do this, we needed to "shut off" the nuclear expression of *RPO41*, and score for respiratory proficiency. Because our diploids were heterozygous at the nuclear *RPO41* locus, simple Mendelian segregation would result in a 2:2 ratio of haploids that contain nuclear *RPO41* to those containing *rpo41Δ::KanMX4*. Because we deleted the *RPO41* ORF using the *KanMX4* marker, we could determine which haploids had *RPO41* deleted by selection on media containing G418.

To isolate haploids from their diploid parents, we performed a standard sporulation and tetrad dissection procedure in which the diploid cells are sporulated in sporulation medium, their cell walls digested with zymolyase, and individual daughter cells segregated on an YPD plate with a microdissection needle. After growth, the colonies on the original dissection plate are then replica-plated onto other selectable media. The original dissection plate, as well as the replica plates to both YPEG and YPD+G418 are shown in Figure 49.

Note that the phenotypes of $G418^R$ and respiratory proficiency are mutually exclusive, *i.e.* all haploids that are $G418^R$ are mit^-, and all strains that are $G418^S$ are Mit^+. This is a consequence of the *KanMX4* cassette and the *RPO41* gene being at the same locus in the diploid, and is a clear example of 2:2 segregation of heterozygote alleles. The mit^- nature of

the *G418*^{R} and therefore *rpo41Δ* strains conclusively shows that the mitochondrial *RPO41*^{m} is not being expressed in such a fashion that it can be manifested phenotypically.

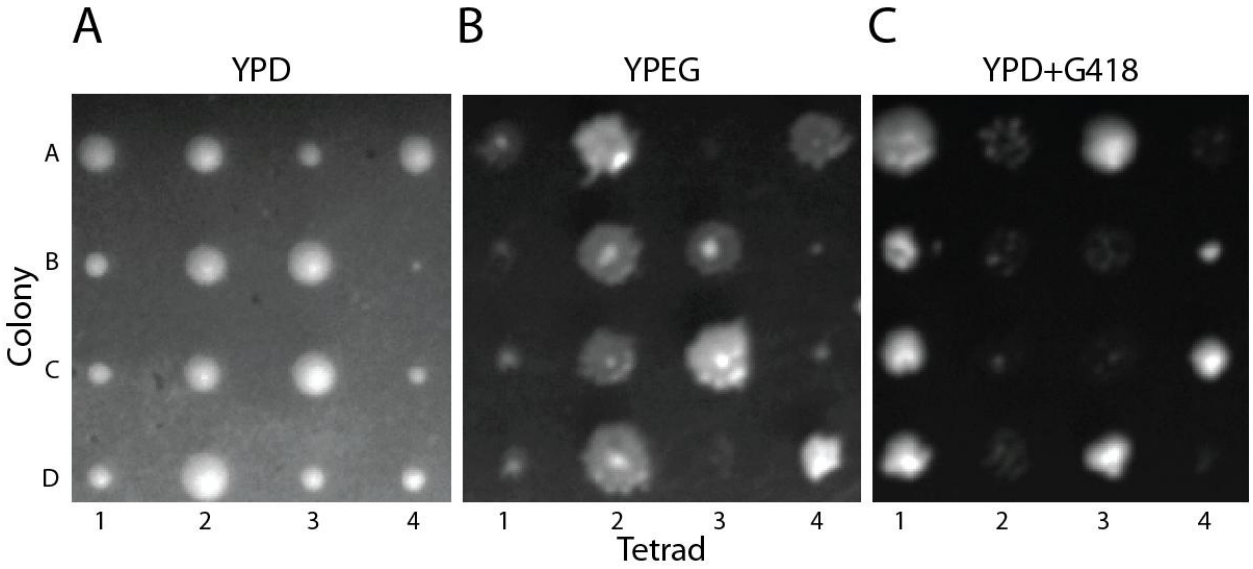

Figure 49: Tetrad dissection and phenotypic segregation of BMZ5-105 haploid offspring. A. Original dissection plate (YPD). **B.** Replica plate of original dissection plate to YPEG, which scores for respiratory-proficient colonies. **C.** Replica plate of original dissection to YPD+G418, which scores for strains with *RPO41* deleted from the nucleus. Note that all colonies that have nuclear *RPO41* deleted also do not respire, and vice versa, indicating that only nuclear *RPO41* (and conversely not mitochondrial *RPO41*^{m}) is expressed to phenotypic levels.

By scoring the haploid segregants of BMZ5-105 and BMZ5-106 for respiration directly, we skipped all of the steps of the central dogma, and looked only for the end result of expression. For a gene to be expressed, it must first be transcribed to mRNA, that mRNA must go through post-transcriptional modifications specific to the organism/organelle, then it must be translated to a protein, and that protein must go through specific post-translational and/or co-translational modifications specific to the organism/organelle. There is mRNA editing in yeast mitochondria, a consequence of editing polycistronic mitochondrial mRNAs[54]. Mitochondrial translation is also a highly coordinated process, in which the ribosome is linked to an imported translational co-factor that is specific to the mRNA being translated. This co-factor serves to tether the ribosome to the inner surface of the mitochondrial membrane, presumably to aid in co-translational insertion of the nascent polypeptide into the inner membrane, as all but one mitochondrially-encoded enzyme (Var1) is membrane-localized[54].

The fact that the introduced *RPO41*^{m} does not complement nuclear-encoded *RPO41* is at first surprising, since other soluble proteins have been expressed in similar ways[166,172]. To elucidate the reason for the failure of *RPO41*^{m} to support mitochondrial function we decided to monitor its expression in the organelle.

RPO41M *IS BEING TRANSCRIBED*

In order to determine if the *RPO41m* mRNA is being made, we performed a reverse-transcriptase PCR (RT-PCR) reaction on whole-cell RNA preparations, using primers specific to the *RPO41m*. Because RT-PCR is most efficient when the amplicon size is on the order of a hundred base pairs, we could not determine the intactness of the mRNA in a single step. Instead we amplified both ends of the mRNA, using two primer sets. The first primer set, RPO41m-start and RPO41m-RT-PCR-rev, amplified from the start codon of *RPO41m* until nucleotide 244. The second primer set, RPO41m-end and RPO41m-RT-PCR-fwd, amplified a 351 bp product ending with the terminal bp of the *RPO41m* ORF (Figure 50). As we are working in a diploid context, we had to design primers specific of the *RPO41m* gene. The differences in sequence between the mitochondrially and nuclear versions of *RPO41* allowed us to amplify *RPO41m* specifically.

Pictures of agarose gels showing the above mentioned RT-PCRs and associated controls are given in Figure 50B&C. Note that amplification occurs only in the diploid strain, and only in reactions in which there was a reverse-transcription step (lane 1). Lane 4 contains genomic DNA of the diploid, and is a positive control. The negative control in which the reverse-transcription step is omitted (RT-, lane 2) is a particularly important one, as RT-PCRs are prone to artifacts resulting from DNA contamination of the RNA preparation. There are no bands present in the RT- control (a consequence of careful optimization of the DNAse step after RNA purification) indicating that the amplification product is reverse-transcribed mRNA.

The second negative control, which consisted of the exact same reactions preformed with a template originating from a wild-type strain (BY4743), shows no amplification, RT dependent or otherwise. Because BY4743 contains *RPO41* (as scored by growth on YPEG), the lack of amplification when using preparations from that strain confirm that the primers used were specific to *RPO41m*, and that nonspecific amplification of *RPO41* (or any other gene present in the wild-type strain) was not the cause of the RT+ band using BMZ5-105 as a template.

In conclusion, then, the blank lanes present in Figure 50B and C, show that the reverse-transcriptase dependent amplification product off of BMZ5-105 RNA preparations are indeed (a) a measure of the RNA present in the cell, and (b) are *RPO41m* specific.

The fact that reverse-transcriptase-specific amplifications occur at both ends of the mRNA implies that a full length mRNA is being made. However, the reaction that took place at the 3' end did not lead to as bright of a band as that at the 5' end. Note that, because the amplicons differ, one cannot compare the densities of the two bands as a means to determine concentration. Nonetheless, similar RT-PCRs were performed later by Jean-Charles Blouzard, a postdoctoral successor to this project, and amplified overlapping regions of the entire synthetic construct. These gels are presented in Figure 50D-F. Two of the diploids in these reactions were BMZ43-1 and BMZ43-2. These strains are isogenic to BMZ5-105, only with the addition of additional selectable markers to the nuclear genome in order to perform mutagenesis (see discussion in Future Directions section). This strain has been characterized in all of the same ways as BMZ5-105, and is phenotypically identical to it (save the aforementioned additional selective markers).

In total seven RT-PCR reactions show that differing fragments of the *RPO41*^{m} mRNA is being transcribed, with full overlap along the ORF and the *cox2* UTRs. This is strong evidence that the full length mRNA is being made. Further steps to prove this are discussed in the Future Directions section of this chapter.

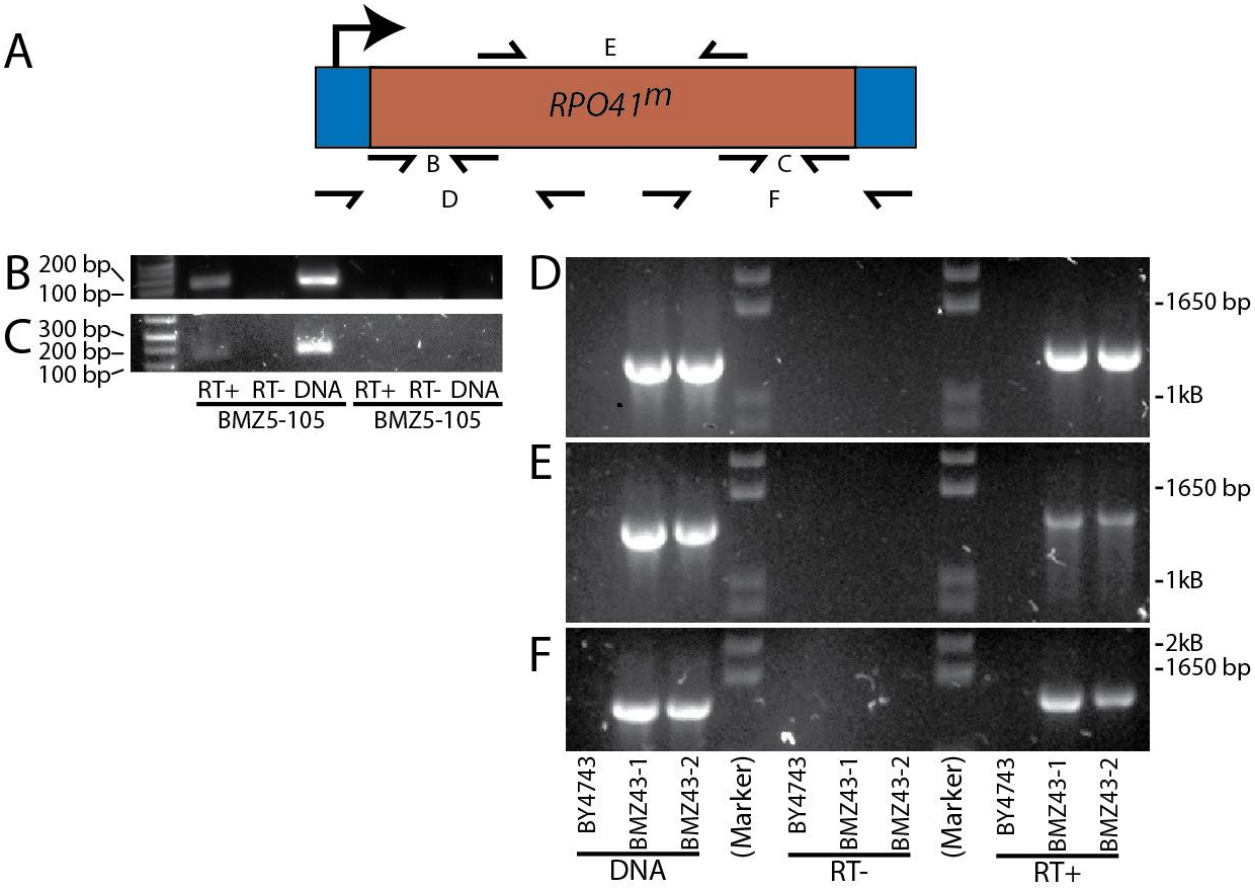

Figure 50: RT-PCR analysis indicates transcription of *RPO41*^{m} A. A schematic of the reverse-transcriptase PCR reactions performed in this study. Brown is the *RPO41*^{m} ORF, blue is the *cox2* UTRs, the full arrow indicates the transcriptional promoter, and the half-arrows indicate primer binding sites and directions of polymerization by the DNA polymerase used in the RT-PCR reaction. The letter between primer binding sites corresponds to the gels shown as labeled. RT-PCR reactions using the *COX2::RPO41*^{m} diploid strain constructed in this study, BMZ5-105, shows reverse-transcriptase-dependent amplification from both the beginning (**B**) and end (**C**) regions of the *RPO41*^{m} ORF. Reactions on BMZ43-1 and BMZ43-2, strains isogenic to BMZ5-105 save nuclear modifications as described in the text, show reverse-transcriptase-dependent amplification over the entire region of the synthetic insertion, including the beginning region containing the *cox2* 5' UTR (**D**), the middle of the *RPO41*^{m} ORF (**E**), and the end of the construct including the *cox2* 3' UTR (**F**). Reactions loaded in the lanes labeled RT+ had a reverse-transcription step prior to PCR amplification, whereas those labeled RT- did not. Lanes labeled DNA were positive controls using genomic DNA as a template, instead of RNA. As expected the WT strain, BY4743, which does not contain *RPO41*^{m}, shows no amplification.

Due to insufficient loading, the RT+ band in C is barely visible in electronic format, but was clear upon manual visualization. Gels D-F, including all preparations and RT-PCRs, were performed by Jean-Charles Blouzard.

Rpo41^{m}-*Specific Protein Expression is Not Detectable*

Detection of the mRNA resulting from transcription of *RPO41^{m}* was facilitated by the fact that *RPO41^{m}* and *RPO41* were different enough in sequence that RT-PCR primers could be designed to assay only the mitochondrially-encoded gene. In contrast, the protein produced via *RPO41^{m}* expression should be of identical sequence to that produced via expression of the regular *RPO41* gene, as our gene design specifically aimed to conserve the primary amino acid sequence of the protein. In order to measure of the Rpo41 protein that resulted from *RPO41^{m}*-specific expression, we utilized the method of *in vivo* labeling of mitochondrial translation products[178].

