AN IN VITRO BIOCHEMICAL STUDY OF THE DROSOPHILA NEGATIVE ELONGATION FACTOR

A Dissertation in
Biochemistry, Microbiology and Molecular Biology

by

Michael J. Fisher

©2017 Michael J. Fisher

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

May 2017
The dissertation of Michael James Fisher was reviewed and approved* by the following:

David S. Gilmour  
Professor of Biochemistry and Molecular Biology  
Graduate Education Co-director  
Dissertation Advisor  
Chair of Committee

Paul Babitzke  
Professor of Biochemistry and Molecular Biology  
Graduate Education Co-director

Katsuhiko Murakami  
Professor of Biochemistry and Molecular Biology

Michael Axtell  
Professor of Biology

*Signatures are on file in the Graduate School
ABSTRACT

The concentration of transcriptionally engaged RNA polymerase II (Pol II) at the 5’ ends of genes, a phenomenon called promoter-proximal pausing, is a common mode of gene regulation in metazoans. This promoter-proximal pausing is mediated together by protein complexes 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity inducing factor (DSIF) and Negative Elongation Factor (NELF). Despite many studies of NELF, the manner in which it imparts pausing activity remains unclear. One feature of interest is the RNA recognition motif (RRM) of the subunit NELF-E, which has previously been shown as necessary for pausing, and its ability to bind to RNA. This binding activity has been implicated in the modulation of pausing activity and the dissociation of NELF from paused polymerase, allowing transcription into the body of the gene. I prepared a series of NELF-expressing baculoviruses which were used to produce NELF and a series of mutations in NELF-E. Using these protein complexes, I studied the effects of NELF-E on pausing in an in vitro transcription system and identified regions required for association of NELF-E with the other three subunits, binding to the Pol II/DSIF elongation complex (EC), and for pausing activity. I discovered that a region previously identified as a leucine-zipper-like domain near the N-terminus of NELF-E is required for association of NELF-E with other NELF subunits. The absence of NELF-E from the complex results in complete loss of pausing activity. Contrary to expectations, I discovered that the RRM is not essential for pausing since deletion of this domain did not affect the pausing activity of NELF in vitro. Instead, I found that amino acids within residues 40 through 113 in NELF-E are required for pausing activity and EC binding, but not the RRM or flanking sequences. This provides incentives for studies on NELF-E to examine residues in the N-terminal half of the protein and discover binding partners for this region which hopefully will explain how NELF promotes polymerase pausing.
TABLE OF CONTENTS

List of Figures .................................................................................................................. vi
List of Tables ...................................................................................................................... viii
Acknowledgements ......................................................................................................... ix

Chapter 1 Introduction ..................................................................................................... 1

Promoter-proximal pausing .............................................................................................. 1
Discovery of the negative elongation factor, NELF ......................................................... 4
Known composition of NELF .......................................................................................... 4
NELF evolutionary lineages ............................................................................................. 9
The roles of NELF-mediated pausing in gene regulation .................................................. 15
The function of NELF in development ............................................................................ 16
NELF-mediated pausing in viruses ................................................................................. 17
Activity by NELF apart from pausing ............................................................................ 18
Medical relevance of NELF ............................................................................................ 19
The NELF-E RRM .......................................................................................................... 21
Scope and significance of dissertation .......................................................................... 22

Chapter 2 Materials and Methods .................................................................................. 23

Sf9 cell general maintenance ......................................................................................... 23
Sf9 cell cryopreservation, recovery, and adaptation ......................................................... 25
Baculovirus amplification and maintenance techniques .................................................. 26
Baculovirus plaque assays ............................................................................................... 27
Design of NELF-E mutant proteins for expression in E. coli ........................................ 32
Expression of NELF-E and NELF-E mutant proteins .................................................... 32
Expression and purification of DSIF .............................................................................. 36
Production of NELF viruses ............................................................................................ 40
Generating a map of the recombined plasmid using restriction enzymes ....................... 48
Expression and purification of NELF and NELF-E mutant complexes ......................... 52
Expression and purification of P-TEFb ......................................................................... 56
P-TEFb kinase reactions ................................................................................................. 56
Promoter-proximal pausing assay using Drosophila nuclear extracts ........................... 58
EMSAs on reconstituted elongation complexes containing NELF and DSIF ................ 63
NELF-E RNA binding EMSAs ...................................................................................... 64
Quantification of bands for EMSAs and sequencing gels ............................................. 65

Chapter 3 Biochemical investigation of NELF binding and activity ................................ 68

NELF-E mutant design strategy ...................................................................................... 68
The RRM of NELF-E is required for binding to a specific sequence of RNA ............... 76
Identification of NELF-E domains required for integration into NELF complexes ...... 80
NELF-E is required, but the RRM is dispensable for pausing in vitro ......................... 82
NELF-E has RNA binding activity outside of the RRM ................................................. 85
A previously undefined region of NELF-E is required for pausing ............................ 89
NELF-E is required for binding to the elongation complex and the RRM is dispensable for this binding ................................................................. 92
Mutating the RRM results in more runoff products in G-less cassette assay .................. 95
RRM mutants retain pausing activity in G-less cassette assay .................................. 98
P-TEFb kinase reactions of DSIF and NELF-E proteins .............................................. 102

Chapter 4 Discussion .................................................................................................. 104
NELF-E is required for pausing but the RRM of NELF-E is not ................................. 104
RNA binding specificity of the RRM is not required for pausing .............................. 105
Implications for structural dependency on NELF-E .................................................. 105
Human NELF-E phosphorylation sites are only conserved in mammals .................. 109
Drosophila P-TEFb does not phosphorylate Drosophila NELF-E ............................... 111
Implications of the role of NELF in metazoans ......................................................... 111
Closing remarks ........................................................................................................ 112

Appendix A Protein - Nucleic Acid Crosslinking Assays ........................................... 114
Introduction ................................................................................................................ 114
Protein to DNA crosslinking experiments ................................................................. 114
Protein to RNA crosslinking experiments ................................................................. 122
Discussion .................................................................................................................. 122

Appendix B Addback of NELF-E into Incomplete NELF complexes ..................... 125
Introduction ................................................................................................................ 125
Production of Mini-NELF-A ....................................................................................... 125
Expression and purification of E-less NELF .............................................................. 128
Discussion .................................................................................................................. 128

Appendix C Testing of NELF construct affinity tags .............................................. 133
Introduction ................................................................................................................ 133
NELF affinity tags ..................................................................................................... 133
Discussion .................................................................................................................. 139

References .................................................................................................................. 141
LIST OF FIGURES

Figure 1-1: Primary structure comparison between human and *Drosophila* NELF subunits...6
Figure 1-2: Protein-protein crosslinking results and model of putative NELF architecture....8
Figure 1-3: A simple cladogram of eukaryote lineages indicating organisms where NELF is present .................................................................12
Figure 1-4: Sequence comparison for NELF-D from eukaryote lineages .........................14
Figure 2-1: Baculovirus plaque assay ........................................................................31
Figure 2-2: SDS-PAGE Coomassie stained gels of recombinant *Drosophila* NELF-E and mutants from *E. coli* ..................................................................................35
Figure 2-3: Partial purification of DSIF from *E. coli* ..................................................39
Figure 2-4: Test expression of NELF using baculovirus .............................................47
Figure 2-5: Plasmid map of pACEBac1_NELF from colony 8 as determined by restriction digest .......................................................................................51
Figure 2-6: Purification of wild-type NELF ................................................................55
Figure 2-7: Coomassie blue stained SDS-PAGE gel of P-TEFb preparation ...............57
Figure 2-8: Image quantification guide ......................................................................67
Figure 3-1: Structural diagrams of NELF-E .................................................................71
Figure 3-2: Secondary structure predictions on NELF-E using Jpred4 .......................75
Figure 3-3: NELF-E RNA binding EMSA using fluorescent-labeled Napt1min RNA ....78
Figure 3-4: The NELF-E RRM is required for RNA binding specificity ....................79
Figure 3-5: Partially purified NELF complexes containing NELF-E mutants .............81
Figure 3-6: The NELF-E RRM is dispensable for pausing *in vitro* ...............................84
Figure 3-7: NELF-E has RNA binding activity outside of the RRM ............................88
Figure 3-8: NELF-E residues within amino acids 40 through 113 are required for pausing...91
Figure 3-9: NELF-E is required for binding to the elongation complex, but the RRM and residues 114-167 or 247-277 are not .........................................................94
Figure 3-10: Mutations in the RRM allow for more runoff .......................................97
Figure 3-11: Capture of RNase T1-cut transcripts from a G-less cassette .....................101
Figure 3-12: Human NELF-E, but not *Drosophila* NELF-E is phosphorylated by P-TEFb ... 103

Figure 4-1: An alternative structural model of NELF................................................108

Figure 4-2: An examination of NELF-E primary structure .......................................110

Figure A-1: Elongation complex binding EMSA using crosslinkable templates ..........119

Figure A-2: Protein-DNA crosslinking experiments on elongation complexes ..........121

Figure A-3: Figure A-3: Protein-RNA crosslinking assay........................................124

Figure B-1: Sequence comparison for Mini-NELF-A .............................................127

Figure B-2: Expression level comparison of E-less NELF constructs and purified complex integration test.........................................................................................130

Figure B-3: NELF-E may be successfully integrated into E-less NELF complexes containing Mini-NELF-A .................................................................132

Figure C-1: NELF-D*FLAG* expresses at very low levels........................................136

Figure C-2: NELF-A-Str constructs cannot be purified using the strep-tag ...............137

Figure C-3: N-terminal strep tag on NELF-A causes expression problems ..........138

Figure C-4: Utilizing the FLAG tag on the C-terminus of NELF-B causes NELF-A to dissociate from the complex .................................................................140
LIST OF TABLES

Table 2-1: Primers used for InFusion cloning of pIDK_NELF-E scanning mutants..................46
Table 2-2: Fragment sizes from BamH I digests of pACEBac1_NELF constructs ..................50
Table 2-3: Example of single sample contents before mixing with nuclear extract ...............60
Table A-1: DNA oligonucleotides for synthesizing crosslinkable templates .....................116
Table C-1: NELF constructs tested with brief description of results.................................135
ACKNOWLEDGEMENTS

This work was made possible with the assistance of my colleagues, advisors, friends and family, all who have given me their support and guidance along the course of this project, this chapter in my life where I have developed both as a scientist and as a human.

I would like to thank my advisor, David Gilmour, for giving me this opportunity to improve myself, not only for providing a place of learning and invaluable assistance, but also for his patience with me during my tenure in his lab. I had constant personal struggles between the end of 2011 and the end of 2014 which had a substantial negative impact on my project but despite that, Dave was willing to give me additional chances to succeed, even when he had to take pains to do so. His efforts were a large contributing factor in helping me get my problems under control. Working in his lab was an irreplaceable experience in my life.

I would like to thank the other members of my committee, Joe Reese, Katsu Murakami, Paul Babitzke, and Mike Axtell for their patience as well. The opportunity afforded by that patience and their firm guidance allowed me to develop my project to its current extent from having almost no results in February 2015 to having enough data to publish by mid-August 2016. I especially appreciate their support and critical feedback in writing this dissertation. Writing does not come naturally to me, but I feel that they taught me a lot about how to be clear and concise.

My time in the lab was not always easy, but the technical and moral support from my fellow lab members proved indispensable in helping me through my difficulties. For this, I would like to thank Bede Portz, Doug Baumann, Yijun Qiu, Saikat Ghosh, Jian Li,
Feiyue (Jacob) Lu, Greg Kothe, Bhavana Achary, Anamika Missra, and Laura Stefanik. I give special thanks to Bede Portz for his concern, advice and support as I grappled with both my scientific problems and my personal ones. I also give special thanks to Yijun: She took a lot of time to repeat some of my experiments when I wasn’t able to come to the lab. She was also the only other person in the lab who knew how to do several of the assays I used in my work after Anamika and Jian left and I would ask her for help when one of the assays would inexplicably fail. I would like to thank the faculty and students in the Center for Eukaryotic Gene Regulation for their advice and permission to use some of their equipment and I would like to thank all of the undergraduates and technicians who worked in our lab over the years as they often made our lab work much easier. I give special thanks to Spencer Hauck, the undergraduate I helped train, as he provided much assistance in the process of setting up and getting the baculovirus expression system to work at every step.

Though my family didn’t always clearly understand my work – there are no other scientists in my family – they were always willing to listen to what I have been working on and offered any support they could. It was from their efforts throughout my life that I had the opportunity to be the first in my family to receive a post-graduate education and a chance at getting a doctorate. For their efforts, I dedicate my dissertation to them.
Chapter 1

Introduction

Promoter-proximal pausing

Promoter-proximal pausing, first characterized on the human \textit{myc} gene [Bentley and Groudine, 1986] and \textit{hsp70} in \textit{Drosophila} [Gilmour and Lis, 1986] is a phenomenon where RNA polymerase II is concentrated at the 5’ end of many genes, even those that are not being actively transcribed [Muse et al, 2007; Geunther et al, 2007; Zeitlinger et al, 2007], leading to the generally accepted model that the rate limiting step in gene expression – determining whether the gene is transcribed – is in many cases not due to pre-initiation complex assembly, but rather the release of these paused polymerases such that elongation can resume [Gilmour, 2009; Fuda et al, 2009].

To learn how pausing works, it must first be understood that RNA polymerase II activity, which its role in the nucleus is to transcribe genes, is not a simple process. Transcriptionally engaged RNA polymerase II mobilizes on the DNA template by Brownian motion causing the polymerase to drift on the template to find a thermodynamically favorable state [Nudler, 2009]. As a result of this mechanism, polymerase can be driven forward by the thermodynamically favorable incorporation of NTPs into the elongating transcript, but may otherwise be able to backtrack at least one nucleotide [Shaevitz et al, 2003]. The misalignment of the 3’ end of the transcript halts polymerization and initiates a
pause. Due to Brownian motion, the polymerase may be able to spontaneously realign the 3’ end of the transcript with the active site and resume transcription [Selth et al, 2010]. Under some instances, transcriptionally engaged polymerase cannot spontaneously realign the 3’ end of the RNA with the active site. This is termed “transcriptional arrest” and requires cleavage of the RNA either intrinsically or with the assistance of TFIIS to resume transcription [Izban and Luse, 1992].

A number of things can influence the equilibrium between active elongation and pausing of RNA polymerase in general, including intrinsic inhibition of elongation by the sequence of the DNA [Bai et al, 2004] and/or the nascent transcript [Wang et al, 1997], DNA binding proteins such as histones [Izban and Luse, 1991] or nucleoside triphosphate (NTP) limitation. For example, a strongly positioned +1 nucleosome strengthens NELF-mediated pausing [Jimeno-González et al, 2015]. There are a variety of proteins known to either enhance or alleviate pausing as well. Those known to alleviate pausing include TFIIS, TFIIF, ELL, and Elongin [Uptain et al 1997]. Factors known to bind to the elongation complex and enhance pausing include DSIF [Wada et al, 1998], a heterodimer of Spt4 and Spt5 conserved in all eukaryotes with homologues in Archaea known as RpoE and NusG, respectively [Ponting, 2002], and NELF (negative elongation factor) [Yamaguchi et al, 1999], a four subunit complex conserved in most higher eukaryotes.

The Spt5 subunit of DSIF consists of a NusG N-terminal domain (NGN) which binds to Spt4 and a series of KOW domains, the number of which vary by
organism [Wenzel et al, 2009; Meyer et al, 2015]. In vitro transcription studies in *Bacillus subtilis* show that NusG is an amplifier of sequence specific pauses [Yakhnin and Babitzke 2010]. Footprinting data shows that NusG protects nucleotides in the non-template strand of the DNA [Yakhnin et al, 2008], in particular, the NusG NGN crosslinks to the non-template DNA strand at the U144 pause site of the *B. subtilis trp* leader [Yakhnin et al, 2016]. Crosslinking of DSIF to the non-template strand of DNA was also confirmed in yeast [Crickard et al, 2016]. Furthermore, the NGN of Spt5 binds to and bridges subunits Rpb1 and Rpb2 of RNA polymerase near the upstream DNA and interacts with the non-template strand [Klein et al, 2010; Bernecky et al, 2016]. Additionally, protein-RNA crosslinking experiments show that *Drosophila* Spt5 interacts with a nascent transcript when it is greater than 18 nucleotides long [Missra and Gilmour, 2010]. This evidence together suggests that DSIF acts through nucleic acid binding, and by doing so, possibly reducing the favorability of forward translocation of the polymerase. Yet, this alone does not confer strong pausing in higher eukaryotes. Both DSIF and NELF are required for pausing in vivo [Wu et al, 2003] and in vitro [Yamaguchi et al, 1999; Yamaguchi et al, 2002]. Paused polymerase bound by DSIF and NELF can be allowed to resume productive elongation through phosphorylation of serine 2 of the carboxy-terminal domain Pol II subunit Rpb1 [Marshall et al, 1996], and the carboxy-terminal region of DSIF subunit Spt5, which converts DSIF from a negative elongation factor to a positive elongation factor [Yamada et al, 2006] and releases NELF from the elongation complex [Wu et al, 2003; Yamaguchi et al, 1999].
**Discovery of the negative elongation factor, NELF**

NELF was first identified following the discovery of DSIF and P-TEFb (positive transcription elongation factor b) in an effort to isolate novel factors that sustain promoter proximal pausing. The initial observation was that there were two classes of elongating RNA polymerase: Polymerase that was located near the transcription start site that would not transcribe further (paused) and polymerase that was able to actively transcribe long distances [Marshall and Price, 1992]. It was found that complexes that could transcribe long distances were able to do so through the action of a DRB-sensitive positive transcription elongation factor (P-TEF), and that the paused complexes could be stimulated to productive elongation by the addition of high salt or heparin. This led to the hypothesis that a negative elongation factor exists that inhibits paused polymerase from entering productive elongation, and that high salt or heparin caused the inhibitor to dissociate [Marshall and Price, 1992]. This negative elongation factor, along with DSIF, was eventually found by screening fractionated nuclear extract for factors which induced DRB sensitivity in elongation complexes [Wada et al, 1998; Yamaguchi et al, 1999].

**Known composition of NELF**

NELF is a four subunit complex consisting of proteins named NELF-A, NELF-B, NELF-C or NELF-D, and NELF-E and is known to directly interact with Pol II and DSIF [Yamaguchi et al, 1999]. NELF-C and NELF-D are essentially the same protein except that NELF-C is 9 amino acids longer which is thought to be the result of an alternate translation start site on the mRNA in humans [Narita...
et al, 2003]. The subunits interact with each other in a quasi-linear manner where NELF-A binds to NELF-D, which binds to NELF-B, which binds to NELF-E. All of these subunits are thought to be required both for binding to a DSIF-containing elongation complex and for pausing activity [Narita et al, 2003].