The *in vivo* labeling technique is based on the fact that, due to their prokaryotic origin, mitochondrial ribosomes are not inhibited by cycloheximide. Conversely, cycloheximide interferes with eukaryotic (cytoplasmic) ribosomes during the translocation step, drastically inhibiting translation[179]. This allows for a relatively simple pulse-chase experiment, in which cytoplasmic translation of a given strain is inhibited by incubation with cycloheximide, and mitochondrial translation products are labeled using methionine containing the radioisotope ^{35}S. Traditional polyacrylamide gel electrophoresis on the cell extracts, and subsequent exposure of the dried polyacrylamide gel to a phosphor screen, allows for visualization of the proteins that arise only as a result of mitochondrial translation, *i.e.* proteins that are encoded by genes on the mitochondrial genome.

We performed this *in vivo* labeling procedure on our diploid strains BMZ5-105 and BMZ5-106, as well as BMZ6-105-6 and BMZ6-106-6. BMZ6-105-6 and BMZ6-106-6 are BMZ5-105 and BMZ5-106 (respectively), transformed with pJM20, a vector which contains *PET111* under the control of a strong promoter. Pet111 activates *COX2* translation by tethering the *COX2* mRNA to the matrix side of the mitochondrial inner membrane, and recruiting the ribosome[180]. The Pet111-*COX2* mRNA interaction occurs at the *COX2* 5' UTR, and thus translation from heterologous mRNAs containing *COX2* 5' UTRs is up-regulated when Pet111 is overexpressed[180]. Because *RPO41^{m}* contains a *COX2* 5' UTR, we expect increased expression of *RPO41^{m}* in Pet111 overexpression strains.

As negative controls, we used strains BMZ33-1 and JRY4012. BMZ33-1 was created by mating MCC123 with NB40-16D. MCC123 is identical to MCC123*ρ^{0}*, except that it has a wild-type mitochondrial genome. Thus the mating between MCC123 and NB40-16D results in a diploid identical to the BMZ5-10x (*RPO41/rpo41Δ::KanMX* [*ρ^{+} RPO41^{m}::COX2*]) diploids except (a) it has wild-type mitochondrial genome, and (b) it is homozygous wild-type at the *RPO41* locus (*RPO41/RPO41* [*ρ^{+}*]). JRY4012 is a haploid W303 derivative and a common wild-type lab strain known for its robust respiration[181] (*RPO41* [*ρ^{+}*]).

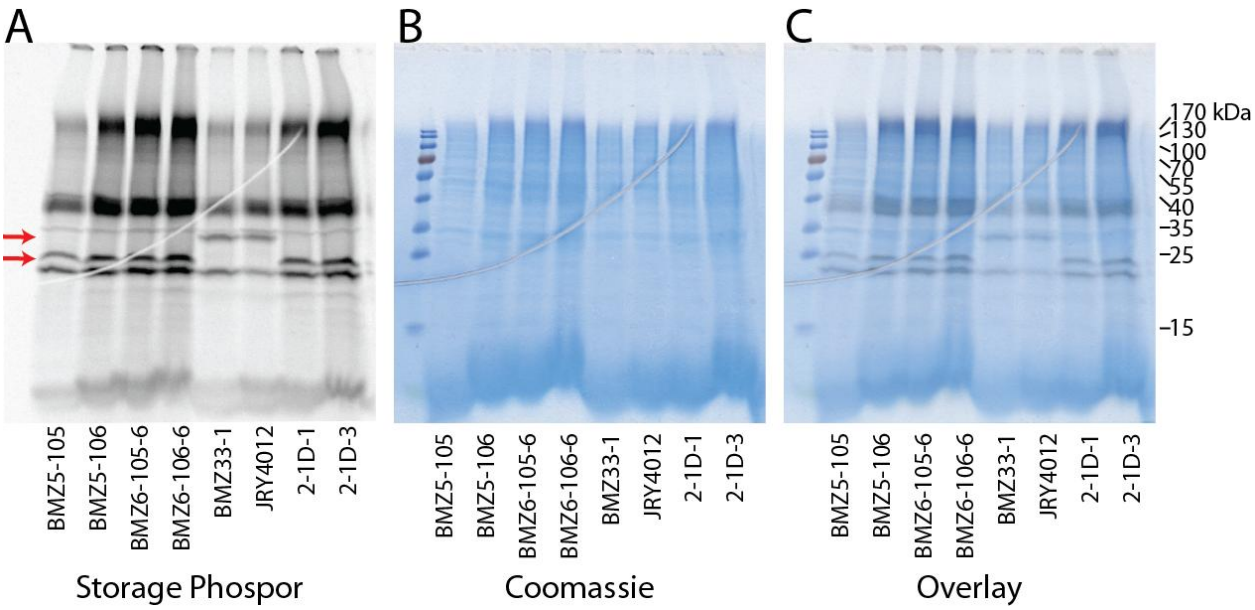

Figure 51: *In vivo* labeling of mitochondrial translation products shows a lack of mitochondrially-based Rpo41 expression. A. Storage phosphor radiogram showing mitochondrial-specific translation products. The red arrows indicate differences in the band pattern between strains with an unmodified mitochondrial genome (BMZ33-1 and JRY4012) and those with a modified mitochondrial genome (all others). **B.** The Coomassie stained gel used for the radiogram as described in A. This shows similar density of each lane indicates that roughly equal amounts of protein were loaded for each sample. C. An overlay of the radiogram and the Coomassie images.

We also performed the *in vivo* labeling protocol for an entirely different, though very similar project. Strains 2-1D-1 and 2-1D-3 were created in the exact same way as BMZ5-105 and BMZ5-106, except (a) they are homozygous wild-type at the *RPO41* locus, and (b) they have a completely different gene – mitochondrially-encoded proteorhodopsin – inserted upstream of *COX2* (in the same place as *RPO41*^{m} is in BMZ5-105 and BMZ1-106)[182]. These inadvertently served as even better negative controls than BMZ33-1 or JRY4012, as is discussed below.

A radiogram of the mitochondrial translation products for the above strains is given as Figure 51A, while the original, Coomassie-stained gel is given in Figure 51B, and the overlay as Figure 51C. Note that there is a difference in the band pattern for the strains in which the mitochondrial genome has been modified as compared to the in which it has not. Unfortunately, the change in the banding pattern is not specific for *RPO41*^{m}, as the mitochondrially-encoded proteorhodopsin strains have an identical pattern to the mitochondrially-encoded *RPO41* strains. Rather, it is specific to modification of the region upstream to *COX2*. This can be explained by a comment in one of the papers on *in vivo* labeling:

> *Cox2 is produced as a precursor protein with a small amino-terminal extension that is cleaved after the insertion of the protein into the inner membrane. The precursor and the mature forms of the protein run as sharp bands below Cox1*

signal. The maturation normally occurs cotranslationally; therefore, in wild-type cells only the mature form of Cox2 can be seen[178]*.*

Thus the difference in banding patters between wild-type strains and those involving mitochondrial insertions presumably arises from differences in the post-translational processing of Cox2 as opposed to heterologous gene expression in the mitochondrially modified strains.

We have exposed the phosphor membranes used in this experiment from one day to six months, and see no change in the banding pattern. Consequently, we can confidently conclude that, at least to the level of radiographic measurements, there is no Rpo41 protein being produced as a result of mitochondrial expression.

FUTURE DIRECTIONS

While the expression of a mitochondrially-encoded mitochondrial RNA polymerase in yeast to the mRNA level is an achievement, the goal of expression to the phenotypic level has not been reached. Obviously, the establishment of a minimal cell from a mitochondrion will require this step, as well as many others, in order to be completed. Here we propose the future steps we foresee being necessary for the expression of *RPO41*^{m} to the phenotypic level.

MUTAGENESIS

There can be many reasons why expression of *RPO41*^{m} to the protein level has not been achieved. If the genetic elements controlling expression (the promoter and the sequences modulating translational efficiency) result in too high of protein levels (overexpression), there will be a selective pressure against protein expression. While we have checked the *RPO41*^{m} mRNA and flanking sequences in the diploids, and found them to be intact, other regulatory elements may be mutated to maintain viability (but abolish *RPO41*^{m} expression) in those strains. These regulatory elements could be as disparate as distal, uncharacterized genetic elements, protein cofactors assisting in translation, or modulation of mitochondrial physiology through communication with the nucleus, an area of intense research. Conversely, if protein expression in the diploids is modulated to levels too low to maintain mitochondrial function (underexpression), strains expressing *RPO41*^{m} alone will be inviable.

Overexpression and underexpression are just the two most obvious mechanisms by which protein expression of *RPO41*^{m} could be hindered. As discussed below, internal transcriptional initiation may affect expression, as well as a myriad of other processes which we are unable to anticipate.

Mutagenesis is a powerful technique in which one can explore a virtually infinite space of genetic changes to a strain. It has been used throughout the biological fields, including yeast mitochondrial genetics[180]. Coupled with powerful selection, then, one could mutate just the correct elements in order to establish a strain with the desired phenotype, should such a strain be possible. It is the "powerful selection" requirement, however, that has challenged us in establishing an effective mutagenesis protocol for *RPO41*^{m} expression.

The end goal of this project, in experimental terms, is to isolate a haploid segregant of BMZ5-105 or BMZ5-106 that is capable of growing on nonfermentable carbon containing G418, *i.e.* one that respires while still containing the *rpo41Δ::KanMX4* cassette conferring G418 resistance. A naïve mutagenesis protocol, then, would be to mutagenize the diploids, and then score the haploid segregants on YPEG+G418. This is not possible, for random sporulation is never complete, and the unsporulated diploid strains would grow on YPEG+G418. Tetrad dissection, which results in plates only containing haploids, is a labor intensive procedure, taking on the order of one hour per ~100 haploids. Mutagenesis requires thousands to millions of selection events in order to successfully isolate the desired strain (if possible). Clearly, then, tetrad dissection is not a viable means to perform mutagenesis.

The fact that the BMZ5-10x diploids will grow on YPEG+G418 as well as its theoretical *rpo41Δ::KanMX4*, MIT$^+$ segregant is a specific example of a larger problem in mutagenesis. In general, a dominant marker that allows for selection against a particular haploid segregant will also select for its diploid parent cell. In this case, the heterozygous *RPO41/ rpo41Δ::KanMX4* allele confers the same phenotype (G418^R) than the desired *rpo41Δ::KanMX4* haploid segregant. What is needed, then, is a recessive marker. Unfortunately, recessive markers generally convey their selectable phenotype when they are in mutant form. Thus mutation of the wild-type form of that gene also results in the selectable phenotype. Consequently, mutagenesis and selection via a recessive marker would be plagued by the background of haploids with merely the recessive marker mutated, or heterozygous diploids in which the wild-type allele has been mutated.

For clarity, we give the canonical example of a gene used as a recessive marker: *CAN1*. The Can1 protein is located in the plasma membrane, and is the arginine permease, allowing the amino acid arginine to enter the cell. Canavanine, an arginine analog, is toxic to *CAN1* strains, as it can enter the cell via Can1, and can then be misincorporated during translation. Thus *CAN1* strains are canavanine sensitive (CAN^S). Conversely, *can1* strains cannot import arginine (they rely on metabolism as the source of intracellular arginine) or canavanine; consequently, they are canavanine resistant (CAN^R). *CAN1/can1* heterozygotes are CAN^S, giving the *can1* allele a recessive phenotype. Unfortunately, any mutation that turns the *CAN1* gene into a nonfunctional allele, either in a heterozygous diploid or a wild-type haploid, will make the strain CAN^R. Obviously, then, canavanine selection is not a plausible means to segregate haploids.

Fortunately a new technique has been developed in order to analyze haploids in a high throughput manner[183]. This "Synthetic Genetic Array" (SGA) technology allows for robust selection of haploids after mutagenesis.

The power of the SGA technology that we have chosen to employ is the *can1Δ::STE2pr-Sp_his5* cassette. A strain containing the *can1Δ::STE2pr-Sp_his5* cassette only expresses the *Sp_his5* gene, which encodes for the *HIS3* analog in *Schizosaccharomyces pombe*, in MAT**a** haploids (the MAT locus is what determines mating types. Haploids can be either MAT**a** or MATα, while diploids are MAT**a**/α. Mating occurs through complex signaling between MAT**a** and MATα cells). This is because the *Sp_his5* gene is driven by the *STE2pr* promoter, which is the promoter that nominally drives expression of the *α* mating type pheromone receptor. The *α* mating type pheromone receptor is only expressed in MAT**a** cells. Cells

with disruptions in the endogenous *his3* gene and containing the *can1Δ::STE2pr-Sp_his5* cassette, then, will only be His^+ when they are MAT**a** haploids. Mutagenesis is highly unlikely to modify the *STE2pr* such that it would work in diploid or MAT**a** cells, greatly reducing such a source of background during selection. Furthermore this *STE2pr-Sp_his5* cassette disrupts the *CAN1* gene, allowing for double recessive selection.

A future direction for this study, then, would be to use SGA technology as a means to performing mutagenesis on our strains. We propose the following specific strategy:

1. Genetically modify BMZ8-7 and NB40-16D to carry the selectable markers described above. This includes the deletion of the *HIS3* gene with a *MphMX* cassette, conferring hygromycin resistance.
2. Mate these strains to isolate a diploid isogenic to the BMZ5-10x strains, except with the added SGA markers.
3. Mutagenize this modified diploid using ethane methyl sulfonate (EMS), a global mutagen, and $MnCl_2$, a mutagen specific to the mitochondrial genome[180] (in separate reactions), along with proper controls.
4. Sporulate the diploid, digest the ascus, and plate the haploids *en masse* to selective media, scoring for Can^R, Hyg^R, $G418^R$, Mit^+, arg^-, His^+ haploids.
5. Perform transcriptional and translational assays (RT-PCR, northern blotting, *in vivo* labeling), as described in the Results and Discussion section of this chapter, on isolated haploids.
6. Characterize the nature of the mutations that allow for $RPO41^m$ expression in the haploids.