Only limited structural data exists for NELF. In humans, NELF-A has a region with weak homology to hepatitis delta antigen (HDAg) between residues 89 and 248 that is important for binding to elongation complexes. NELF-A also contains a region critical for binding to NELF-D between residues 125 and 188 [Narita et al, 2003]. NELF-E contains a leucine-zipper motif near the N-terminus, an RRM-type RNA recognition motif, and in some vertebrates, an arginine/aspartate rich domain called RD (Figure 1-1)[Lyamouri et al, 2002; Fujinaga et al, 2004; Yamaguchi et al, 2002]. NELF-B, in humans, is leucine rich and contains three repeats of the LXXLL motif [Ye et al, 2001], which is present in many transcription coactivators and mediates binding to nuclear receptors [Heery et al, 1997]. Recently, a crystal structure of a partially reconstituted NELF complex, consisting of the N-terminal half of NELF-A and the C-terminal two thirds of NELF-D, has been solved. This structure shows that NELF-A contains an ordered helical N-terminal domain followed by an extension, along which interacts with NELF-D. NELF-D forms an unusual horseshoe-like structure consisting of a series of alpha-helices, among which are three HEAT repeat motifs [Vos et al, 2016]. HEAT repeat motifs are structural motifs known to function in protein-protein interactions [Andrade et al, 2001] with a largely
Figure 1-1 Primary structure comparison between human and *Drosophila* NELF subunits. Both NELF-B and NELF-D are largely conserved between humans and *Drosophila*. Known regions in NELF shared by both humans and *Drosophila* include the NELF-A hepatitis delta antigen (HDAg) homologous domain (red box), the leucine zipper-like motif in NELF-E (blue box), and the RD domain in human NELF-E (green box). *Drosophila* has a roughly 70kDa insertion in NELF-A and lacks the RD domain in NELF-E. Figure adapted from a previous study [Wu et al, 2005].
degenerate sequence, but has highly conserved proline, aspartate, and arginine residues at fixed positions [Groves et al, 1999].

Protein-protein crosslinking experiments on NELF, described in Vos et al, 2016, show widespread inter and intramolecular interactions within the NELF complex, but uncovered several notable interactions. NELF-B, NELF-E, and NELF-D were found to extensively crosslink with the HDAg-like domain in NELF-A. This region was previously determined to be important for binding to the elongation complex and for NELF-D binding, residues 125-188 in particular [Narita et al, 2003]. However, this region is known to be highly susceptible to proteolysis so much of this crosslinking activity was attributed to the natural flexibility of this region [Vos et al, 2016].

There was extensive reactivity between a roughly 30aa size region near the N-terminus of NELF-E and multiple targets in NELF-B within the N-terminal, middle, and C-terminal regions of the protein, and a small region near the C-terminus of NELF-D. This observation supports an earlier model that NELF-E predominantly binds to NELF-B [Narita et al, 2003], but also indicates that NELF-E is interacting with the remaining two subunits as well (Figure 1-2D-F). This led to a rough model where The N-terminal part of NELF-E is sandwiched between the N and C-terminal parts of NELF-B, with NELF-D stuck to the N-terminal half of NELF-B. Thus, the RRM of NELF-E is bound to the rest of the complex by a flexible linker that is made longer by the presence of RD, an arginine/aspartate repeat domain, in mammals and reptiles (figure 1-2G).
Figure 1-2: Protein-protein crosslinking results and model of putative NELF architecture. Crosslinks are shown as binary intermolecular (black, straight lines) or intramolecular (colored, curved lines) interactions within the NELF tetramer (A-F). This data was used to generate a simple model of the NELF quaternary structure (G). NELF-E is bound between the N- and C-terminal regions of NELF-B while NELF-D is bound to the N-terminal domain of NELF-B nearby. Figure adapted from a previous study [Vos et al, 2016].
NELF evolutionary lineages

To draw conclusions about the role of NELF in biology, it is necessary to examine NELF in all lineages that have it, not just the few model organisms that have been studied to date. NELF subunits are conserved throughout metazoans with the one exception being the Nematode lineage. Within Nematoda, a few members have either degenerate forms of NELF, a few notable examples being Brugia malayi and Pristionchus pacificus (Figure 1-4, compare B. malayi conserved residues), or are missing NELF altogether like Caenorhabditis elegans. It has been recently noted that NELF may in fact be an ancient protein complex that evolved in early protozoans [Vos et al, 2016]. Lineages that have organisms which express NELF, in addition to metazoans, include protozoa, fungi, and plants as it is present in Mucor ambiguus (fungus mold) (Figure 1-3), Dictyostelium discoideum (slime molds) [Chang et al, 2012], and Chlorella varabilis (green algae) [Vos et al, 2016]. Since NELF has existed since the diversification of early protozoa, learning how ubiquitous NELF containing organisms are is best done by studying those early eukaryotic clades. Protozoan lineages have remained difficult to sort into monophyletic taxa. In this analysis and for the sake of simplicity, Protozoan lineages were sorted as described [Cavalier-Smith, 2010]. A simplified diagram of these lineages is shown in figure 1-3. NELF subunits are found in several isolated opisthokont lineages including examples in the Amoebozoa lineage (e.g. the slime molds Dictyostelium sp.), and can be detected in some Fungi (e.g. Mucor ambiguus). Furthermore, a few examples of NELF can be found in isolated bikont lineages (e.g. Chlorella
variabilis, Phytophthora infestans), which suggests that the first instances of NELF appeared in early Protozoan lineages before the diversification of unikonts and bikonts (Figure 1-3). Sequence alignments of NELF-D from each of these organisms are shown in figure 1-4. It is important to note, however, that the majority of sister taxa of these examples have no detectible NELF subunits. That being said, genomic and proteomic databases are still being improved at a rapid rate. To produce a more precise estimate of when the first instances of NELF came into existence, it will be important to look for examples of NELF subunits in protozoans that are neither unikonts nor bikonts.

Another question to address is whether or not NELF mediated pausing exists in these other lineages. NELF has remained in these isolated groups over hundreds of millions of years. This suggests selective pressures maintained its presence, implying that NELF is functional in these organisms. Indeed, a recent study speculated that there would be pausing in Dictyostelium – a nucleosome resides adjacent to where Pol II would be predicted to pause. However, no mapping data for Pol II was provided so it remains to be determined if promoter proximal pausing occurs.

This general analysis of the lineage of NELF brings us to a few notable observations and conclusions. Though some species do not have NELF, their ancestors did (S. cerevisiae and C. elegans being very commonly studied examples). Organisms that do have NELF have fairly well conserved NELF proteins. Degenerate NELF genes that do not show typical patterns of conservation are rare. The presence of NELF is greatly enriched within Metazoa.
Instances of NELF in lineages outside Metazoa exist only in smaller, isolated taxa. Finally, in order to determine if NELF-mediated pausing exists outside of Metazoa, then there need to be studies that have Pol II mapping data in organisms that have NELF.
Figure 1-3: A simple cladogram of eukaryote lineages indicating organisms where NELF is present. Lineages are marked in non-italic bold text. Examples of organisms known to contain NELF subunits are named at the ends of the lineages in italic text (most metazoans except *C. elegans* have NELF). Based on the current understanding of eukaryotic phylogeny, it can be deduced that NELF first arose in organisms before the divergence of bikonts and unikonts (indicated with an asterisk). The possibility remains that NELF evolved at earlier lineages, which could be discovered with improving proteomics databases for basal eukaryotes.
Figure 1-4: Sequence comparison for NELF-D from eukaryote lineages. The primary amino acid sequences of NELF-D from *Phytophthora infestans* (fungi), *Mucor ambiguus* (fungi), *Brugia malayi* (nematode), *Chlorella variabilis* (green algae), *Dictyostelium discoideum* (slime mold), *Drosophila melanogaster* (fly), and *Homo sapiens* (mammal) are shown in alignment best fit using the Multiple Sequence Comparison by Log-Expectation algorithm (MUSCLE) [Edgar, 2004]. Identical residues are highlighted in black. NELF-D was used for this analysis because it was the easiest subunit to identify using proteomics and basic local alignment search tools. The NELF-D primary sequence is consistently conserved in organisms that have it, despite some organisms being separated by over 1.3 billion years on the evolutionary time scale.
The roles of NELF-mediated pausing in gene regulation

Permanganate footprinting, PRO-seq, and ChIP-seq assay genome-wide analysis show that polymerase is transcriptionally engaged and stably associated between 20 and 60 bases downstream of the transcription start site on thousands of genes in *Drosophila* [Kwak et al, 2013; Nechaev et al, 2010; Li et al, 2013]. Similar pausing distribution was also found in mammalian cells [Min et al, 2011]. Depletion of NELF in *Drosophila* cells resulted in a decrease in paused Pol II [Muse et al, 2007], implying that this concentration of polymerase at the 5’ end of these genes is the result of NELF-mediated pausing.

NELF-mediated promoter-proximal pausing in metazoans fulfils three main functions in gene regulation: Deactivation, synchronization, and rapid induction. Deactivation is caused by the suppression of Pol II elongation near the promoter by binding of DSIF and NELF to the elongation complex. This has been shown to be important for shutting off the hsp70 gene in *Drosophila* through pausing-mediated direction of CBP, a factor which acetylates heat shock factor (HSF), inhibiting its DNA binding activity [Ghosh et al, 2011]. NELF is also known to inhibit TFIIS [Palangat et al, 2005], which could cause arrested polymerase to dwell for longer periods of time.

Almost paradoxically, NELF-mediated pausing is necessary for the rapid activation of some genes. Limiting the activation of genes to the release of a single factor, NELF permits faster gene activation than recruiting general transcription factors to the promoter and initiating transcription. Indeed, NELF mediated pausing allows rapid induction of inflammatory response genes.
Adelman et al, 2009], innate immune response genes [Xu et al, 2012] and is required for stimulus specific rapid induction of some immediate early genes (IEGs) including arc [Saha et al, 2011], junB [Aida et al, 2006] and c-fos [Fujita et al, 2009]. In neurons, stimulated IEGs under the control of NELF are activated faster than non-paused IEGs [Saha and Dudek, 2013]. NELF-mediated pausing can also block promoter-proximal nucleosome assembly, leaving chromatin in an open, transcriptionally permissive state, which enhances transcription on affected genes [Gilchrist et al, 2008]. Following this theme, NELF enhances the occupancy of Pol II and TFIIB and is also necessary to sustain histone modifications associated with active transcription at genes under its control [Sun and Li, 2010].