As stated above, the first step in the mutagenesis screen is to employ SGA technology in strains BMZ8-7 and NB40-16D. We have done this by inserting the *HphMX* cassette in place of the *his3* genes in NB40-16 and BMZ35-31, yielding strains BMZ32-27 and BMZ32-29, respectively. The *CAN1* locus was then disrupted in strain BMZ35-31 to make strain BMZ42-5, which was subsequently mated with strains BMZ32-27 and BMZ32-29 and to yield strains BMZ43-1 and BMZ42-5, respectively. These diploid strains are isogenic to the BMZ5-10x strains, and have even been characterized by RT-PCR by the successor to this project.

We have begun the third step in the above strategy, mutagenesis of the BMZ42-x strains using both EMS, and $MnCl_2$. Careful optimization of the dosage levels and exposure time of the mutagen, the duration and media of the recovery step(s), the sporulation and ascus digestion procedure, and the order and dosage of selection steps is underway.

Note that, as per step 6 in the above mutagenesis strategy, the isolation of a mutant $G418^R$ Mit^+ haploid through mutagenesis has an added benefit in that it may offer the opportunity to investigate the very mutations that allow for $RPO41^m$ expression, and may lead to new insights pertaining to mitochondrial expression.

Nonetheless, mutagenesis is not a panacea. It is a blunt instrument, and is no way guaranteed to work. Furthermore, even if we do isolate a viable mutant, we may not be able to characterize it sufficiently. Thus careful consideration and elucidation of our strains in a directed manner should – and is – be performed concurrently.

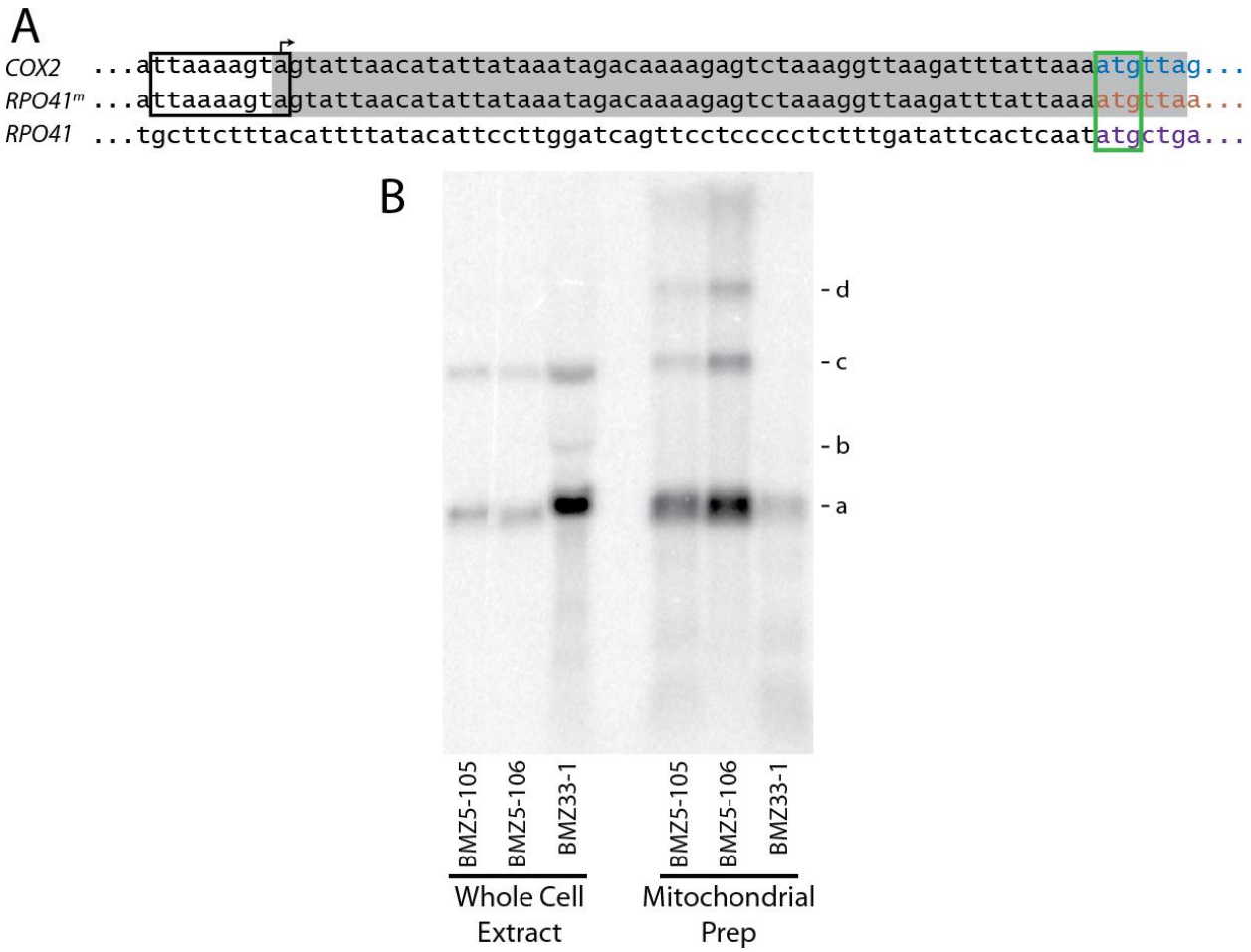

Figure 52: Northern blots on strains transcribing *RPO41m* yield unclear results. A. The probe used in analysis of *RPO41m* RNA is complementary to the region shaded in gray. This region is found in the *COX2* 5' UTR present in the wild-type *COX2* gene (top sequence) and the synthetic *RPO41m* construct (middle sequence), a region with no homology to that in the nuclear *RPO41* gene (bottom sequence). The *COX2* promoter is boxed in black, with the first transcribed nucleotide denoted with an arrow indicating the direction of transcription. The start codon for each sequence is boxed in green, with the gene ORF in blue (*COX2*), brown (*RPO41m*), or purple (*RPO41*). **B.** The storage phosphor image of the northern blot performed on RNA isolated from whole cell extracts (lanes 1-3) or purified mitochondria (lanes 4-6) of *RPO41m* expression strains (BMZ5-105 and BMZ5-106) and a negative control (BMZ33-1). Band (a) presumably comes from the *COX2* mRNA, as it is present in all strains. Band (b) is faint, and present only in the whole-cell extract of BMZ33-1; its identity is unexplained. Band (c) is present in all extracts except the mitochondrial extract of the negative control, and is of uncertain nature. Band (d) is present only in the mitochondrial extracts of the *RPO41m* expression strains, and may correspond to the *RPO41m* mRNA.

TROUBLESHOOTING THE NORTHERN BLOT

The fact that all of the mitochondrial-specific RT-PCRs described above worked over the entire synthetic *RPO41^m* construct (including *COX2* UTRs), is very strong evidence that the full *RPO41^m* mRNA is being made in BMZ5-105, BMZ5-106, BMZ43-1, and BMZ43-2. RT-PCR cannot, however, determine if there is only one species of mRNA being made. Furthermore, RT-PCR cannot determine the stability or integrity of the resultant mRNA, even if it is initially transcribed in full; post- or co- translational processing and degradation of the primary transcript may result in altered mRNAs. The northern blot technique allows one to distinguish if a particular sequence is present in mRNAs of differing length through the use of electrophoresis and hybridization.

We performed a northern blot using a probe that hybridizes to the mRNA of the *COX2* 5' UTR. This was done so that each sample would have a separate, internal positive control – the *COX2* mRNA itself. A diagram of the regions homologous to the probe is shown in Figure 52A, while the actual radiogram of the northern blot we performed is given as Figure 52B. We performed the northern blot procedure using RNA preparations from both whole cells and purified mitochondria from three strains: BMZ5-105, BMZ5-106, and BMZ33-1.

The bottom band in the radiogram (band "a" in Figure 52B) is present in all lanes, and is therefore most likely that of the *COX2* mRNA. The top major band (band "d") is present only in the mitochondrial extracts of the BMZ5-10x strains, and is in much lower abundance than the lowest (presumably *COX2*) bands in the same lanes. This difference in abundance may explain why these *RPO41^m*-specific bands are not present in the whole-cell extracts of the same strains, as approximately equal total amounts of total RNA was loaded to each lane, and thus the absolute amount of each mitochondrial mRNA is much less in the whole-cell lanes. Nonetheless, this band is clearly *RPO41^m* specific.

There is a second major band (band "c") that is present in all whole-cell extracts, as well as the mitochondrial extracts of the BMZ5-10x strains. The identity of this band is unclear, but because it is present in the whole-cell preparation of the BMZ33-1 strain, it cannot be *RPO41^m* specific. Thus the only *RPO41^m* specific product seen in Figure 52B is the top one, which is a distinct band of a single molecular weight.

Finally there is a band only visible in the whole-cell extract of the negative control strain BMZ33-1 (band "b"). It is not present in the mitochondrial extract of the same strain, indicating that it is of nuclear origin. However, it is not present in the whole cell extracts of BMZ5-105 or BMZ5-106, which share nearly identical nuclear genotypes. Thus the identity of said band is an anathema.

Taken in concert with the RT-PCR results, the results of our northern blot assays do strengthen the contention that transcription of *RPO41^m* results in a single (and presumably full-length) mRNA. Nonetheless, the above mentioned uncertainties in the results of the northern blot must be explained if we are to be sure that *RPO41^m* mRNA is being made correctly. To this end, we have begun to explore alternate hybridization probes for northern hybridization, a project that is continuing with the successor of this project. As we learned from the above-described *in vivo* labeling assays, one should also include a control with a different gene inserted upstream of *COX2* (e.g. the proteorhodopsin present in strains 2-1D-1 and 2-1D-3). This is especially imperative when one considers the presence of a cryptic promoter within the *RPO41^m* sequence (described below).

Removal of a Cryptic Promoter from the RPO41^{M} *Sequence*

When we designed *RPO41^{m}*, we were careful not to add the yeast mitochondrial consensus promoter to our sequence. This promoter is stated in the literature as a "highly conserved sequence of nine nucleotides," specifically ATATAAGTARY[23]. The *COX2* promoter, TTAAAAGTA, is not strictly a consensus promoter. This sequence exists in the *COX2* 3' UTR region of our synthetic construct by design. Unfortunately, it also exists within the synthetic *RPO41^{m}* ORF, 794 bp downstream of the first promoter, and 738 bp downstream of the *RPO41^{m}* start codon. This internal promoter is labeled as such in the full synthetic sequence shown in Appendix A.

The internal *COX2* promoter does not have any proximal translational regulatory elements present in the *COX2* 5' UTR. Therefore, transcription off of the internal promoter should not affect expression at the protein level. Nonetheless, especially in the light of the above-mentioned uncertainties with the mRNA analysis of *COX2*-related transcripts, this internal promoter should be removed and the entire project repeated with a new transformation plasmid.

Determination of the Maximal Transcriptional Capacity of the Mitochondrial Ribosome

Our experimental design was rooted in the fact that other proteins have been expressed to the phenotypic level in exactly the same manner (with the same strains and the same loci) as we have done, save the particular identity of *RPO41^{m}*. The inability, then, for our strains to express *RPO41^{m}* to the phenotypic level must lie in the differences between *RPO41^{m}* and the other genes successfully inserted at the same locus.

At 4186 bp, the length of the *RPO41^{m}* mRNA is longer than any ORF in the mitochondrial genome, including those that have been inserted. As described above, whether the full-length mRNA is being transcribed is a crucial question to this project, and is an active part of our current research. However, we do not anticipate the length of the *RPO41^{m}* mRNA as being a limiting factor in its expression, for most mitochondrial genes are transcribed in polycistronic transcripts, the largest extending over 10,500 bp[184]. However, Cox1, the largest protein endogenously encoded on the mitochondrial genome, has a molecular weight of 59 kDa and is composed of 534 AAs. The other proteins heterologously expressed in the mitochondria include Arg8 (47 kDa, 423 AAs)[58], GFP (28 kDa, 247 AAs)[166], and Rip1 (23 kDa, 215 AAs)[137]. In contrast, Rpo41 is 153 kDa and 1,351 AAs in length; it would be the largest protein ever to be expressed in the mitochondria by over a factor of two! The mitochondrial ribosome may simply not be able to transcribe such a large protein. It is crucial, then, to investigate the maximal capacity of the yeast mitochondrial translation system.

The successor to this project, Jean-Charles Blouzard, has designed and begun the initial cloning of the constructs shown in Figure 53. These constructs consist of repeats of the *ARG8^{m}* gene, followed by two *GFP^{m}* genes, both of which have been expressed in mitochondria[166,185]. This will allow for analysis of expression of genes of varying length (as determined by the number of *ARG8^{m}* repeats in the gene), and will be able to be quantitatively assayed via the fluorescence of the product of the *GFP^{m}* gene – green

fluorescent protein. Through these assays, we hope to determine the maximal protein length that can be translated by mitochondrial ribosomes. We also may be able to answer other questions concerning the dynamics and kinetics of mitochondrial translation.

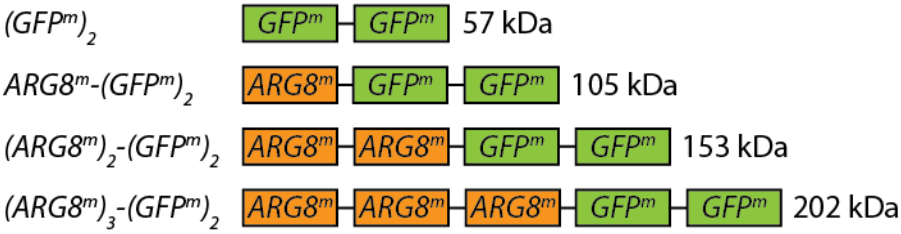

Figure 53: Constructs to determine the maximal translational capacity in mitochondria. Constructs consisting of various repeats of the *ARG8m* gene (orange), followed by two copies of the *GFPm* gene (green), are currently being cloned, and will be inserted in the place of *RPO41m* in pRPO41m. Transformation of the resultant plasmids, insertion of these constructs into the mitochondrial genome, and their subsequent expression will be selectable on media lacking arginine. Those constructs in which *GFPm* are expressed will be screenable using fluorescence, allowing for determination of the maximal protein size that can be translated by the mitochondrial translational machinery. The mass of the expressed proteins is given to the right of each construct.