Several signal transduction pathway genes are regulated by NELF. These include components of the MAPK pathway, Erk2 [Gilchrist et al, 2012], and MAP kinase phosphatase-1 for example [Fujita et al, 2007]. Another signal transduction pathway gene under the control of NELF is Rel, which expresses a subunit of NF-κB, a complex which regulates rapid immune response [Gilchrist et al, 2012].

**The function of NELF in development**

In *Drosophila*, NELF is found at the 5’ ends of pair-rule transcription factor genes [Wang et al, 2007], implicating a role in development. Rather than a repressive role, NELF mediated pausing was found to be necessary for the activation [Wang et al, 2010] and synchronized expression of developmental genes [Lagha et al, 2013] including segmentation and dorsal-ventral genes
Four out of the eight Hox genes have NELF-mediated pausing. Interestingly, NELF seems to be required for interactions with insulators Fab7 and Fab8, which is thought to cause higher order chromatin domain structuring, causing a sort of partition between the Hox genes and enhancers for adjacent genetic loci [Chopra et al, 2009].

NELF is also important for developmental processes in mammals. Knockdown of NELF-B induces expression of developmental genes in mice, causing spontaneous cellular differentiation [Amleh et al, 2009]. In humans, defects in NELF can result in developmental abnormalities. For example, truncations in NELF-A are a cause of Wolf-Hirschhorn syndrome [Bergemann et al, 2005].

**NELF-mediated pausing in viruses**

NELF restricts human immunodeficiency virus (HIV) transcription in latently infected cells. Knockdowns of NELF-E significantly diminish pausing activity in HIV and inhibit re-silencing of HIV proviruses [Jadlowsky et al, 2014]. One of the earliest discovered examples of pausing is on the long-terminal repeat (LTR) of HIV. In the absence of the transactivator Tat, NELF pauses Pol II on the LTR and represses transcription [Kao et al, 1987; Wada et al, 1998; Yamaguchi et al, 1999]. An important feature of the LTR is the presence of a stretch of nascent RNA produced at the 5’ end of the gene called the trans-acting responsive element (TAR) [Rosen et al, 1985; Rana and Jeang, 1999] which interacts with NELF-E [Fujinaga et al, 2004; Yamaguchi et al, 2002] and is thought to enhance its pausing activity [Pagano et al, 2014]. Tat recruits P-TEFb
to the gene by binding to TAR [Wei et al, 1998] and much like other genes with NELF, P-TEFb phosphorylates Pol II and DSIF, which releases NELF and allows polymerase to enter productive elongation [Ping and Rana, 2001].

Tat was originally discovered as an anti-termination factor on the LTR [Rao et al, 1987]. The nature of this anti-termination activity was not known until much later: Early termination on the LTR is mediated by NELF coupled with the termination factor Pcf11 which serves in part to further repress HIV LTR transcription by terminating early transcripts [Natarajan et al, 2013]. By eviction of NELF from the LTR on the recruitment of Tat/P-TEFb, Pcf11 activity is diminished and as a result, Tat acts as an anti-terminator.

NELF interacts with other factors on HIV genes as well. A complex of NCoR1, GPS2, and HDAC3, a known transcriptional corepressor [Zhang et al, 2002], physically associates with NELF and helps repress transcription, presumably through histone deacetylation by HDAC3 [Natarajan et al, 2013].

**Activity by NELF apart from pausing**

NELF has also been shown to play a role in 5’ and 3’ end RNA processing. Capping enzyme has been found to overcome transcriptional repression by NELF *in vitro* [Mandal et al, 2004]. NELF interacts with Cap Binding Complex (CBC) which together function in 3’ processing on histone genes [Narita et al, 2007]. There is also evidence that NELF is involved in alternative splicing, at least of transcripts from androgen-dependent promoters [Sun et al, 2007]. NELF is found on the 3’ end of the U2 snRNA gene in humans where it assists in termination [Egloff et al, 2009].
Interestingly, NELF seems to serve a function in regulating signal transduction pathways beyond gene expression. Both NELF-B and NELF-D bind to androgen receptor (AR), both inhibiting androgen receptor-controlled transcription [Sun et al, 2007; Yang et al, 2010]. NELF-D interferes with AR by directly binding to it and by targeting it for degradation by facilitating its ubiquitination [Yang et al, 2010]. NELF-D also binds directly to p21-activated kinase (PAK1) and A-Raf, both components of the MAPK signal transduction pathway, and inhibits them [Cheng et al, 2009; Liu et al, 2004]. Tyrosine phosphorylation of NELF-D by c-Src diminishes binding to PAK1 and A-Raf, reversing this inhibition [Wu et al, 2012].

**Medical relevance of NELF**

The suppression of transcription and maintenance of latency in viruses is a common function of NELF. This phenomenon holds true in Kaposi’s sarcoma-associated herpesvirus, where knockdown of NELF induces lytic phase gene expression [Toth et al, 2012], and in HIV proviral latency, where NELF keeps HIV inactive through Pol II pausing and recruiting factors which terminate short transcripts or keep local chromatin in a transcriptionally non-permissive state. One of the main challenges with curing HIV is depleting the pool of latent virus infecting the host’s T-cells [Chun and Fauci, 1999]. Histone deacetylase inhibitors have been shown to activate viral transcription in cellular models, suggesting a possible means of eliminating latent virus [Demonté et al, 2004]. However, clinical trials with the HDAC inhibitor valproic acid [Göttlicher et al, 2001] have seen limited success [Sagot-Lerolle et al, 2008; Archen et al, 2010].
Still, the central role of NELF in HIV latency remains a subject of interest for viral treatment to this day [Agosto et al, 2015].

NELF has been implicated in various forms of cancer. NELF-B, also known as COBRA1 (cofactor of BRCA1), is involved in BRCT1-mediated chromatin decondensation and interacts with the tumor suppressor BRCA1, which has been implicated in breast cancer [Ye et al, 2001]. NELF-D and NELF-B are down-regulated in breast cancer and are shown to inversely correlate with breast cancer progression [Zou et al, 2010; Sun et al, 2008]. In stark contrast, NELF is up-regulated in several other forms of cancers including upper gastrointestinal carcinomas [McChesney et al, 2006] and hepatocarcinomas, the latter in which increased NELF levels have been shown to increase the metastatic potential and cell invasion capabilities of the cancer [Iida et al, 2012]. Knockdown of NELF-B indeed decreases cell proliferation and metastatic potential in hepatocarcinoma cells [El Zeneini et al, 2017]. The discrepancy may be explained by additional functions of NELF in breast cancer, specifically in regards to estrogen-activated transcription. Estrogen-responsive genes are often regulated post-initiation, frequently by NELF-mediated pausing [Kininis et al, 2009]. NELF-B binds to estrogen receptor alpha (ERα) and represses ERα-mediated transcription [Aiyar et al, 2004]. This inhibits estrogen-induced transcription of trefoil factor 1 (TFF1) in breast cancer [Aiyar et al, 2007]. In breast cancer, TFF1 stimulates cell migration [Prest et al, 2002]. In contrast, transcriptional regulation of TFF1 is estrogen independent in hepatocarcinoma cells [McChesney et al, 2006].
The NELF-E RRM

Our current understanding of how NELF imparts its pausing activity and the function of each subunit has been almost entirely focused on subunits A and E. Of particular interest is the ability for NELF to bind to RNA through the RNA recognition motif (RRM) of NELF-E. RRMs are among the most common and most studied RNA binding motifs in vertebrates [Maris et al, 2005]. They are often coupled with proteins associated with post-transcriptional gene regulation [Dreyfuss et al, 2002]. Structurally, RRMs are typically around 90 amino acids long and forms a $\beta_4\alpha_2\beta_4\alpha_2$ motif, the four $\beta$-sheets stacked as a plane against the two $\alpha$-helices. RNA binds to the surface of the beta sheet, though not always with sequence specificity [Cléry et al, 2008].

Deletion of the NELF-E RRM has been shown to significantly increase runoff transcripts from cell-free in vitro transcription assays compared to wild-type. This has been presumed to be due to a loss of pausing activity [Yamaguchi et al, 2002]. It was proposed that RNA binding by NELF-E through the RRM serves to stabilize the interaction between NELF and DSIF/Pol II in an elongation complex [Yamaguchi et al, 1999]. This hypothesis was recently reinforced by a model that the RRM binding to specific RNA sequences in part determines where polymerase pauses [Pagano et al, 2014]. Furthermore, it has been shown that human NELF-E is phosphorylated by P-TEFb on a serine residue immediately adjacent to RD which diminishes binding of HIV-TAR RNA to NELF-E [Fujinaga et al, 2004], presumably through the RRM [Yamaguchi et al, 2002]. The RRM has also been shown to interact with enhancer RNAs (eRNAs) to facilitate the
release of NELF from paused polymerases, essentially “luring” NELF away, allowing entry into productive elongation [Schaukowitch et al., 2014].

**Scope and significance of dissertation**

The objective of my research is to discover how NELF causes or maintains promoter proximal pausing. To do this, I engineered baculoviruses which express NELF and mutants of NELF and tested these complexes using *in vitro* biochemical assays. Specifically, I investigated the role of the RRM of NELF-E in a nuclear extract-based in vitro transcription system that recapitulates the NELF and DSIF-dependent promoter proximal pausing, which has been extensively characterized on the hsp70 promoter in *Drosophila* [Li et al., 2013]. I discovered that a previously unidentified region on the N-terminal half of NELF-E is required for pausing in vitro while the RRM is not. The discovery that the RRM is dispensable for pausing runs contrary to the current paradigm in the scientific community which implicates both the RRM and its RNA binding activity as being required for pausing activity. With this discovery, efforts may be refocused on alternate mechanisms by which NELF causes pausing with interest on the newfound importance of the N-terminal half of NELF-E.
Chapter 2

Materials and Methods

The purpose of this chapter is to provide a description of the methods used to perform experiments outlined in other chapters of this dissertation as well as to inform the reader of strategies that can be used for working with baculovirus expression systems, in vitro transcription, and protein purification.