Usage of Different Expression Strategies

The method we have used in this study – insertion of *RPO41m* just upstream of the *COX2* locus, is just one of a few methods by which heterologous genes have been inserted and expression in yeast mitochondria. For example, a number of proteins, including *GFP*[166] and *ARG8*[58], have been inserted in upstream of the *COX3* locus. We chose to avoid this locus at the outset, as the UTRs for that region are much larger than those used for insertion of genes upstream of *COX2*. In fact, the *COX2* UTRs are both the smallest and the best characterized, which was a major reason for our choosing of that region as the initial region for our project. Furthermore, the *COX2* UTRs were added to *RPO41m* during the synthesis of the DNA, and because we wanted to change as little as possible from the sequences that have been shown in the literature to work (albeit with BARSTAR instead of Rpo41), we did not add any restriction sites between the *RPO41m* ORF and the *COX2* UTRs. Therefore, replacing the *COX2* UTRs with the *COX3* UTRs requires site-directed mutagenesis and cloning. Nonetheless, insertion of the *RPO41m* upstream of the *COX3* locus may result in *RPO41m* expression to the phenotypic level through changes the levels of expression, differences in post-transcriptional and post-translational processing events, or by other unanticipated mechanisms.

The expression of a mitochondrial form of the gene encoding for the Rieske Iron Sulfur Protein (*RIP1m*) was performed via an insertion of the gene between the *COX1* and *ATP8*[137]. In order to avoid instability resulting from having repeated sequences in the mitochondrial genome, the authors used *COX1* sequences from *Saccharomyces douglasii* as homologous regions for recombination, while maintaining the endogenous *COX1* promoter for initiation of transcription of the *RIP1m* gene. A schematic of the insertion even they used is given in

Figure 54. We have these strains and plasmids in our laboratory (though they will require some modification), and are currently in the process of performing recombineering[186] experiments to switch the *RIP1*^{m} ORF with that of *RPO41*^{m}.

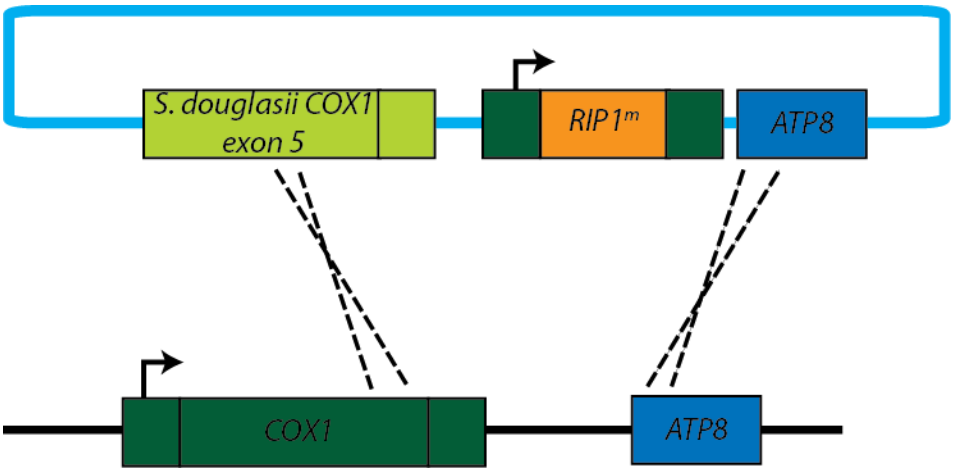

Figure 54: An alternative location for insertion of exogenous genes in the mitochondrial genome. The strategy used for insertion and expression of *RIP1*^{m} (orange) in the mitochondrial genome[137]. *RIP1*^{m} was inserted in between *COX1* and *ATP8* (dark blue). The *COX1* gene from *S. douglasii* (light green) was used in the transformation plasmid (light blue) instead of the *S. cerevisiae COX1* gene in order to mitigate instability of the mitochondrial genome due to repeated sequences. *RIP1*^{m} is flanked by *COX1* UTRs (dark green boxes). Replacement of *RIP1*^{m} with *RPO41*^{m} will result in an alternative expression system for *RPO41*^{m}.

MATERIALS AND METHODS

STANDARD TECHNIQUES AND STARTING MATERIALS.

Standard molecular biology techniques (restriction digestion, ligation, transformation, etc.) are as commonly described[101]. Plasmid minipreps from *E. coli* were performed using the Zymo Research (Orange, CA) Zyppy Plasmid Miniprep Kit. Gel extractions were performed using the Qiagen (Valencia, CA) QIAquick Gel Extraction Kit. Sequencing was performed as non-premix reactions by Sequetech, Inc. (Mountain View, CA). All enzymes were from New England Biolabs (Ipswich, MA), unless otherwise noted. All chemicals were purchased from Sigma-Aldrich (St. Louis, MO).

Unless otherwise noted, PCR reactions contained 1×Thermopol buffer, 0.5 µM of each primer, and 0.2 mM each NTP. Taq DNA polymerase (25 U/mL) was used for analytical PCRs, while Vent (20 U/mL) was used for cloning. Thermocycling consisted of an initial 5 min. denaturation step at 95°C, followed by thirty cycles of 95°C for 30 s, the annealing temperature for 30 s, and 72°C for the extension time. A final extension step of 72°C for 10 min. followed the last cycle. Extension times and annealing temperatures are given for each reaction.

Yeast protocols were as published[121]. Yeast transformations were performed via the high-efficiency LiAC protocol[187].

A list of primers used in this study is given in Table 9. All primers were synthesized by Integrated DNA Technologies, Inc. (San Diego, CA).

Yeast strain genotypes are given in Table 8.

Strains Y8205 and JRY4012 were kind gifts from J. Rine.

Strains BY4743 and BY4743 YFL036W were kind gifts from A. Deutschbauer.

Strains MCC123, MCC123ρ^0, NB40-16D, HMD7, and TF183 were kind gives of T. Fox, as were plasmids pHM102 and pJM20.

Plasmid pJJ1148 and was a kind gift from J. Jaehning.

Rich media for yeast growth contained 1% yeast extract, 2% bacto-peptone, and 2% carbon source. These media are designated "YP" in the text, with a third letter designating the carbon source. YPD contained 2% glucose, YPG contained 2% glycerol, YPEG 2% ethanol and 2% glycerol, YPGal contained 2% galactose, and YPRaff contained 2% raffinose. YPDA was YPD as above, supplemented with adenine to 100 mg/L.

Synthetic media for yeast growth contained 0.67% yeast nitrogen base without amino acids, amino acids as described[121] and carbon sources as described for YP media. These media are denoted "S" below, with the carbon source designation as a third letter, e.g. SD for synthetic media containing glucose, SG for synthetic media containing glycerol, etc. Dropout media is further delineated by the specific amino acids excluded from the mixture, e.g. SD-Ura is synthetic media containing all amino acids except uracil.

For solid media, agar (1.5% w/v) was added to all solutions before autoclaving.

Yeast antibiotic selection was done at 200 µg/mL for G418, 300 µg/mL for hygromycin (hyg), 60 µg/mL for canavanine (can), and 1 mg/mL for 5-flouroorotic acid (5FOA).

Table 8: Yeast strains used in this study

Name	Genotype	x =
2-1D-x[a]	MAT***a***/α *leu2-3,112/LEU2 lys2/LYS2 ura3-52/ura3 ade2/ADE2 kar1-1/KAR1 arg8::hisG/ARG8* [*ρ+ PR^{m}*]	1, 3
BMZ1-x[b]	MCC123*ρ^{0} rpo41Δ::KanMX4*	1-3
BMZ2-2[b]	NB40-16D pJJ1148 (*RPO41*)	
BMZ3-x[b]	BMZ2-2 *rpo41Δ::KanMX4*	1, 2
BMZ3-x-1[b]	NB40-16D *rpo41Δ::KanMX4*	1, 2
BMZ4-x[b]	BMZ1-1 [*ρ^{-}* pRPO41^{m}]	1, 3, 5-8
BMZ6-105-6[b]	BMZ5-105 pJM20 (*PET111*)	
BMZ6-106-6[b]	BMZ5-106 pJM20 (*PET111*)	
BMZ5-x (1-72)[b]	MAT***a***/α *leu2-3,112/LEU2 lys2/LYS2 ura3-52/ura3 ade2/ADE2 kar1-1/KAR1 arg8::hisG/ARG8 rpo41Δ::KanMX4/rpo41Δ::KanMX4* pJJ1148-ura^{-} [*ρ+ RPO41^{m}*]	1-72
BMZ5-x (105-107)[b]	MAT***a***/α *leu2-3,112/LEU2 lys2/LYS2 ura3-52/ura3 ade2/ADE2 kar1-1/KAR1 arg8::hisG/ARG8 rpo41Δ::KanMX4/RPO41* [*ρ+ RPO41^{m}-rev*]	105-107
BMZ8-x[b]	BMZ1-1 [*ρ^{-}* pRPO41^{m}-rev]	1, 3, 4, 5, 7
BMZ32-11[b]	NB40-16D *his3Δ::hphMX*	
BMZ35-9[b]	BMZ8-7 *his3Δ::hphMX*	
BY4743[167]	MAT***a***/α *his3Δ/his3Δ leu2Δ/leu2Δ LYS2/lys2Δ met15Δ/MET15 ura3Δ/ura3Δ*	
BY4743 YFL036W[167]	BY4743 *rpo41Δ::KanMX4*	
HMD7[172]	MATα *lys2 [ρ^{+}cox2-107]*	
JRY4012[181]	MAT***a*** *ura3 lys2 his3-11,15 leu2-3,112 trpl-l can1-100*	
MCC123[172]	*MAT**a** ade2 ura3 kar1-1*	
MCC123ρ^{0} [172]	MCC123 [*ρ^{0}*]	
NB40-16D[172]	MATα *leu2-3,112 lys2 ura3-52 arg8::hisG [ρ^{+} cox2-62]*	
TF183[165]	MAT***a*** *ura3-52 inol-13* [ρ^{-} *COX2* (pMT36)]	
Y8205[183]	MATα *can1Δ::STE2pr-Sp_his5 lyp1Δ::STE3pr-LEU2 his3Δ leu2Δ ura3Δ*	

[a]JM Walter, BM Zamft, A Vila-Sanjurjo, J. Liphardt, CJ Bustamante, unpublished
[b]This study

Table 9: Primers used in this study

Name	Sequence (5' - 3')
CAN_A	ctatcaatgaaaatttcgaggaaga
CAN_D	gacgtgaagataacgaaaaatgagt
cox2-11	tatgaatgatgtaccaacaccttatg
cox2-5	attcacatctccttcggccg
cox2-6	tatgaatgatgtaccaacaccttatg
cox2-8	ggagaacaattaaagtaaagagctgctggtc
his3-hphMX-fwd	tatactaaaaaatgagcaggcaagataaacgaaggcaaagcggatccccgggttaattaa
his3-hphMX-rev	tatatatatcgtatgctgcagctttaaataatcggtgtcacgatgaattcgagctcgttt
KanB	ctgcagcgaggagccgtaat
KanC	tgattttgatgacgagcgtaat
pRPO41-colony-rev-2	ctgatcctttcgctgtaaatcctg
RPO41_A	tctccgtgaagatattggtaaaaag
RPO41_B	ttgaaagggtcaaactctaattctg
Rpo41_C	gaaatgttggtgaaaagatggttac
RPO41-fwd	ggagaacaattaaagtaaagagctgctggtc
RPO41m-end	ttatgaaaaaaaatattgtgaatttcttaatact
RPO41m-RT-PCR-fwd	ggtaaattaaaaagatcaacagatttagctcaaaaaattattagaatta
RPO41m-RT-PCR-rev	gataagaaattgaagatttagcatcttttctatttttattaaataatgga
RPO41m-start	atgttaagaccagcttataaatcattagtaaaaacatctttattac

DESIGN AND CLONING OF A MITOCHONDRIAL EXPRESSION VECTOR

The Rpo41 protein sequence was entered into an Excel spreadsheet and reverse-translated using the most frequently used codon for that amino acid, as found in the Codon Usage Database[b]. EcoRI sites and consensus mitochondrial promoters were removed and, when necessary, the second or third most optimal codon was used. The 5' and 3' untranslated leader sequences (UTRs) of *COX2*, as described in the literature[172], were added to the 5' and 3' ends of the sequence, respectively. EcoRI sites were added to the distal ends, and this composite sequence was submitted to GenScript (Piscataway, NJ) for secondary structure analysis and synthesis. GenScript returned the synthetic DNA cloned into the EcoRI sites of pUC57, named pUC57-Rpo41^{m}. The sequence of the *RPO41^{m}* insert was confirmed via sequencing both by GenScript and by ourselves.

pUC57-RPO41^{m} and pHM102[172] were digested with EcoRI. The 4.2 kB fragment of pUC57-RPO41^{m} and the ~5 kB fragment from pHM102 were gel extracted, ligated together, and transformed into *E. coli* DH5α cells, to yield pRPO41^{m} and pRPO41^{m}-rev. pRPO41^{m} contains *RPO41^{m}* with *cox2* UTR flanks in the same orientation relative to *COX2* within the plasmid, while they are in the opposite orientation in pRPO41^{m}-rev.

[b] http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4932.mitochondrion

DNA from positive transformants was isolated by miniprep, analyzed by restriction digestion using EcoRI and PstI, and confirmed by sequencing.

Creation of a mitochondrial transformation strain

We deleted *RPO41* in strain MCC123ρ^0 by standard procedures, as outlined by the *Saccharomyces* genome deletion project[167]. We performed a PCR reaction on genomic DNA from strain BY4743 YFL036W using primers RPO41_A and RPO41_D, with an annealing temperature of 45°C, and an extension time of 5 min. This amplified the *KanMX4* cassette present in that strain, with 235 bp and 342 bp flanks identical to the upstream and downstream flanks of the *RPO41* gene, respectively.

This PCR product was transformed into MCC123ρ^0 using the standard high-efficiency transformation protocol[188]. Successful transformants were selected on YPD+G418 plates, streaked on the same media, and single colonies grown in 5 mL liquid YPD+G418 culture. This culture was used to make archival glycerol stocks, as well as for isolation of genomic DNA.

To confirm insertion of the *KanMX4* gene into the *RPO41* locus, genomic DNA from transformants, as well as wild-type (BY4743) and *rpo41*-deleted (BY4743 YFL036W) strains was subjected to analytical PCR using RPO41-fwd and either KanB or Rpo41_B as reverse primers, using an annealing temperature of 45°C, and an extension time of 2 min. These *rpo41* deleted strain was denoted BMZ1-1, BMZ1-2, and BMZ1-3.

Creation of a Final Mating Strain with Conditional RPO41 *Expression*

Strain NB40-16D was transformed with plasmid pJJ1148 and selected on SD-Ura plates, to yield strain BMZ2-2. This strain was then transformed with the same *KanMX4* cassette used to generate strain BMZ1-1. Ura$^+$ G418^R transformants were then screened using the above described analytical PCR and strains that contained both the *KanMX4* and *RPO41* genes was labeled BMZ3-1 and BMZ3-2.

BMZ3-1 and BMZ3-2 were grown for several rounds on YPD, and replica plated onto SD+5FOA and SD-Ura plates. Two 5FOA^R ura$^-$ strains were picked. These strains, BMZ3-1-1 and BMZ3-2-1, were subjected to analytical PCR for confirmation of the presence of the *KanMX4* cassette and the absence of the *RPO41* gene.

A streak of strain BMZ3-1 and BMZ3-2 were then replica plated to a freshly spread lawn of strain TF183, and incubated at 30°C for two days. These mating plates were then replica plated to YPEG plates, and incubated at 30°C for two days to score for growth. Diploid strains that were able to respire were identified by growth on these YPEG plates (a "marker rescue"), indicating the presence of the initial ρ^+ genome present in strain NB40-16D.

Mitochondrial Transformation via High Velocity Microprojectile Bombardment

BMZ1-1 mitochondria were transformed with plasmids pRPO41^m and pRPO41^m-rev as per published protocols[173]. Thirty micrograms of dry, 0.6 μM tungsten particles ("microcarriers," Bio-Rad, Hercules, CA) were resuspended 1 mL 70% EtOH, vortexed at maximum speed for 5 min., and incubated at RT for 15 min. This mixture was then washed three times by pelleting the solution for a few seconds, aspirating the supernatant, and

resuspending in sterile water. The microcarriers were then pelleted one more time, resuspended in 500 µL sterile 50% glycerol, and stored at -20°C for later use.

To coat microcarriers with DNA, 4.25 µg of pRS316 and 32 µg of either pRPO41^{m} or pRPO41^{m}-rev were premixed, and then 100 µL of the prepared microcarriers were added. After vigorous vortexing, 40 µL 100 mM spermidine and 100 µL 2.5 M ice-cold $CaCl_2$ were added, and the reaction incubated on ice, with occasional vortexing, for 10 min.

After incubation, the coated microcarriers were pelleted by centrifugation for a few seconds, and washed by resuspension in 1 mL 100% EtOH. During each wash, clumps of microcarriers were broken up using the pipet tip, and by a microcentrifuge pestle (USA Scientific, Orlando, FL). Washes were repeated until no clumps were visible. The solution was then pelleted one more time and resuspended in 60 µL of 100% EtOH.

The coated, washed microcarriers were precipitated on macrocarrier disks (Bio-Rad) that had been pre-sterilized in 80% EtOH and pre-seated onto macrocarrier holders, by aliquoting 10 µL of the microcarrier solution to the center of each macrocarrier and allowing them to dry.

The transformation strain BMZ1-1 was grown to saturation in 50 mL in YPRaff, pelleted by centrifugation at 1000×g for 5 min., washed in sterile water, and plated to a final concentration of 1×10^{8} cells per plate on SD-Ura plates containing G418 and 1 M sorbitol. Cells were spread using glass rollers, paying careful attention to spread the cells uniformly, and left face-up to dry.

Transformation plates were bombarded with coated microcarriers using the Helios® gene gun and PDS-1000/He™ system (Bio-Rad) as per the manufacturer's instructions, using at 1350 PSI rupture disk pre-sterilized in isopropanol, and the maximum vacuum pressure (~29.5 in. Hg.).

The transformation plates were incubated at 30°C until colonies (nuclear transformants) were observed (3-4 days).

Identification of Mitochondrial Transformants

A marker rescue procedure similar to that used with strain BMZ3-1 above was used to distinguish mitochondrial transformants. Nuclear transformants were replica-plated to YPD plates containing a freshly-spread lawn of strain HMD7 and incubated at 30°C for three days. These mating plates were then replica plated to YPEG plates and incubated at 30°C for three days. The corresponding nuclear transformants that rescued were then re-streaked on SD-Ura+G418 plates, and the process of mating on YPD and scoring for growth on YPEG repeated until the full streaks rescued. These stains, labeled BMZ4-5 (containing pRPO41^{m} in its mitochondria) and BMZ8-7 (containing pRPO41^{m}-rev) were then grown in liquid SD-Ura+G418, stored as archival glycerol stocks, and assayed as described in the Results and Discussion section.

Cytoduction/Mating

Cytoduction was essentially performed as described in[173]. One milliliter of synthetic ρ^{-} strains (either BMZ4-5 or BMZ8-7) were mixed with 0.2 mL final mating strains (NB40-16D or BMZ3-1). Negative controls included the individual strains themselves, as well as

mixtures of the final mating strains with BMZ1-1. Each mixture was then centrifuged at 600×g for 30 s, the supernatant aspirated, and the cell pellet resuspended in ~200 µL leftover supernatant. This culture was then spread to a YPDA plate, and incubated at 30°C for 3 hr. Plates were then scraped and the mixture was used to inoculate 4 mL YPDA. This culture was incubated at 30°C for 12 hr., and then spread in serial dilutions to YPEG plates.

To determine if the colonies that grew on YPEG plates were diploid or haploid, the plates were replica plated onto a set of dropout plates to check the auxotrophic markers. Those that were prototrophic for all markers were identified as diploids and discarded; the rest were subjected to PCRs specific to the strain.

Diploids were grown in liquid YPEG+G418 and stored in archival glycerol stocks. Strains originating from BMZ4-5 were labeled BMZ5-1 through BMZ5-72; strains originating from BMZ8-7 were labeled BMZ5-105, BMZ5-106, and BMZ5-107.

Confirmation of RPO41^{m} *Insertion into the Mitochondrial Genome*

To determine insertion of *RPO41^{m}* in the reverse orientation to *COX2*, primers cox2-5 and pRPO41-colony-rev-2 were used with no annealing step, and an extension step of 2.5 minutes at 64°C. Template DNA was from yeast genomic preps of strains BMZ5-105, BMZ5-106, BMZ5-107 and negative controls BMZ8-7, NB40-16D, and BY4743, as well as miniprepped DNA of plasmid pRPO41^{m}-rev.

Inverse PCR

Various amounts of genomic DNA from strain BMZ5-1 – from 0 to 7.5 µL – were cut using 25 U ApaI restriction enzyme (along with a no-enzyme control) in 1×NEB buffer 4 and 1×BSA. Digests were incubated at 37°C for 2 hr., and cleaned using the Qiagen PCR Purification Kit with 30 µL elution in water. Twenty-five microliters of this cleaned product was used as a template in two standard Taq PCR reactions with either (i) cox2-6 and cox2-11, or (ii) cox2-8 and cox2-11 as primers. Both reactions used a 30 s, 48°C annealing step, and a 5 min., 60°C extension step. Ten microliters of each PCR reaction was analyzed via electrophoresis using 1% TAE-agarose gel. Reactions with bands on the gel were cleaned using the Qiagen PCR purification kit, and sent for sequencing using the primers used in the PCR.

Reverse Transcriptase PCR

Strains BY4743 and BMZ5-105 were grown in 5 mL cultures of YPEG and YPEG+G418, respectively. Whole cellular RNA was isolated from these strains using the Zymo Research (Irvine, CA) YeaStar RNA Kit as per the manufacturer's instructions. Contaminating DNA was digested by adding 7 µL 10×Roche DNAse buffer (Roche Diagnostics, Indianapolis, IN) and 3 µL Roche DNAse I to 60 µL of the eluate from the YeaStar RNA Kit, This reaction was incubated at 37°C for 1 hr., and then incubated at 75°C for 10 min. The concentration of the resultant RNA was determined by NanoDrop (ThermoScientific, Wilmington, DE), and the RNA preparations were diluted to 10 ng/µL. RT-PCR was performed using the Qiagen OneStep RT-PCR Kit. A master-mix for each reaction was made containing buffer, dNTPs, primers, RNase Out, and the reverse-transcriptase/HotStart Taq DNA polymerase enzyme

mix as per the manufacturer's instructions, as well as the RNA template to a final concentration of 1.6 ng/μL. The final volume of this master mix was 40 μL, which was then split into two 20 μL reactions. One aliquot of each condition was incubated at 50°C for 30 min. (the RT step), 6°C for 5 min. (the pause step), 95°C for 15 min., 58°C for 30 s, 60°C for 30 s, and then subjected to 29 cycles of 95°C for 30 s, 58°C for 30 s, and 60°C for 1 min. The no-reverse-transcriptase controls were performed by taking the second aliquot from the 40 μL master mix, and adding it to the thermocycler during the pause step above.

An identical (positive control) reaction to the above reaction (including the initial RT step), was performed using whole cell genomic DNA as a template, which was prepared as described above. The final concentration of the DNA template was 7.5 ng/μL.

IN VIVO *LABELING OF MITOCHONDRIAL TRANSLATION PRODUCTS*

The below protocol for *in vivo* labeling of mitochondrial translation products is based on that described in the literature[178]. All growth/incubation steps involving whole cells were performed at 30°C with shaking at 200 RPM. Pelleting of cells was performed by centrifugation at 5,000×g for 5 min., while pelleting of the cell supernatant was performed 16,000×g for 30 min.

Cultures of strains were grown in 2 mL YPGal (for nonrespiring strains) or YPEG (for respiring strains) until saturation (~ 2 days), upon which time 2 mL of fresh, pre-warmed medium was then added and the cells grown for an additional 2.5 hr. The cells were then pelleted, washed twice with 2 mL sterile distilled water, resuspended in 2 mL synthetic media (with appropriate carbon source) lacking methionine and tryptophan, and incubated for an additional 30 min. A freshly-prepared solution of 10 mg/mL cycloheximide in 100% EtOH was then added to a final concentration of 0.1 mg/mL. After 5 min. of incubation, 0.1 mCi of ^{35}S-methionine (Perkin-Elmer, Waltham, MA) was added to the cultures. After one hour of incubation, 2 mL of chase solution (2% casamino acids, 2 mg/mL Na_2SO_4) was added, and the cultures incubated for an additional 10 min. The cultures were then pelleted, washed one time with chase solution, two times with ice-cold MTE (250 mM mannitol, 20 mM Tris-Cl, pH 7.4, 1 mM EDTA), and resuspended in 200 μL MTE. All subsequent steps were performed on ice or at 4°C in a refrigerated centrifuge.

Borosilicate beads (1 mm in diameter) were added to each culture, and each tube was vortexed at the maximum speed for 5 min. The beads were then allowed to settle, and the liquid aspirated and saved. The process of MTE addition, vortexing, and aspiration was repeated four times, for resulting in a total of 800 μL of aspirant.

The aspirant was then pelleted and the supernatant was discarded. The pellet was resuspended in 500 μL MTE, repelleted, and resuspended to a final volume of 20 μL in MTE. These final whole cell lysates were then run in an 8-15% Tris-Tricine polyacylamide gel (Bio-Rad Inc., Hercules, CA) as per the standard protocols, with the sole exception that the samples were not boiled before loading. The gels were then stained using Coomassie stain, and dried using the ThermoScientific Owl gel drying kit, as per the manufacturer's instructions. The dried gels were then used to expose the storage phosphor screens (GE Healthcare, Piscataway, NJ) for times ranging from one day to six months. The phosphor screens were then imaged using a Typhoon Variable Mode Imager (GE Healthcare, Piscataway, NJ) using the "storage phosphor" setting.

Creation of Synthetic Genomic Array-Type Strains

The *his3Δ::hphMX* cassette (with flanks homologous to those of the *HIS3* ORF) was PCR amplified using primers his3-hphMX-fwd and his3-hphMX-rev, and plasmid pAG32[189] as a template. The *can1Δ::STE2pr-Sp_his5* cassette (with flanks homologous to those of the *CAN1* ORF) was amplified using primers CAN_A and CAN_D, and genomic DNA from strain Y8205 as a template. Both PCR reactions used an annealing temperature of 47°C and an extension time of 3 min.

The *his3Δ::hphMX* PCR product was used to transform strains NB40-16D and BMZ8-7 to yield strains BMZ32-27, BMZ32-29 (derived from NB40-16), and BMZ35-31 (derived from BMZ8-7). The *can1Δ::STE2pr-Sp_his5* PCR product was used to transform BMZ35-31 to yield strain BMZ42-5.

Strain BMZ42-5 was mated with BMZ32-27 to yield strain BMZ43-1; and BMZ42-5 was mated with strain BMZ32-29 to yield strain BMZ43-2.

All transformants were tested via analytical PCR using the same reaction as described above for BMZ5-105 and BMZ5-106. Strains BMZ32-x and BMZ35-31 were tested phenotypically by growth on media containing hygromycin, and lack of growth on media lacking histidine. Strain BMZ42-5 was phenotypically verified by these above two assays, as well as for growth on media lacking arginine and containing canavanine. All BMZ8-7-derived strains were tested for mitochondrial integrity by mating with HMD7 and subsequent growth on YPEG. All NB40-16D-derived strains were tested for mitochondrial integrity by mating with TF183 and subsequent growth on YPEG.

Chapter 5: CONCLUSION

The goal of any dissertation is to demonstrate the breadth and depth of one's graduate work. The breadth of my graduate work, as I have tried to display here, contains the ends of the spectrum of the way scientists study and even think about biological systems. Consider only the imagery that has flashed through one's head while reading the research presented here. In Chapter 2 we thought of picomoles – a hundred billion molecules or so – of Rpo41 sitting on DNA as stable elongation complexes, some of them dead or slow, some waiting at the starting gates. We thought of these molecules running in aggregate – uncoordinated, asynchronous, towards the end of the template, and considered, on average, how fast this community went, and how stable they were as they were going. We thought of hundreds of billions of RNA molecules slithering through the netting of a polyacrylamide gel, glowing beta particles from their 5' ends.