Sf9 cell general maintenance

Sf9 cells are a lepidopteran cell line derived from the fall army worm *Spodoptera frugiperda*. They are typically grown in serum-free media at 27°C. Several brands of serum-free media exist, but the cells are highly sensitive to minute changes in media, and of those I tried, Sf-900 II SFM (Gibco) works most consistently. Cells may be grown on a solid substrate (in T-flasks or cell culture dishes), or in suspension in shaker flasks or spinner flasks. For growing cells on solid substrate, the cells tend to grow well in high densities and will have a much longer doubling time below 50% plate coverage (around 5x10^5 cells/mL). For healthy cells in exponential phase, grow each new passage starting at 50% density and split the culture every three to four days. General practice is to cover the cells with enough media to prevent the cells from drying out, but not so much as to limit gas diffusion. To achieve this balance, media is kept at a 1:5 ratio to
surface area (mL to cm²). For example, a 75 cm² T-flask would use 15 mL media. Sf9 cells adhere to the culture vessel, so it is necessary to dislodge the cells in order to pass them. For T-flasks, this may be done by tapping the flask against a table hard enough to knock some of the cells from their substrate. For culture dishes, cell scrapers must be used instead.

Shaker cultures are maintained in baffled DeLong flasks with metal caps. For flasks up to 500 mL, the cultures are shaken at 120 RPM. For 2 L flasks, cultures are shaken at 75 RPM. Media to flask volume ratios should be between 0.15:1 and 0.2:1, so a 2 L flask could accommodate a culture volume between 300 and 400 mL. Cell densities should be kept between 5x10⁵ and 3x10⁶ cells per mL to maintain the culture in exponential phase. Shaker cultures must use serum free media, they will not work with serum supplemented media – cells tend to form dense mats, clump together, and die. After use, the DeLong flasks may be cleaned with soap and water then rinsed with copious amounts of deionized water, autoclaved for at least 30 min on the gravity cycle, then dried in a 45°C drying oven for at least 24 hr. Before starting a new culture in the cleaned flasks, it is advisable to further sanitize it by bombarding it with UV light from the tissue culture hood’s UV lamp. This produces a noticeable improvement in keeping the shaker cultures free of contamination by rapidly growing adventitious agents.

For spinner cultures, cells are grown and maintained in exponential phase in 150 cm T-flasks in TNM-FH (*Trichoplusia ni* media formulation Hink) media with L-glutamine supplemented with sodium bicarbonate (0.35g/L) and 10% fetal bovine serum (FBS) until being ready to transfer to spinner cultures. Spinner
flasks are cleaned thoroughly with soap and water, autoclaved half full with deionized water (20 min liquid cycle), then autoclaved again dry (dry cycle, 20 min exposure, 30 min dry time). The sanitized spinner flasks are then further dried in a 45ºC drying oven. To prepare a spinner culture, cells are transferred from T-flasks and mixed with fresh TNM-FH 10% FBS to a density of 5 x 10^5 cells/ml. The media must be supplemented with serum – serum-free cultures cannot withstand the shear forces imparted by the impeller. Spinner flasks are to be filled no higher than the top of the impeller (For 250 mL flasks, 200 mL maximum) to encourage adequate aeration but not lower than half of the height of the impeller to ensure adequate suspension. Impeller speed is set between 50 and 80 rpm, 60 rpm being ideal. Cells are sustainable without service in this state for up to 7 days.

**Sf9 cell cryopreservation, recovery, and adaptation**

Cells in exponential phase growth can be preserved and frozen in liquid nitrogen for long-term storage. To do this, take some cells from a suspension culture (or dislodge them from a stationary culture) and concentrate them to between 1x10^7 and 2x10^7 cells per mL. Resuspend the cells in the same type of media they have been growing in plus 7.5% cell-culture grade dimethyl sulfoxide (DMSO), aliquot them into 1mL cryopreservation vials, and transfer them to a 4ºC refrigerator for 30 min. Then, freeze the cells by putting them in a cryogenic freezing container and putting it in a -80ºC freezer for 4 hr. After the cells are frozen, transfer to a liquid nitrogen sample storage tank for long term storage.
To thaw and recover the cells, take a vial out of the long term storage container and thaw it out by putting it in a 37\(^\circ\)C water bath for about 5 min or until the remainder of the ice has just finished melting. Then, pipette the thawed cells into a cell culture dish with a surface area at least 75 cm\(^2\). Add media to the appropriate volume (see previous section). Incubate cell culture at 27\(^\circ\)C for 4 hr, then refresh the media to get rid of the remainder of the cryopreservation media. The cells have mostly attached at this point, so simply pipetting or pouring off the media is adequate.

After thawing the cells and removing the preservation media, it may be necessary to let the cells grow for longer than four days before passage to allow them to re-enter exponential phase. This theme also holds true for adapting the cells to almost any new growth condition like changes in media formulation or culturing technique (adapting a stationary culture to grow in a shaker or spinner flask). It is not unusual for cells to take 1 to 3 weeks to adapt and enter exponential phase growth.

**Baculovirus amplification and maintenance techniques**

There are two common methods for amplifying baculovirus titer in tissue culture cells: amplification of low-titer [10 to 10000 pfu/mL (plaque forming units)] virus to mid-titer (10000 to 1x10\(^7\) pfu) virus in suspension cultures and amplification of mid-titer virus to high-titer (higher than 1x10\(^7\) pfu) virus in stationary cultures.

To generate a stock of mid-titer virus, inoculate a suspension culture (shaker or spinner) with a small amount of live virus. Incubate the cultures
normally for 3 to 4 days to infect all of the cells in the culture with virus. Then, harvest the virus by centrifuging the cells and passing the supernatant though a 0.2 µm filter. This filtered supernatant liquid is the mid-titer virus. Filtering the virus reduces the titer somewhat, but is necessary to stabilize it for long-term storage. I did not usually check the titer, but when I did, the virus stock produced this way typically had a titer around 1x10^6 pfu/mL as determined by plaque assay.

To generate high titer virus, infect a stationary cell culture with mid-titer virus and wait until all of the cells have been infected (3 to 4 days). Decant and filter the media. This should produce virus with a titer greater than 1x10^8 pfu/mL.

**Baculovirus plaque assays**

The basic idea of this procedure is to infect a disperse monolayer of cells with serial dilutions of virus, and then overlay the cells with agar media. At some dilution level, the viruses will leave small clearings of dead cells which are called "plaques". By counting the number of these plaques and back-calculating the dilution and amount of virus added to the culture, it is possible to determine the titer of the viral stock in "plaque forming units" per mL or "pfu". While not strictly necessary for baculovirus-based protein expression systems, determining the viral titer can be useful for optimizing the initial virus to cell ratio, the multiplicity of infection (MOI), which influences protein yields.

The procedure I used is derived from a published procedure [Litts, 2003]. This procedure works for cells that are grown in T-flasks in TNM-FH media supplemented with 10% FBS. For a 6 well culture dish, 2 mL of 5x10^5
resuspended cells/mL were added and allowed to attach to the plate for 3 min. I made sure that the cells were in a homogeneous monolayer at the bottom of the plate before proceeding, or the plaques would have been difficult to detect. Serial dilutions of virus between $10^{-3}$ and $10^{-7}$ of the original stock were made in 1 mL TNM-FH media without FBS. The media was aspirated off of the cells and virus was slowly added to each well from most diluted to least diluted. The plate(s) were incubated in the tissue culture hood for one hr and gently rocked by hand every 15 min. Many procedures call for using a rocker or shaker to do this, but this should not be done. I found that rockers only do two things: dislodge cells and make the monolayer blotchy, which makes the assay unreadable, or makes the plaques themselves smeary, which also makes the assay unreadable. Gently rocking a few times by hand is more than adequate to distribute the virus.

During the incubation, I prepared 3% low melting point agarose in ddH$_2$O. After boiling, I made sure the agarose did not get contaminated by any non-sterile material or exposed to air for longer than was absolutely necessary. After briefly cooling, I transferred 10 mL of the sterile 3% agarose to a 50 mL conical tube and kept it in a 37°C water bath to make sure it did not solidify. I warmed a vial of 2x TNM-FH + 20% FBS in the 37°C water bath as well. The water bath was brought to the tissue culture hood and, while maintaining aseptic technique (paper towels help keep excess water from getting everywhere), I added 10 mL 2x TNM-FH + 20% FBS to the vial, mixed it well, and transferred the vial back to the 37°C water bath. The vial was kept in the water bath until ready to use.
After the virus was incubated with the cells for one hr, I aspirated off the media and added 2 mL molten agarose media. This had to be done quickly enough so that the cells did not dry out and the agarose did not solidify before it needed to. Also, I took care not to pour the agarose media directly on top of the cells as this could kill them. Instead, I poured the agarose down the side of the well to give it some time to cool down a bit.

The plates were returned to the 27ºC incubator and incubated for not less than three days. During this incubation time, I periodically checked the cultures, making sure that the agarose overlay was not showing signs of drying out. If it showed signs of drying out (it usually did unless the incubation was done in a 98%+ humidity controlled incubator), I added 1 mL TNM-FH media + 10% FBS to the top of the overlay.

Plaques start to become visible after three days. Staining with neutral red makes these plaques easier to see. I prepared a 1% neutral red solution and an agarose media solution as described earlier in this section. I added 1/100 volume of the 1% neutral red solution to this agarose overlay, aspirated liquid media from each of the culture wells (if any), then add 1 mL of the red agarose to each well. I let the agarose solidify and incubated the cultures at 27ºC for 4 hr. Plaques were generally visible after the incubation. If not, I returned the cultures to the incubator for an additional 1 to 2 days as sometimes it took a bit longer for plaques to become large and clear enough to see. An example of what these plaques look like can be found in figure 2-1.
**Figure 2-1: Baculovirus plaque assay.** Plaque assay to determine the titer of viral stock NELF V3 from Apr-22-2013. Virus was diluted from a range of $10^{-4}$ to $10^{-8}$ (labeled 4 through 8 in red on image A). The well labeled “ctrl” served as a negative control and contained no virus. Plaques are readily detectible in well “4” (Panel B, one is circled as an example). They appear as small clearings in the cells covering the bottom of the well. I detected 23 plaques at the time this picture was taken. This implies a viral titer of $2.3\times10^5$ pfu/mL. Note: plaques continue to develop if plates are left to incubate for a longer time. If they are difficult to visualize, incubate for longer.
**Design of NELF-E mutant proteins for expression in *E. coli***

The coding region of NELF-E with a modified 6x his tag on the N-terminus (amino sequence MGSSHHHHHHSSGS) was cloned into a pST50 bacterial expression vector. Mutants of the NELF-E protein focused on the RRM, where either the RRM (residues 161-250) was deleted (ΔRRM) or critical amino acid residues were charge reversed (N173E, K174D) or substituted with alanines (F169A, S171A, K199A, R201A, and F203A). This substitution mutant was called the 7 point mutant (7mut). To make the ΔRRM mutant, I amplified the entire pST50_NELF-E plasmid using the primer set no_RRM_fwd (5’ – AGCCAGAGGTCTGGTCATCTATTGCGGCCAG – 3’) and no_RRM_rev (5’ - ACCAGACCTCTGGCGCTGTGTCCATG - 3’), then circularized the product using InFusion cloning (Clontech). To make the 7 point mutant, I amplified sections of the NELF-E ORF from the pST50 vector using 5’His-NELF-E_XhoI (5’-AAGctcgagATATACATATGGGCAGC-3’) and NELF-E_SM2 (5’ – AGCACCAGGCACCTCGCTTGATCTCCATGGAGAC – 3’) for the upstream segment. For the downstream segment, primers NELF-E_SM1(Sequence: 5’-GCCGTGGCCGCCGAGATGTCACGAGGATTTC – 3’) and T7_terminator were used. Products from both amplifications were gel purified, combined and amplified with 5’His-NELF-E_XhoI and T7_terminator, cut with NdeI and BsrGI, gel purified, and ligated into purified pST50 vector cut with the same enzymes.

**Expression of NELF-E and NELF-E mutant proteins**

NELF-E and NELF-E mutants were expressed using Rosetta(DE3)pLysS *E. coli* transformed with a plasmid containing recombinant *Drosophila* NELF-E
genes under control of a T7 promoter. The bacteria were cultured in 500 mL Terrific Broth at 37°C with 200 RPM shaking in 2 L Erlenmeyer flasks until an OD₆₀₀ absorbance of 0.3 was reached. The cultures were then incubated at 18°C under the same conditions for 1 hr before inducing with 250µM IPTG. The cells were incubated under these conditions for 16 hr and harvested by centrifuging at 3000x g for 5 min at 4°C. The cells were resuspended in 50 mL low salt lysis buffer (50 mM Sodium phosphate pH 7.0, 10% glycerol, 0.1% Nonidet P-40, 25 mM NaCl, 5 mM imidazole, 5 mM 2-mercaptoethanol, 100 µM PMSF) then flash frozen in liquid nitrogen to complete lysis. The cell lysate was then probe-sonicated for approximately 5 min until no longer viscous and then centrifuged at 16000x g for 20 min at 4°C to clear the lysate. The cleared lysate was then filtered through a 0.2 micron cellulose acetate filter and incubated in batch with 1 mL bed volume TALON Superflow resin (Clontech) on a rotator at 4°C for 4 hr. The resin was centrifuged at 250x g for 5 min to collect it at the bottom of the tube. The supernatant (flowthrough) was removed and the resin was washed in batch with 25 mL of ice-cold wash buffer (same as lysis buffer, but with 500 mM NaCl and an additional 250 mM KCl, 20 mM imidazole, and 1% Triton X-100 instead of the Nonidet P-40 – Nonidet P-40 is no longer manufactured) and then incubated on a rotator for 1 hr at 4°C. The resin was centrifuged as before, and all but 1 mL of the supernate was decanted. The resin was resuspended and transferred to disposable 10 mL gravity columns (Bio-Rad) and allowed to pack by gravity flow. The packed resin was then washed with a total of 20 mL ice cold wash buffer, then 5 mL of ice cold pre-
elution buffer (same as lysis buffer, but with 500 mM NaCl instead of 25 mM and 1% Nonidet P-40 instead of 0.1%). The protein was eluted in 1 mL steps with pre-elution buffer supplemented with 300 mM imidazole.

Protein concentrations were quantified by loading known volumes of the protein preparations on an SDS-PAGE along with a known amount of BSA in a range of 100 ng to 800 ng to act as a standard. The gel was stained with Coomassie brilliant blue R solution (9% acetic acid, 45% ethanol, 0.5% Coomassie brilliant blue R-250), destained briefly with 9% acetic acid, 45% ethanol, and further destained with 5% acetic acid, 7% ethanol. The stained proteins were then quantified using a Gel Doc XR system (Bio-Rad) using the BSA samples to generate a standard curve. With the exception of the seven-point mutation, NELF-E prepared this way is mostly free of contaminants, breakdown products, or truncations (Figure 2-2 A). It is not clear why the seven-point mutation causes more contaminants to appear, but their ratio to the intended product can be reduced by incubating the bacteria during the induction phase at 10°C instead of 18°C (Figure 2-2 B compare lanes 3 and 4). The ~70kDa contaminant that shows up in the wild type and 7mut preparations is probably DnaK, a chaperone that assists in protein folding of newly expressed proteins (Szabo et al, 1994) known to co-purify with recombinant proteins (Birringer et al, 2006).
Figure 2-2: SDS-PAGE Coomassie stained gels of recombinant *Drosophila* NELF-E and mutants from *E. coli*.  

**A:** N-terminal His-tagged wild type NELF-E (WT) run on a 12% 29:1 SDS-PAGE gel.  

**B:** His-tagged NELF-E mutants lacking the RNA recognition motif (ΔRRM) or containing a series of point mutations (7mut) were run on a 12% 29:1 SDS-PAGE gel. A mock purification of NELF-E, prepared the same way as the other proteins but using *E. coli* containing empty expression vector was also run on this gel (lane 2). The 7mut version was expressed at 18°C (lane 3) and 10°C (lane 4). Expression at 10°C results in fewer break-down products and truncations while having a minimal impact on target protein levels. The 10°C preparation was used in all procedures calling for 7mut NELF-E protein.
Expression and purification of DSIF

DSIF was expressed using Rosetta(DE3)pLysS E. coli transformed with a plasmid containing recombinant Drosophila 6xHis-tagged Spt4 and Spt5-FLAG-tagged genes under control of a T7 promoter. The bacteria were cultured in 500 mL 2x YT broth at 37°C with 200 RPM shaking in 2 L Erlenmeyer flasks until an OD$_{600}$ absorbance of 0.3 was reached. The cultures were then incubated at 18°C under the same conditions for 1 hr before inducing with 250 µM IPTG. Lysate preparation was done as described in the previous section. The cleared lysate was then filtered through a 0.2 micron cellulose acetate filter and the salt concentration adjusted to 300 mM by slowly adding 1 M NaCl high salt binding buffer + PMSF (50 mM Na$_2$HPO$_4$ pH 7.0, 10% glycerol, 0.1% Nonidet P-40, 1 M NaCl, 5 mM 2-mercaptoethanol, and 100 µM PMSF). The salt-adjusted lysate was loaded at a rate of 1 mL/min onto a 5mL bed volume TALON Superflow resin (Clontech) 1 cm interior diameter Amicon FPLC column which was pre-equilibrated with 300 mM binding buffer (50 mM Na$_2$HPO$_4$ pH 7.0, 10% glycerol, 0.1% Nonidet P-40, 300 mM NaCl, 5 mM 2-mercaptoethanol, and 100 µM PMSF). The resin was washed with 10 column volumes of ice-cold wash buffer (same as binding buffer, but with 500 mM NaCl and an additional 250 mM KCl). The resin was then washed with 2 column volumes of binding buffer and the protein was block-eluted with 4 column volumes of binding buffer supplemented with 250mM imidazole and collected in four 5 mL fractions.

The eluted protein solution fractions (20 mL) from the TALON column were combined and diluted to a salt concentration of 200 mM NaCl by slowly
adding 10 mL of 1x HGE buffer (25 mM HEPES pH 7.6, 15% glycerol, 0.1 mM EDTA, 1 mM DTT, and 0.1 mM PMSF). This was loaded at 0.7 mL/min onto a Mono Q 10/100 GL (GE) pre-equilibrated with 250mM HGKE (25mM HEPES pH 7.6, 15% glycerol, 250 mM KCl, 0.1 mM EDTA, 1 mM DTT, and 0.1 mM PMSF). Protein was eluted with a gradient from 250 mM KCl to 400 mM KCl over 45 min at a flow rate of 1 mL/min and collected in 2.5 mL fractions. DSIF eluted at approximately 300 mM KCl. Fractions were run on an SDS-PAGE gel and stained with Coomassie blue (Figure 2-3). Care needs to be taken to separate fractions containing Spt5 truncations from full length Spt5 as they tend to elute at similar salt concentrations.
Figure 2-3: Partial purification of DSIF from *E. coli*:  

**A:** Coomassie blue stained SDS-PAGE gel showing crude protein extracts from *E. coli* not induced to express DSIF, or induced to express DSIF, the protein that flowed through the TALON column binding step, and samples from the four 5 mL fractions.  