In the latter part of Chapter 2 and the entirety of Chapter 3, we imagined a lone Rpo41 or Pol II molecule, bathed in laser light, drawn and quartered by two tethers. We discussed the processes of the nucleotide addition cycle, and imagined a lone polymerase in thermal motion, shaking and vibrating between two states. We watched in our mind's eye a hairpin form near the exit channel of the transcribing polymerase, and saw as the two bumped together from time to time, preventing a would-be backtrack. The pictures in our minds zoomed in to the very point at which a new nucleotide enters the core of the polymerase, or of the hopping of an RNA secondary structure between open and closed states: images whose frames measure a few ångströms on each side.

Chapter 4 brought us from the bulk biochemical assays that belong in the scale of the first part of Chapter 2, to even bigger considerations. We went from considering picomoles of molecules to thousands of cells. We imagined the shmooing of haploid yeasts, the mixing of their cytoplasms, the fusing of their nuclei, and the insertion of foreign genomic material into a mitochondrial genome. We considered hundreds of billions of Rpo41 molecules transcribing billions of genomes, their products being translated by a similar number of mitochondrial ribosomes, and the resultant proteins slithering through polyacrylamide, streaming beta particles once again.

The thinking required to perform such experiments also spanned a spectrum: the spectrum from the thinking of a "pure biologist" to a "pure physicist," though the existence of such pure forms of thinking is a fallacy. In Chapter 4 we used no math, save arithmetic and the occasional exponential growth calculation. Instead, we were consumed by considerations of creating a well-controlled, well designed experiment involving a vast number of components and variables. Indeed, with the scale of an experiment comes uncertainty, and with uncertainty comes the need to spend more time in consideration and preparation. The experiments involved in Chapter 4 hinged on cells, containing more than 4905 independent components (only the number of characterized ORFs in the

Saccharomyces Genome Database[c]). We aimed to investigate how far we could manipulate these cells, and used a system in which we attempted to move the location of a gene encoding for a core physiological enzyme from one genome to another.

In Chapter 2 and Chapter 3 we used an isolated system, composed of no more than four general components (the RNAP, DNA, RNA, and RNase). Thus the means of thinking about experiments were in some ways a bit easier than for those in Chapter 4 – if something went wrong, for example, we investigated the effect of changing one component at a time – though the experimental design and use of proper controls was no less essential. However, the relative simplicity of the system allowed us to investigate the fundamental principles underlying transcription, to design kinetic models incorporating these principles, and to test these models experimentally. This involved not only months of experimental design and data collection, but perhaps an order of magnitude more mathematics, computer programming, and data analysis – a process that will commence with the writing of the final word of this dissertation.

The knowledge gained from such disparate investigations was no less diverse. In Chapter 2 we characterized the yeast mitochondrial RNA polymerase, Rpo41, by bulk methods and, for the first time, by single molecule assays. We found that Rpo41 is able to transcribe without its initiation factor, if a transcription bubble is pre-formed using synthetic DNA and RNA oligonucleotides. We found that NTP is limiting for bulk runoff transcription in the sub-micromolar range. We further optimized the buffer conditions for Rpo41 transcription, finding that a $MgCl_2$ concentration in the 5-20 mM range, a KCl concentration of 40 mM, and 10 mM DTT resulted in the greatest amount of transcript runoff.

The single molecule studies in Chapter 2 resulted in the first evidence that Rpo41, a single-subunit RNAP from the bacteriophage lineage, does pause and backtrack. We developed a novel method of statistically determining the pause-free velocity of the polymerase in a way that does not rely on pause-picking algorithm, and in turn found the kinetic values of nucleotide incorporation by the enzyme: K_m = 22±13 μM^{-1} and v_{max} = 25±2.5 bp/s.

Chapter 3 used the system established in Chapter 2, as well as previously established single molecule assays of Pol II, to investigate modifications to the model of transcriptional pausing. By performing single-molecule transcriptional assays by both enzymes on templates of varying base pair composition, we found that GC content attenuates pause entry and aids pause recovery; pause durations and pause densities both decrease on templates of relatively higher GC content. We investigated models by which the template could modulate the kinetics of pausing, and found that one in which an energy penalty is introduced into the backward rate of backtracking predicts the difference in mean pause duration, mean pause density, and the distribution of pause durations amongst templates. We proposed that this energy penalty is due to secondary structure of the nascent RNA preventing backward movement during backtracking, and supported this by showing that,

[c] http://yeastgenome.org/cache/genomeSnapshot.html

in the presence of RNase A, the characteristics of pausing on all templates by both enzymes are indistinguishable from their respective pause characteristics on the AT-rich template without RNase A.

In Chapter 4 we investigated the principles of *in vivo* mitochondrial transcription in yeast, and attempted to modify that system to make the organelle transcriptionally independent. We moved the gene encoding for the core RNA polymerase, Rpo41, from the nuclear genome to that in the mitochondria itself, and found that it is being transcribed. We have come upon a limit of its expression at the translational level, and have begun further studies of the very limitations of the mitochondrial ribosome and associated factors.

The above accomplishments are not enough. Indeed, nothing is ever enough for science. Only when science and theology meet, a situation which could very well be used as the definition of an asymptote, do they end. But the studies described within this dissertation, in addition to adding to the body of scientific knowledge, open doors to new investigations and to new understandings of the physical and biological world. The details of the future directions that can be taken from this research has been described in each chapter, and likely only scratches the surface of the complexity of the systems we purport to understand a bit better here. Only time will tell where this research goes, but there can be little doubt that no matter where it does, given enough time to incubate in the halls of academia, the insight gained will be grand.

Here we conclude a journey that has taken the much greater part of a decade. We have spanned the spectrum of biological and biophysical experiment and thought. We have pushed the boundaries of knowledge an iota further, and I have grown intellectually from a student to a research scientist. Though it was often difficult, frustrating, and sometimes downright enraging, the process of scientific investigation and its endpoint – the *creation* of knowledge, the seeing of a gel band trace or that no one in history has ever seen before – has been worth it.

As Sir Isaac Newton once said,

> *I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me*[190]*.*

We know not what tomorrow brings, but today we stake our claim to a grain of sand on Newton's shore.

WORKS CITED

1. Crick, F.H. On protein synthesis. *Symposia of the Society for Experimental Biology* **12**, 138-63(1958).

2. Wöhler, F. Ueber künstliche Bildung des Harnstoffs. *Annalen der Physik und Chemie* **88**, 253-256(1828).

3. Kauffman, G.B. & Chooljian, S.H. Friedrich Wöhler (1800-1882), on the Bicentennial of His Birth. *The Chemical Educator* **6**, 121-133(2001).

4. Fessner, W. *Biocatalysis : from discovery to application*. (Springer: New York, 2000).

5. Kresge, N., Simoni, R.D. & Hill, R.L. Otto Fritz Meyerhof and the Elucidation of the Glycolytic Pathway. *J. Biol. Chem.* **280**, e3-(2005).

6. Cohn, S.A. The Mechanochemistry of Kinesin. *Molecular and Chemical Neuropathology* **12**, 83-94(1990).

7. Szybalski, W. In vivo and in vitro initiation of transcription. *Advances in experimental medicine and biology* **44**, 23-4(1974).

8. Weiss, S.B. & Gladstone, L. A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. *Journal of the American Chemical Society* **81**, 4118-4119(1959).

9. Davenport, R.J. et al. Single-molecule study of transcriptional pausing and arrest by E. coli RNA polymerase. *Science (New York, N.Y.)* **287**, 2497-500(2000).

10. Yin, H. et al. Transcription against an applied force. *Science (New York, N.Y.)* **270**, 1653-7(1995).

11. Forde, N.R. et al. Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by Escherichia coli RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11682-7(2002).

12. Rivetti, C., Guthold, M. & Bustamante, C. Wrapping of DNA around the E.coli RNA polymerase open promoter complex. *The EMBO journal* **18**, 4464-75(1999).

13. Galburt, E.A. et al. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**, 820-823(2007).

14. Hodges, C. et al. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science (New York, N.Y.)* **325**, 626-8(2009).

15. Galburt, E.A., Grill, S.W. & Bustamante, C. Single molecule transcription elongation. *Methods (San Diego, Calif.)* **48**, 323-32(2009).

16. Thomen, P., Lopez, P. & Heslot, F. Unravelling the Mechanism of RNA-Polymerase Forward Motion by Using Mechanical Force. *Physical Review Letters* **94**, 1-4(2005).

17. Thomen, P. et al. T7 RNA Polymerase Studied by Force Measurements Varying Cofactor Concentration. *Biophysical journal* **95**, 2423-2433(2008).

18. Skinner, G.M. et al. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. A single-molecule view of the transcription cycle. *The Journal of biological chemistry* **279**, 3239-3244(2004).

19. Masters, B.S., Stohl, L.L. & Clayton, D.A. Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7 . *Cell* **51**, 89-99(1987).

20. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature reviews. Microbiology* **9**, 85-98(2011).

21. Cheetham, G.M., Jeruzalmi, D. & Steitz, T.A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80-83(1999).

22. Savkina, M. et al. Multiple functions of yeast mitochondrial transcription factor Mtf1p during initiation . *The Journal of biological chemistry* (2009).

23. Tracy, R.L. & Stern, D.B. Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Current genetics* **28**, 205-16(1995).

24. Skinner, G.M. et al. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. A single-molecule view of the transcription cycle. *The Journal of biological chemistry* **279**, 3239-44(2004).

25. Nayak, D., Guo, Q. & Sousa, R. A promoter recognition mechanism common to yeast mitochondrial and phage t7 RNA polymerases . *The Journal of biological chemistry* **284**, 13641-13647(2009).

26. Cliften, P.F., Jang, S.H. & Jaehning, J.A. Identifying a core RNA polymerase surface critical for interactions with a sigma-like specificity factor. *Molecular and cellular biology* **20**, 7013-23(2000).

27. Gong, P. & Martin, C.T. Mechanism of instability in abortive cycling by T7 RNA polymerase. *The Journal of biological chemistry* **281**, 23533-23544(2006).

28. Kapanidis, A.N. et al. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science (New York, N.Y.)* **314**, 1144-7(2006).

29. Revyakin, A. et al. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science (New York, N.Y.)* **314**, 1139-43(2006).

30. Zhou, Y. et al. Dissociation of halted T7 RNA polymerase elongation complexes proceeds via a forward-translocation mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 10352-7(2007).

31. Mason, P.B. & Struhl, K. Distinction and Relationship between Elongation Rate and Processivity of RNA Polymerase II In Vivo. *Molecular cell* **17**, 831-840(2005).

32. Yin, Y.W. & Steitz, T.A. The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase. *Cell* **116**, 393-404(2004).

33. Guo, Q. & Sousa, R. Translocation by T7 RNA polymerase: a sensitively poised Brownian ratchet. *Journal of molecular biology* **358**, 241-54(2006).

34. Abbondanzieri, E.A. et al. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460-5(2005).

35. Svetlov, V. & Nudler, E. Jamming the ratchet of transcription. *Nature structural & molecular biology* **15**, 777-9(2008).

36. Bar-Nahum, G. et al. A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183-93(2005).

37. Banerjee, S. et al. Rho-dependent transcription termination: more questions than answers. *Journal of microbiology (Seoul, Korea)* **44**, 11-22(2006).

38. Richardson, J.P. Transcription termination. *Critical reviews in biochemistry and molecular biology* **28**, 1-30(1993).

39. Song, H. & Kang, C. Sequence-specific termination by T7 RNA polymerase requires formation of paused conformation prior to the point of RNA release. *Genes to Cells* **6**, 291-301(2001).

40. Larson, M.H. et al. Applied force reveals mechanistic and energetic details of transcription termination. *Cell* **132**, 971-82(2008).

41. Lykke-Andersen, S. & Jensen, T.H. Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* **89**, 1177-82(2007).

42. Schinkel, A.H., Groot Koerkamp, M.J. & Tabak, H.F. Mitochondrial RNA polymerase of Saccharomyces cerevisiae: composition and mechanism of promoter recognition. *The EMBO journal* **7**, 3255-3262(1988).

43. Matsunaga, M. & Jaehning, J.A. Intrinsic promoter recognition by a "core" RNA polymerase. *The Journal of biological chemistry* **279**, 44239-44242(2004).

44. Jang, S.H. & Jaehning, J.A. The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. *The Journal of biological chemistry* **266**, 22671-22677(1991).

45. Mangus, D.A., Jang, S.H. & Jaehning, J.A. Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes . *The Journal of biological chemistry* **269**, 26568-26574(1994).

46. Schubot, F.D. et al. Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. *Protein science : a publication of the Protein Society* **10**, 1980-1988(2001).

47. Gruber, T.M. & Gross, C.A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual review of microbiology* **57**, 441-66(2003).

48. Selth, L.A., Sigurdsson, S. & Svejstrup, J.Q. Transcript Elongation by RNA Polymerase II. *Annual review of biochemistry* **79**, 271-93(2010).

49. Gustafsson, C.M. et al. Identification of new mediator subunits in the RNA polymerase II holoenzyme from Saccharomyces cerevisiae. *The Journal of biological chemistry* **273**, 30851-4(1998).

50. Lipinski, K.A., Kaniak-Golik, A. & Golik, P. Maintenance and expression of the S. cerevisiae mitochondrial genome-From genetics to evolution and systems biology. *Biochimica et biophysica acta* **1797**, 1086-98

51. Biswas, T.K. & Getz, G.S. Position-specific inhibition of yeast mitochondrial transcription by a poly(T) sequence. *Journal of Molecular Biology* **275**, 547-560(1998).

52. Mueller, D.M. & Getz, G.S. Transcriptional regulation of the mitochondrial genome of yeast Saccharomyces cerevisiae. *The Journal of biological chemistry* **261**, 11756-64(1986).

53. Pfeuty, a et al. Mitochondrial upstream promoter sequences modulate in vivo the transcription of a gene in yeast mitochondria. *Mitochondrion* **6**, 289-98(2006).

54. Costanzo, M.C. & Fox, T.D. Control of mitochondrial gene expression in Saccharomyces cerevisiae. *Annual review of genetics* **24**, 91-113(1990).

55. Ulery, T.L., Jang, S.H. & Jaehning, J.A. Glucose repression of yeast mitochondrial transcription: kinetics of derepression and role of nuclear genes. *Molecular and Cellular Biology* **14**, 1160-1170(1994).