**B:** SDS-PAGE gel with samples from fractions purified from the Mono Q column, fractions 10 through 19. Fraction 12 was chosen for subsequent assays requiring DSIF.
Production of NELF viruses

Donor and acceptor plasmids were selected based on convenience of restriction sites in each plasmid and the source DNA for each NELF coding sequence. Coding sequence for NELF-A was excised from pST53-NELF-A using XbaI and HindIII and ligated into the acceptor plasmid pACEBac1 cut with the same enzymes. The sequence for NELF-D was excised from pST44-NELF_BDE using BglII and SacI and ligated into the donor plasmid pIDC cut with BamHI and SacI, the BglII and BamHI sites having compatible overhangs. The sequence coding for NELF-E was amplified from pST44-NELF_BDE using the primers 5'His-NELF-E_XhoI and 3'NELF-E_Nhel, purified and cut with XhoI and Nhel, then ligated into the donor plasmid pIDK cut with the same enzymes.

NELF-B was PCR amplified from pST44-NELF_BDE using the primers 5'NELF-B_Nhel and 3'NELF-B_KpnI, cut with Nhel and KpnI and ligated into the donor plasmid pIDS cut with the same enzymes. This turned out to be a mistake because an out of frame start codon was immediately upstream of the Nhel site, which was repaired by PCR amplifying two sequences from the plasmid using the primer pairs p10/NELF-B_Overlap_U and NELF-B_Overlap_D/NELF-B_200R, purifying the products, using them as a template and amplifying with the outside primers p10 and NELF-B_200R. This final product was purified then cut with PacI and Sall, then ligated into the pIDS-NELF-B plasmid cut with the same enzymes. This edited out the aberrant start codon.

All cloning operations using pID vectors required the use of DH10βpir116 E. coli (Fisher) due to their dependence on the R6Kγ origin of replication.
Plasmids were then recombined by adding 1 µg of each to a 20 µL reaction containing 10 U cyclic recombinase (NEB #M0298) and 1x Cre reaction buffer (33 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂ pH 7.5 @ 25°C). This reaction was incubated at 37°C for 1 hr, heat inactivated by heating to 70°C for 10 min, then transformed into DH5α E. coli. The transformants were plated on double selection LB agar plates containing combinations of two of the following: 34 µg/mL chloramphenicol, 100 µg/mL spectinomycin, 30 µg/mL kanamycin. Transformants were tested for single integrations of two donor plasmids. Plasmid DNA was isolated from a transformant containing NELF-A, -B, and –D genes.

One microgram of the plasmid containing NELF-A, -B, and –D genes was added to another recombination reaction containing 1 µg of pIDK-NELF-E, 10 U Cre, and 1x Cre reaction buffer and incubated as before. This reaction was transformed into DH5α and plated on LB agar plates containing all three of the previously mentioned antibiotics. Twenty-four colonies from this plate were used to produce plasmid preparations, which were then cut with a variety of enzymes both to determine how many of each gene cassette was present and to produce a map of each construct.

Once I identified the plasmid harboring all four NELF genes, I used it to transform DH10 E. coli containing the EmbacY bacmid. 400 µL of chemically competent bacteria (CaCl₂ method) were transformed with 500 ng of plasmid with a 45 sec heat shock at 42°C. The bacteria were then transferred to a 10 mL cell culture tube, to which 2 mL of SOC media was added, and allowed to recover by
incubating with shaking at 37ºC for 4 hr. The recovered bacteria were centrifuged, resuspended in 200 µL SOC media, and plated on blue/white screen plates (LB media, 1.5% agar, 15 µg/mL gentamicin, 15 µg/mL tetracycline 20 µg/mL X-gal, and 100 µM IPTG). The bacteria were allowed to grow on the plates for 48 hr.

Bacteria from white colonies were used to inoculate 3 mL cultures in LB media + 15 µg/mL gentamicin, 15 µg/mL tetracycline which were allowed to grow overnight. Bacmid DNA was isolated by completing the first few steps of a mini-scale plasmid purification. Specifically, bacteria from 3mL cultures were centrifuged at 3000 x g for 5 min and the pellet was resuspended in 200 µL solution I (50 mM glucose, 25 mM Tris-Cl pH 8.0, 10 mM Na₂EDTA), then 400 µL of solution II (0.2M NaOH, 1% SDS) was added. The reactions were incubated for 2 min on ice, then 300 µL of ice cold solution III (3 M potassium acetate, 2 M glacial acetic acid) was added to form a flocculent precipitate. The tube was gently mixed by inverting several times and care was taken not to shear the bacmids. These samples were then centrifuged at 16000 x g for 10 min. After pelleting the precipitate, the supernatant was passed through a simple rayon fiber filter (rayon stuffed into a pipette tip) and incubated with 1 µg RNase A for 30 min. The reaction was then precipitated by adding 600 µL isopropanol and gently mixing. The solution was centrifuged at 16000 x g for 20 min at 4ºC and the supernatant was discarded. The pellet was then washed with 75% ethanol. At this point, the crude DNA preparation may be stored in the freezer for extended periods at -20ºC. To complete the crude DNA preparation, the pellet
was centrifuged again at 4°C, 16000x g for 3 min and the supernatant ethanol was decanted. The pellet was air dried for 5 min before gently resuspending in 40 µL 0.22 μm sterile filtered TE (10 mM Tris pH 8.0, 1 mM EDTA). This was the working stock of viral genomic DNA.

To generate live virus, the working stock of DNA was transfected into Sf9 cells in exponential-phase growth. The cells were grown in TNM-FH (Trichoplusia ni media - formulation Hink) (Sigma) supplemented with 0.35 g/L sodium bicarbonate and 10% FBS (fetal bovine serum) in a 75 cm² T-flask. 8x10⁵ cells were transferred into a well of a 6 well sterile culture plate. These cells were left to sit in the tissue culture hood for 20 min undisturbed to allow them to attach to the bottom of the well. The media was aspirated off of the cells and 2 mL serum-free TNM-FH media was added. At this point, the cells were prepared for the addition of the viral genome with transfection agent.

To prepare the viral DNA and transfection agent (Cellfectin II), 100 µL of TNM-FH media was pipetted into two separate sterile microcentrifuge tubes. To one tube, 8 µL of Cellfectin II agent was added and gently mixed into the media. To the other tube, 2 µL of the working stock of viral DNA was added and mixed into the media. The total volume of liquid in one tube was added to the other and mixed by pipetting. The DNA/transfection agent solution was incubated at room temperature for 20 min, then added slowly, dropwise, to the surface of the media in the 6 well cell culture plate, trying to get an even distribution of the transfection solution. The cell culture was returned to the 27°C incubator and allowed to
incubate for 5 hr. The media was then aspirated off of the cells and 2 mL fresh TNM-FH supplemented with 10% FBS was added.

The cells were incubated at 27ºC for 3 days, then viewed under a microscope capable of detecting green fluorescent protein (the protein expressed by the viruses is yellow fluorescent protein, but the GFP filter detects it as well). Cells which fluoresced, about 1 in 1000, were detected at this point and the supernate media was designated as the 0th generation virus (\(V_0\)). To produce the 1st generation virus and the first batch of NELF, 1 mL of resuspended cell slurry from the \(V_0\) sample was added to 200 mL of Sf9 cells at 5x10^5 cells/mL in TNM-FH + 10% FBS in a 250 mL spinner flask. This culture was incubated for three days before checking for fluorescence. After three days, >80% of the cells fluoresced, and the cells were pelleted and tested for the presence of NELF (Figure 2-4). The supernate was sterile filtered through a 0.2 micron cellulose acetate filter and stored at 4ºC. This was designated the 1st generation virus (\(V_1\)).

Subsequent NELF virus preparations, including viruses containing mutant NELF-E, were prepared mostly the same way, though a few changes were made to streamline the virus production process. First, the gene expression cassette of NELF-D was cut from pIDC_NELF-D using BstXI and PI-SceI and cloned into pIDS_NELF-B cut with PI-SceI. This new plasmid was called pIDS_D/B. The Cre recombination reactions each used 100ng of pACEBac1_NELF-A, pIDS_D/B, and pIDK_NELF-E plasmids. NELF-E mutants were cloned into the pIDK vectors as follows. The ΔRRM and 7mut mutants were cloned by PCR
amplifying the ORF from the pST50 vectors used to express these mutants in *E. coli* (described earlier in this chapter) using the primers 5'His-NELF-E_XhoI and 3'NELF-E_Nhel, purified and cut with XhoI and Nhel, then ligated into the donor plasmid pIDK cut with the same enzymes. NELF-E scanning mutants were cloned into pIDK by amplifying the NELF-E ORF using two primers unique to the mutant being prepared and two primers that match the sequence flanking the NELF-E ORF. The primers used that match sequence flanking the ORF were pIDK-Spel_U and pIDK-BstXI_D. A list of the primers used and the sequences for those primers can be found in table 2-1. Two PCR products were generated per mutant: one product made with pIDK-Spel_U and the upstream mutant primer (U) and the other product made with pIDK-BstXI_D and the downstream mutant primer (D). These products were then cloned into pIDK cut with SpeI and BstXI using 3-way InFusion (clontech) cloning.
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pIDK-SpeI_U</td>
<td>AGGTATCGATACTAGTATACGGACCTTTTAATTCAACCC</td>
</tr>
<tr>
<td>pIDK-BstXI_D</td>
<td>GAAGTTATCTGCCAGGCACATGGGTTAACC</td>
</tr>
<tr>
<td>NE1-44_D</td>
<td>CGGATCCTTGACCTTAAGCGTGTCACCG</td>
</tr>
<tr>
<td>NE1-44_U</td>
<td>AAGGTCAAGGATCCGCTGCTGTGATGATG</td>
</tr>
<tr>
<td>NE40-86_D</td>
<td>CAAGCCGAAAGGCAGGAAAGCAGCC</td>
</tr>
<tr>
<td>NE40-86_U</td>
<td>TGCCCTTTCGGCTTGGCGCCTTG</td>
</tr>
<tr>
<td>NE83-121_D</td>
<td>CTCCGAAATCATTAGGAGAAGGACCGC</td>
</tr>
<tr>
<td>NE83-121_U</td>
<td>TTAATGATTTTCGGAGGTCTGTGATCTTGGG</td>
</tr>
<tr>
<td>NE114-167_U</td>
<td>AGGAAGATGTCGTTCTGCTAGACGA</td>
</tr>
<tr>
<td>NE114-167_D</td>
<td>GAACGACATCTTCTGCTCCGGCAACAAG</td>
</tr>
<tr>
<td>NE247-277_U</td>
<td>TAGAGCAAGAAGGGCATCGTTAATGGGC</td>
</tr>
<tr>
<td>NE247-277_D</td>
<td>TGCCTTCTTGCTTAAATGGCTAGCAGC</td>
</tr>
</tbody>
</table>
Figure 2-4: Expression of NELF using baculovirus. A western blot of crude cell lysates. Samples from either non-infected or infected cells were loaded in four consecutive lanes. The blot was cut into individual lanes and incubated with the antibodies against the *Drosophila* NELF subunit indicated. Expected product sizes are marked on the right of the gel image. NELF-A was thought to have broken down over time inside of the cells since this test expression was from cells 96 hours post infection. NELF-B appears as a doublet and consistently does so for other preparations. The reason for this was never determined, but could be from post translational modifications or from starting protein translation at the third amino acid (M3) rather than M1.
Generating a map of the recombined plasmid using restriction enzymes