56. Amiott, E.A. & Jaehning, J.A. Sensitivity of the Yeast Mitochondrial RNA Polymerase to +1 and +2 Initiating Nucleotides. *Journal of Biological Chemistry* **281**, 34982-34988(2006).

57. Amiott, E. a & Jaehning, J. a Mitochondrial transcription is regulated via an ATP "sensing" mechanism that couples RNA abundance to respiration. *Molecular cell* **22**, 329-38(2006).

58. Steele, D.F., Butler, C.A. & Fox, T.D. Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 5253-5257(1996).

59. Rogowska, A.T. et al. Balance between transcription and RNA degradation is vital for Saccharomyces cerevisiae mitochondria: reduced transcription rescues the phenotype of deficient RNA degradation. *Molecular biology of the cell* **17**, 1184-93(2006).

60. Malecki, M. et al. In vivo and in vitro approaches for studying the yeast mitochondrial RNA degradosome complex. *Methods in enzymology* **447**, 463-88(2008).

61. Naithani, S. et al. Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of Saccharomyces cerevisiae. *Molecular biology of the cell* **14**, 324-33(2003).

62. Green-Willms, N.S. et al. Pet111p, an inner membrane-bound translational activator that limits expression of the Saccharomyces cerevisiae mitochondrial gene COX2. *The Journal of biological chemistry* **276**, 6392-7(2001).

63. Fiori, A., Perez-Martinez, X. & Fox, T.D. Overexpression of the COX2 translational activator, Pet111p, prevents translation of COX1 mRNA and cytochrome c oxidase assembly in mitochondria of Saccharomyces cerevisiae. *Molecular microbiology* **56**, 1689-704(2005).

64. Greenleaf, A.L., Kelly, J.L. & Lehman, I.R. Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3391-4(1986).

65. Wang, Y. & Shadel, G.S. Stability of the mitochondrial genome requires an amino-terminal domain of yeast mitochondrial RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 8046-8051(1999).

66. Contamine, V. & Picard, M. Maintenance and integrity of the mitochondrial genome: a plethora of nuclear genes in the budding yeast. *Microbiology and molecular biology reviews : MMBR* **64**, 281-315(2000).

67. MacAlpine, D.M. et al. Replication and preferential inheritance of hypersuppressive petite mitochondrial DNA. *The EMBO journal* **20**, 1807-17(2001).

68. Zamaroczy, M. de et al. The origins of replication of the yeast mitochondrial genome and the phenomenon of suppressivity. *Nature* **292**, 75-8(1981).

69. Fangman, W.L., Henly, J.W. & Brewer, B.J. RPO41-independent maintenance of [rho-] mitochondrial DNA in Saccharomyces cerevisiae. *Molecular and cellular biology* **10**, 10-5(1990).

70. Mehta, A.D. Single-Molecule Biomechanics with Optical Methods. *Science* **283**, 1689-1695(1999).

71. Moffitt, J.R. et al. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9006-11(2006).

72. Moffitt, J.R. et al. Recent advances in optical tweezers. *Annual review of biochemistry* **77**, 205-28(2008).

73. Ashkin, A. et al. Observation of a single-beam gradient force optical trap for dielectric particles. *Optics letters* **11**, 288(1986).

74. Holmberg, A. et al. The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **26**, 501-10(2005).

75. Prakash, O. & Eisenberg, M.A. Biotinyl 5'-adenylate: corepressor role in the regulation of the biotin genes of Escherichia coli K-12. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 5592-5(1979).

76. Schatz, P.J. Use of Peptide Libraries to Map the Substrate Specificity of a Peptide-Modifying Enzyme: A 13 Residue Consensus Peptide Specifies Biotinylation in Escherichia coli. **11**, 1138-1143(1993).

77. Matsunaga, M., Jang, S.-heon & Jaehning, J.A. Expression and purification of wild type and mutant forms of the yeast mitochondrial core RNA polymerase , Rpo41. *Protein Expression and Purification* **35**, 126-130(2004).

78. Kane, J.F. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Current opinion in biotechnology* **6**, 494-500(1995).

79. *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*. (Merck and Co., Inc: Whitehouse Station, NJ, 2006).

80. Mangus, D.A. & Jaehning, J.A. Transcription in Vitro with Saccharomyces cerevisiae Mitochondrial RNA Polymerase. *Methods in Enzymology* **264**, 57-66(1996).

81. Hodges, C. et al. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science (New York, N.Y.)* **325**, 626-8(2009).

82. Hodges, H.C. Single-Molecule Studies of Gene Expression. 108(2009).

83. Bintu, L. Dynamic Interactions and Molecular Rearrangements Occurring when RNA Polymerase II Meets the Nucleosome. 70(2010).

84. Bustamante, C. et al. Entropic elasticity of lambda-phage DNA. *Science (New York, N.Y.)* **265**, 1599-600(1994).

85. Jr. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **46**, 68 - 78(1951).

86. Michaelis, L. & Menten, M. Die kinetik der invertinwirkung. *Biochem. Z* **49**, (1913).

87. Martin, W. et al. An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biological chemistry* **382**, 1521-1539(2001).

88. Gray, M.W. Mitochondrial Evolution. *Science* **283**, 1476-1481(1999).

89. Cermakian, N. et al. On the Evolution of the Single-Subunit RNA Polymerases. *Journal of Molecular Evolution* **45**, 671-681(1997).

90. Cermakian, N. Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Research* **24**, 648-654(1996).

91. Shutt, T.E. & Gray, M.W. Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends in Genetics* **22**, 90-95(2006).

92. MacArthur, R.H. & Wilson, E.O. *The theory of island biogeography*. *Monographs in Population Biology* **1**, 203(Princeton University Press: 1967).

93. Schäfer, B. RNA maturation in mitochondria of S. cerevisiae and S. pombe. *Gene* **354**, 80-5(2005).

94. Shibata, T. & Ling, F. DNA recombination protein-dependent mechanism of homoplasmy and its proposed functions. *Mitochondrion* **7**, 17-23

95. Lecrenier, N. New features of mitochondrial DNA replication system in yeast and man. *Gene* **246**, 37-48(2000).

96. Sugimoto, K., Kohara, Y. & Okazaki, T. Relative roles of T7 RNA polymerase and gene 4 primase for the initiation of T7 phage DNA replication in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 3977-81(1987).

97. Davanloo, P. Cloning and Expression of the Gene for Bacteriophage T7 RNA Polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 2035-2039(1984).

98. Sousa, R. et al. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature* **364**, 593-9(1993).

99. Zawadzki, V. & Gross, H.J. Rapid and simple purification of T7 RNA polymerase. *Nucleic acids research* **19**, 1948(1991).

100. Tannous, B.A. T7 RNA polymerase as a self-replicating label for antigen quantification. *Nucleic Acids Research* **30**, 140e-140(2002).

101. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning*. *New York* **3**, 5463(Cold Spring Harbor Laboratory Press: 1989).

102. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5(1970).

103. Odijk, T. Stiff Chains and Filaments under Tension. *Macromolecules* **28**, 7016-7018(1995).

104. Depken, M., Galburt, E. a & Grill, S.W. The origin of short transcriptional pauses. *Biophysical journal* **96**, 2189-93(2009).

105. Komissarova, N. & Kashlev, M. RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. *The Journal of biological chemistry* **272**, 15329-38(1997).

106. Komissarova, N. & Kashlev, M. Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 1755-60(1997).

107. Kramers, H. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284-304(1940).

108. Kireeva, M.L. & Kashlev, M. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 8900-5(2009).

109. Herbert, K.M. et al. Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell* **125**, 1083-1094(2006).

110. Tadigotla, V.R. et al. Thermodynamic and kinetic modeling of transcriptional pausing. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4439-44(2006).

111. Klopper, A.V., Bois, J.S. & Grill, S.W. Influence of secondary structure on recovery from pauses during early stages of RNA transcription. *Physical review. E, Statistical, nonlinear, and soft matter physics* **81**, 030904(2010).

112. Kireeva, M., Kashlev, M. & Burton, Z.F. Translocation by multi-subunit RNA polymerases. *Biochimica et biophysica acta* **1799**, 389-401

113. Sugimoto, N. et al. Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry* **34**, 11211-11216(1995).

114. Breslauer, K.J. Predicting DNA Duplex Stability from the Base Sequence. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3746-3750(1986).

115. Mathews, D.H. et al. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology* **288**, 911-40(1999).

116. Wuchty, S. et al. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**, 145-65(1999).

117. Nichols, N.M. & Yue, D. Ribonucleases. *Current protocols in molecular biology* **Chapter 3**, Unit3.13(2008).

118. Gu, W. Increased accommodation of nascent RNA in a product site on RNA polymerase II during arrest. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6935-6940(1996).

119. Komissarova, N. Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proceedings of the National Academy of Sciences* **94**, 1755-1760(1997).

120. Geis, M. et al. Folding kinetics of large RNAs. *Journal of molecular biology* **379**, 160-73(2008).

121. Adams, A., Gottschling, D.E. & Kaiser, C.A. *Methods in Yeast Genetics*. 198(Cold Spring Harbor Laboratory Press: 2005).

122. Báez-Viveros, J.L. et al. Metabolic engineering and protein directed evolution increase the yield of L-phenylalanine synthesized from glucose in Escherichia coli. *Biotechnology and bioengineering* **87**, 516-24(2004).

123. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nature biotechnology* **23**, 612-6(2005).

124. Dueber, J.E. et al. Synthetic protein scaffolds provide modular control over metabolic flux. *Nature biotechnology* **27**, 753-9(2009).

125. Glass, J.I. et al. Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 425-30(2006).

126. Hutchison, C.A. et al. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* **286**, 2165-2169(1999).

127. Chaudhuri, R.R. et al. Comprehensive identification of essential Staphylococcus aureus genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC genomics* **10**, 291(2009).

128. Pal, C. et al. Chance and necessity in the evolution of minimal metabolic networks. *Nature* **440**, 667-670(2006).

129. Azuma, Y. & Ota, M. An evaluation of minimal cellular functions to sustain a bacterial cell. *BMC Systems Biology* **3**, 111(2009).

130. Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17669-17674(2004).

131. Noireaux, V. et al. Toward an artificial cell based on gene expression in vesicles. *Physical biology* **2**, P1-8(2005).

132. Chakrabarti, A.C. et al. Production of RNA by a polymerase protein encapsulated within phospholipid vesicles. *Journal of Molecular Evolution* **39**, 555-559(1994).

133. Yu, W. et al. Synthesis of functional protein in liposome. *Journal of bioscience and bioengineering* **92**, 590-593(2001).

134. Kita, H. et al. Replication of genetic information with self-encoded replicase in liposomes. *ChemBioChem* **9**, 2403-2410(2008).

135. Walde, P. Building artificial cells and protocell models: experimental approaches with lipid vesicles. *BioEssays : news and reviews in molecular, cellular and developmental biology* **32**, 296-303(2010).

136. Scheffler, I.E. *Mitochondria*. 367(Wiley-Liss: New York, 1999).

137. Golik, P. et al. The Rieske FeS protein encoded and synthesized within mitochondria complements a deficiency in the nuclear gene. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8844-8849(2003).

138. Beckmann, J.D. et al. Isolation and characterization of the nuclear gene encoding the Rieske iron-sulfur protein (RIP1) from Saccharomyces cerevisiae. *The Journal of biological chemistry* **262**, 8901-9(1987).

139. Towpik, J. Regulation of mitochondrial translation in yeast. *Cellular & molecular biology letters* **10**, 571-94(2005).

140. Bayona-Bafaluy, M.P., Fernandez-Silva, P. & Enriquez, J.A. The thankless task of playing genetics with mammalian mitochondrial DNA: a 30-year review. *Mitochondrion* **2**, 3-25(2002).

141. Fernandez-Silva, P., Enriquez, J.A. & Montoya, J. Replication and transcription of mammalian mitochondrial DNA. *Experimental physiology* **88**, 41-56(2003).

142. Shoji, Y. et al. Tissue distribution of mutant mitochondrial DNA in mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS). *J Inherit Metab Dis.* (1993).

143. Keränen, T. & Kuusisto, H. NARP syndrome and adult-onset generalised seizures. *Epileptic Disorders* **8**, 200-203(2006).

144. Howell, N. LHON and other optic nerve atrophies: the mitochondrial connection. *Developments in Ophthalmology* **37**, 94-108(2003).

145. Mellitus., E.C. on the D. and C. of D. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes care* **26 Suppl 1**, S5-20(2003).

146. Wallace, D.C. Mitochondrial diseases in man and mouse. *Science* **283**, 1482-1488(1999).

147. Skladal, D., Halliday, J. & Thorburn, D.R. Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain : a journal of neurology* **126**, 1905-12(2003).

148. Bibb, M.J. et al. Sequence and gene organization of mouse mitochondrial DNA. *Cell* **26**, 167-180(1981).

149. Anderson, S. et al. Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-465(1981).

150. Polli, I., with Zardi L Human mitochondrial DNA. *Journal of Molecular Biology* **36**, 419-423(1968).

151. Yoon, Y.G. & Koob, M.D. Efficient cloning and engineering of entire mitochondrial genomes in Escherichia coli and transfer into transcriptionally active mitochondria. *Nucleic acids research* **31**, 1407-1415(2003).

152. Gaizo, V. Targeting proteins to mitochondria using TAT. *Molecular Genetics and Metabolism* **80**, 170-180(2003).

153. Lithgow, T. Targeting of proteins to mitochondria. *FEBS letters* **476**, 22-26(2000).

154. Del Gaizo, V. & Payne, R.M. A novel TAT-mitochondrial signal sequence fusion protein is processed, stays in mitochondria, and crosses the placenta. *Molecular therapy : the journal of the American Society of Gene Therapy* **7**, 720-730(2003).

155. Ho, A. et al. Synthetic protein transduction domains: enhanced transduction potential in vitro and in vivo. *Cancer research* **61**, 474-477(2001).

156. Dietz, G.P. & Bahr, M. Delivery of bioactive molecules into the cell: the Trojan horse approach. *Molecular and cellular neurosciences* **27**, 85-131(2004).

157. Sanford, J.C. The biolistic process. *Trends in biotechnology* **6**, 299-302(1988).

158. Hafez, I.M., Maurer, N. & Cullis, P.R. On the mechanism whereby cationic lipids promote intracellular delivery of polynucleic acids. *Gene therapy* **8**, 1188-1196(2001).

159. Bergstrand, N. et al. Interactions between pH-sensitive liposomes and model membranes. *Biophysical chemistry* **104**, 361-379(2003).

160. D'Souza, G.G. et al. DQAsome-mediated delivery of plasmid DNA toward mitochondria in living cells. *Journal of controlled release : official journal of the Controlled Release Society* **92**, 189-197(2003).

161. Mireau, H., Arnal, N. & Fox, T.D. Expression of Barstar as a selectable marker in yeast mitochondria. *Molecular genetics and genomics : MGG* **270**, 1-8(2003).

162. Fox, N.B.T.D. In vivo analysis of mutated initiation codons in the mitochondrial COX2 gene of Saccharomyces cerevisiae fused to the reporter gene ARG8 m reveals lack of downstream reinitiation. *Molecular And General Genetics* 1036-1046(2000).

163. Wiesenberger, G., Costanzo, M.C. & Fox, T.D. Analysis of the Saccharomyces cerevisiae mitochondrial COX3 mRNA 5' untranslated leader: translational activation and mRNA processing. *Molecular and cellular biology* **15**, 3291-300(1995).

164. Butow, R.A. & Fox, T.D. Organelle transformation: shoot first, ask questions later. *Trends in biochemical sciences* **15**, 465-468(1990).

165. Fox, T.D., Sanford, J.C. & McMullin, T.W. Plasmids can stably transform yeast mitochondria lacking endogenous mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7288-7292(1988).

166. Cohen, J.S. & Fox, T.D. Expression of green fluorescent protein from a recoded gene inserted into Saccharomyces cerevisiae mitochondrial DNA. *Mitochondrion* **1**, 181-189(2001).

167. Baudin, A. et al. A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Research* **21**, 3329-3330(1993).

168. Ephrussi, B. Action de l'acriflavine sur les levures. I. La mutation "petite colonie."*Ann. Inst. Pasteur* **76**, 351-367(1949).

169. Chen, X. & Clarkwalker, G. The Petite Mutation in Yeasts: 50 Years On. *International Review of Cytology* **194**, 197-238(1999).

170. Fox, T.D. et al. Analysis and Manipulation of Yeast Mitochondrial Genes. *Methods in Enyzmology* **194**, 149-165(1991).

171. Di Talia, S. et al. The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* **448**, 947-951(2007).

172. Mireau, H., Arnal, N. & Fox, T.D. Expression of Barstar as a selectable marker in yeast mitochondria. *Molecular genetics and genomics* **270**, 1-8(2003).

173. Bonnefoy, N., Remacle, C. & Fox, T.D. Genetic Transformation of Saccharomyces cerevisiae and Chlamydomonas reinhardtii Mitochondria. *Methods in Cell Biology* **80**, 525-548(2007).

174. Tingle, M.A., Küenzi, M.T. & Halvorson, H.O. Germination of yeast spores lacking mitochondrial deoxyribonucleic acid. *Journal of bacteriology* **117**, 89-93(1974).

175. Zakharov, I.A. & Yarovoy, B.P. Cytoduction as a new tool in studying the cytoplasmic heredity in yeast. *Molecular and cellular biochemistry* **14**, 15-8(1977).

176. Conde, J. & Fink, G.R. A mutant of Saccharomyces cerevisiae defective for nuclear fusion. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 3651-5(1976).

177. Ochman, H., Gerber, A.S. & Hartl, D.L. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-3(1988).

178. Funes, S. & Herrmann, J.M. Analysis of mitochondrial protein synthesis in yeast. *Methods in molecular biology (Clifton, N.J.)* **372**, 255-263(2007).

179. Pestka, S. Inhibitors of ribosome functions. *Annual review of microbiology* **25**, 487-562(1971).

180. Poutre, C.G. & Fox, T.D. PET111, a Saccharomyces cerevisiae nuclear gene required for translation of the mitochondrial mRNA encoding cytochrome c oxidase subunit II. *Genetics* **115**, 637-647(1987).

181. Fox, C.A. et al. The origin recognition complex has essential functions in transcriptional silencing and chromosomal replication. *Genes & development* **9**, 911-24(1995).

182. Walter, J.M. Effect of proteorhodopsin on the physiology of heterologous organisms. 131(2008).

183. Tong, A.H.Y. & Boone, C. High-Throughput Strain Construction and Systematic Synthetic Lethal Screening in Saccharomyces cerevisiae. *Methods in Microbiology* **36**, 1-19(2007).

184. Schäfer, B. RNA maturation in mitochondria of S. cerevisiae and S. pombe. *Gene* **354**, 80-85(2005).

185. Fox, T.D. & Bonnefoy, N. In vivo analysis of mutated initiation codons in the mitochondrial COX2 gene of Saccharomyces cerevisiae fused to the reporter gene ARG8 m reveals lack of downstream reinitiation. *Molecular And General Genetics* 1036-1046(2000).

186. Sawitzke, J.A. et al. Recombineering: In Vivo Genetic Engineering in E. coli, S. enterica, and Beyond. *Methods in Enyzmology* **421**, 171-199(2007).

187. Gietz, D. et al. Improved method for high efficiency transformation of intact yeast cells. *Nucleic acids research* **20**, 1425(1992).

188. Gietz, R.D. & Woods, R.A. Yeast transformation by the LiAc/SS Carrier DNA/PEG method. *Methods in molecular biology (Clifton, N.J.)* **313**, 107-20(2006).

189. Goldstein, A.L. & McCusker, J.H. Three new dominant drug resistance cassettes for gene disruption inSaccharomyces cerevisiae. *Yeast* **15**, 1541-1553(1999).

190. Brewster, D. & Westfall, R.S. *Memoirs of the Life, Writings, and Discoveries of Sir Isaac Newton: Volume 2*. 584(Adamant Media Corporation: Chestnut Hill, MA, 1965).

Appendix A: SEQUENCE OF *RPO41^{M}* SYNTHETIC CONSTRUCT

```
    1 EcoRI
    |
  1 gaattcataaattttaattaaaagtagtattaacatattataaatagaca
                >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                COX2 5' UTR
                   >>>>>>>>>
                   cox2 promoter

 51 aaagagtctaaaggttaagatttattaaaatgttaagaccagcttataaa
                                  >>>>>>>>>>>>>>>>>>>>
                                  RPO41m
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    COX2 5' UTR

101 tcattagtaaaaacatctttattacaaagaagattaatttcatcaaaagg
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

151 ttcaaaattatttaaaccttcacctgattcaacatcaacaattttaattt
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

201 ctgaagatcctttagtaacaggttcatcacctacatcatctacaacatca
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

251 ggtattatttcatcaaatgattttccattatttaataaaaatagaaaaga
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

301 tgctaaatcttcaatttcttatcaatggaaaaatccttcagaattagaat
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

351 ttgatccttttaataaatctcatgcttcagctgtaacatcaataacaaga
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

401 acaagagatgtaatacaattatgatctttattagaagcttgtttacaatc
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m
```

```
451 aaatttaataaaaagagctttttcaattttagaatcattatatttagtac
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

501 ctgaacataaacaaagatttattgaagattataatatgtatttaaattca
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

551 ttttcaaaaaatgatcctaattttcctattttaaaaatgaatgaaaaatt
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

601 aacaaatgatttagaaacatcatttaaagatgtaaattataatgataaaa
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

651 cattagcaattatgattcatcatgctttaaattttcattcaacaacatca
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

701 tcaatattattaaaacctattatttctgcttatttaaaaatatcagtaaa
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

751 tggtattagagaaatttttttcttgtttagatattttaacaatttcagatt
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

801 taaatattttaataaatgatttaaaagtaattacacctagtcaattacct
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m
                        >>>>>>>>>
                        internal cox2 promoter

851 aattcagtaagaccaattttagaatcattaactttatctcctacacctgt
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

901 aaataatattgaaaatgaagaaggtttaaataaagtagaagctgaaaatg
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m

951 attcaaaattacataaagcttcaaatgctagttcagattcaattaaaaaa
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    RPO41m
```

```
1001 ccttcattagatcctttaagagaagtatcatttcatggttcaacagaagt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1051 attatcaaaagatgctgaaaaattaattgctgtagatacaattggaatga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1101 gagtaattagacatacattattaggtttatcattaacaccagaacaaaaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1151 gaacaaatttcaaaatttaaatttgatgcaaatgataatgtattaaagat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1201 gaaaccaacaaaaaatgatgataataataattcaattaatttttttgaaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1251 tttataattcattaccaacattagaagaaaaaaaagcatttgaatcagct
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1301 ttaaatatttttaatcaagatagacaaaaagtattagaaaatagagctac
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1351 tgaagctgcaagagaaagatgaaaacatgattttgaagaagcaaaagcta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1401 gaggtgatatttcaattgaaaaaaatttaaatgtaaaattatgaaaatga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1451 tataatgaaatgttaccattagtaaaagaagaaattaatcattgtagatc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1501 attattatcagaaaaattatcagataaaaaaggtttaaataaagtagata
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

1551 caaatagattaggttatggtccttatttaacattaattgatcctggtaaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

RPO41m

1601 atatgtgtaattacaattttagaattattaaaattaaattctactggtgg
>>>
RPO41m

1651 tgtaattgaaggtataagaactgctagagctgtaatttcagttggtaaag
>>
RPO41m

1701 ctattgaaatggaatttagatcagaacaagtattaaaatctgaatcacaa
>>
RPO41m

1751 gctttcagagatgtaaataaaaaatcacctgaatttaaaaaattagtaca
>>
RPO41m

1801 aaatgctaaatcagtatttagatcatcacaaattgaacaatctaaaattt
>>
RPO41m

1851 tatgacctcaatctattagagctagaattggttcagtattaatttcaata
>>
RPO41m

1901 ttaattcaagttgctaaagtatctgtacaaggtgtagatcctgtaacaaa
>>
RPO41m

1951 agctaaagtacatggtgaagctcctgcttttgctcatggttatcaatatc
>>
RPO41m

2001 ataatggttcaaaattaggtgtattaaaaattcataaaacattaattaga
>>
RPO41m

2051 caattaaatggtgaaagattaattgcttctgtacaacctcaattattacc
>>
RPO41m

2101 tatattagtagaacctaaaccatgagtaaattgaagatcaggtggttatc
>>
RPO41m

```
2151 attatacacaatcaacattattaagaacaaaagattcaccagaacaagta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2201 gcttatttaaaagctgcttctgataatggtgatattgatagagtttatga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2251 tggtttaaatgtattaggtagaacaccttgaacagttaatagaaaagtat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2301 ttgatgtagtatcacaagtatgaaataaaggtgaaggtttttagatatt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2351 cctggtgcacaagatgaaatggttttaccacctgctcctccaaaaaattc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2401 tgatccttcaattttaagagcttgaaaattacaagtaaaaacaattgcta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2451 ataaattttcttctgatagatcaaatagatgtgatacaaattataaatta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2501 gaaattgcaagagctttttaggtgaaaaattatattttcctcataattt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2551 agattttagaggtagagcttatcctttatctcctcatttaatcatttag
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2601 gtaatgatatatcaagaggtttattaatttttgacatggtaagaaatta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2651 ggtccttcaggtttaaaatgattaaaaattcatttatcaaatttatttgg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2701 ttttgataaattacctttaaaagatagagtagcttttacagaatctcatt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
     RPO41m

2751 tacaagatattaaagattcagcagaaaatcctttaacaggtgatagatga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2801 tgaacaacagctgataaaccttgacaagctttagctacttgttttgaatt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2851 aaatgaagttataaaaatagataatcctgaagaatttatttcacatcaac
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2901 ctgttcatcaagatggtacttgtaatggtttacaacattatgctgcttta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

2951 ggtggtgatgttgaaggtgctacacaagtaaatttagtaccatcagataa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3001 accacaagatgtttatgctcatgtagctagattagtacaaaaaagattag
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3051 aaattgctgcagaaaaaggagatgaaaatgctaaaattttaaaagataaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3101 attacaagaaaagtagttaaacaaacagtaatgacaaatgtatatggtgt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3151 tacttatgtaggtgctacatttcaaattgctaaacaattatctcctattt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3201 ttgatgatagaaaagaatcattagatttttctaaatatttaacaaagcat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3251 gtattttcagctattagagaattatttcattctgcacatttaattcaaga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m
```

```
3301 ttgattaggtgaatctgcaaaaagaattagtaaatctattagattagatg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3351 tagatgaaaaatctttcaaaaatggtaataagcctgattttatgtcatca
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3401 gtaatttgaacaacaccattaggtttacctattgtacaaccttatagaga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3451 agaatcaaaaaaacaagtagaaacaaatttacaaactgtatttatttctg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3501 atcctttcgctgtaaatcctgtaaatgcaagaagacaaaaagctggttta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3551 ccacctaattttattcattctttagatgcttcacatatattattatcagc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3601 tgctgaatgtggtaaacaaggattagattttgcttctgttcatgattctt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3651 attgaactcatgcatctgatattgatactataaatgtagtattaagagaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3701 caatttattaaattacatgaagtagatttagtattaagattaaaggaaga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3751 atttgaccaaagatataaaaattatgtaaaaattggtaaattaaaaagat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3801 caacagatttagctcaaaaaattattagaattagaaaagatttatcaaga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     RPO41m

3851 aaattaggtagaagtacaacattagctgatgaaatttattttgaaaaaaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
      RPO41m

  3901 aagacaagaattattaaattcacctttaattgaagatagaaatgttggtg
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      RPO41m

  3951 aaaaaatagttacaacagtatcattatttgaagatattacagatttagat
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      RPO41m

  4001 gcattagaattagaaaatggtggagatgaaaattctggtataagtgtttt
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      RPO41m

  4051 attacctttaagattacctgaaattccacctaaaggtgattttgatgtta
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      RPO41m

  4101 cagtattaagaaattcacaatatttttttcataattaatatttacttat
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      RPO41m
                                          >>>>>>>>>>>>>>>>
                                          COX2 3' UTR

  4151 tattaatatttttaattattaaaaataataataataataattataat
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      COX2 3' UTR
                                                        >>>
                                                        mito
transcriptional terminator

  4201 aatattcttaaatataataaagatatagatttatattctattcaatcacc
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      COX2 3' UTR
      >>>>>>>>>>
      mito transcriptional terminator

          4255 EcoRI
          |
  4251 ttatgaattc
      >>>>
      COX2 3' UTR
```