Since the recombination reaction is done with cyclic recombinase which will randomly recombine donor plasmids into any Cre-lox site, it was necessary to map the number and orientation of the recombined genes. To do this, I first reformatted the plasmid sequence files so that the loxp site was at the ends of the sequence file. The loxp sites are the junctions between plasmids in a recombined construct and serve to, in conjunction with the restriction site, mark the distance between the restriction site and the next plasmid. Then, I chose restriction enzymes that would reliably cut at least one of the plasmids used at least one time, preferably far away from the center of the reformatted sequence file. Then, I cut the recombined plasmid with combinations of each of the selected restriction enzymes. This resulted in DNA fragments indicating the distances between the cut sites. By doing this test for each plasmid involved, it was possible to piece together, like a puzzle, the locations and orientations of all source plasmids in the recombined product.

As an example of this strategy, when I solved the map of the pACEBac1 NELF-8 plasmid (Drosophila NELF-A, NELF-B, NELF-D, His-NELF-E), I cut the plasmid with BamHI, which cut once, off-center, in all four source plasmids. Of the 24 colonies screened, only 6 had 4 cuts total: colonies 8, 12, 18, 19, 23, and 24. Fragment sizes from these digests with BamHI can be found in Table 2-2. I compared these fragments with a series of models constructed from knowledge of where BamHI cut in the source plasmids and chose a set of best fit models. I determined the sequence of the source plasmids in my final construct was

48

In my case, there was ambiguity between some of my models. To remove the ambiguity, I cut the plasmid with PaeI and XbaI and found that NELF-B was head-to-tail with NELF-E and NELF-D was head-to-tail with NELF-A. The final map of the pACEBac1-NELF-8 plasmid is shown in figure 2-5.
Table 2-2: Fragment sizes from BamH I digests of pACEBac1-NELF constructs.

<table>
<thead>
<tr>
<th>Colony</th>
<th>8</th>
<th>12</th>
<th>18</th>
<th>19</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>8200</td>
<td>7900</td>
<td>9000</td>
<td>8500</td>
<td>7800</td>
<td>7200</td>
</tr>
<tr>
<td></td>
<td>4500</td>
<td>4500</td>
<td>9000</td>
<td>5500</td>
<td>7800</td>
<td>5000</td>
</tr>
<tr>
<td></td>
<td>3500</td>
<td>3100</td>
<td>4000</td>
<td>5000</td>
<td>3500</td>
<td>3700</td>
</tr>
<tr>
<td></td>
<td>2200</td>
<td>2700</td>
<td>2500</td>
<td>2500</td>
<td>3500</td>
<td>2200</td>
</tr>
</tbody>
</table>

Constructs isolated from transformant colonies indicated were cut with BamH I, which resulted in four fragments. All sizes are approximate on a 0.8% agarose gel compared with a 1 kb molecular weight ladder. Duplicate size numbers indicate that the fragments presented as a doublet. Only fragment sums totaling approximately 18400 +/- 1000 bp were considered as candidates for recombined complexes with genes at 1:1 stoichiometry, which were colonies 8, 12, and 24.
Figure 2-5: Plasmid map of pACEBac1_NELF from colony 8 as determined by restriction digest. NELF open reading frames and orientation are marked by their gene products (NELF A/B/D/E). polh = polyhedrin promoter. p10 = p10 promoter. SV40 = Simian virus 40 terminator. HSV-TK = Herpes simplex virus thymidine kinase terminator. LoxP = locus of crossing-over P site. Tn7L/R = Transposon Tn7 left (L) and right (R) ends.
Expression and purification of NELF and NELF-E mutant complexes

Finding a consistent and universally applicable purification scheme for preparing a supply of NELF as well as all of my mutants was a critical challenge for the success of my project. NELF was expressed using the baculovirus expression system described earlier in this chapter. The purification strategy hinged primarily on affinity purification by using a 6x histidine tag on either the N-terminus of NELF-E or the C-terminus of NELF-D for complexes lacking NELF-E. Since NELF-E typically expresses in excess of the other subunits and the His tag selects for complexes which only contain NELF-E, including incomplete complexes, further purification steps were required to bring the subunit stoichiometry closer to 1:1. This was achieved by passing the first eluate over a quaternary amine resin such as Source Q, Mono Q, or Poros HQ. The yield of NELF at this point was low even for larger cell cultures and usually needed to be concentrated to be used. Additional purification steps often resulted in substantially reduced yield and were not done to obtain the NELF used in these assays.

NELF-expressing virus-infected shaker cultures were incubated under normal growth conditions for 72 hr, then harvested by centrifugation at 100x g 5 min, washed by gently resuspending in 25 mL 1x phosphate buffered saline (PBS), centrifuged again, resuspended in a 50:50 cell:buffer 1x PBS slurry, then lysed by adding Sf9 cell lysis buffer (20 mM HEPES pH 7.6, 500 mM NaCl, 1% Triton X-100, 10% glycerol, 5 mM 2-mercaptoethanol, 0.1 mM PMSF, 1.6 µg/ml benzamidine HCl, 1 µg/ml aprotinin, 1 µg/ml pepstatin A, and 1 µg/ml Leupeptin)
to 25 mL. Cell lysis was completed by processing with a 40 mL Dounce homogenizer, 10 strokes with a tight-fitting pestle. The raw lysate was then ultracentrifuged at 100,000 g at 4°C for one hour. The cleared lysates were then incubated in batch with 0.5 mL bed volume nickel-NTA agarose (Qiagen) for 4 hr on a rotator at 4°C. The resin was centrifuged at 100x g for 5 min and the supernatant decanted. The resin was washed in batch with 25 mL of ice-cold wash buffer (same as lysis buffer, but with an additional 250 mM KCl and 20 mM imidazole) incubated on a rotator for 1 hour at 4°C. The resin was centrifuged as before, and all but 1 mL of the supernatant was decanted. The resin was resuspended and transferred to a disposable 10 mL gravity column (Bio-Rad) and allowed to pack by gravity flow. The packed resin was then washed with a total of 20 mL ice cold wash buffer, then 5 mL of ice cold pre-elution buffer (same as lysis buffer, but with 300 mM NaCl instead of 500 mM. The protein was eluted in 1 mL steps with pre-elution buffer supplemented with 300 mM imidazole. The first fraction, containing the majority of the NELF, was diluted with 1 mL HGE buffer and loaded onto a MonoQ 5/50 (GE) column pre-equilibrated with 100 mM HGKE (25 mM HEPES pH 7.6, 15% glycerol, 0.1 mM EDTA, 100 mM KCl, 1 mM DTT, and 0.1 mM PMSF). The protein was then step-eluted (5 column volume, 50 mM KCl steps) and collected in 1 mL fractions. Active NELF eluted at the 250 mM KCl step, the most concentrated fractions were either concentrated further using a 5000 Da MWCO Vivaspin 500 concentrator or left as is, provided the NELF concentration was at least 10 ng/μL. This procedure was done to prepare all NELF and NELF mutant complexes used in the experiments shown in chapter
3. It is possible to further purify wild-type NELF using a Mono S column (figure 2-6), but I was either not able to get NELF to bind to the Mono S column for some of the NELF-E mutant constructs or lost too much of the NELF protein by using the column, so this was not made standard procedure.
Figure 2-6: Purification of wild-type NELF. NELF complex consisting of all four Drosophila NELF subunits (A, B, D, and E) was purified using metal affinity resin followed by a 300 mM salt block elution from Source Q resin. Further purifying the NELF using a gradient over a Mono S column greatly increases the purity of the protein (compare Mono S 24% lane to Source Q 30% lane on the Coomassie stained gel), though this was only shown to work with a wild-type complex. Standard purification procedures end after using Q-moiety resin, so the NELF complex protein purity is typically similar to what is seen in the source Q 30% lane.
Expression and purification of P-TEFb

Protein expression and purification of P-TEFb using baculovirus was performed the same as with NELF expressed from baculovirus except that TALON Superflow (Clontech) metal affinity resin was used instead of nickel-NTA. Subsequent steps in purification were not necessary as the product of this purification strategy is sufficient for activity assays (Figure 2-7).

P-TEFb kinase reactions

All kinase reactions were 10 µL and consisted of 1x kinase buffer (NEB), 10µCi γ-32P ATP (7000Ci/mmol) (MP), and various amounts of protein. Reactions were incubated for 1 hr at room temperature, then stopped by adding an equal volume of 2x SDS-PAGE sample buffer, heated to 95ºC for 5 min, then run on a 12% 29:1 acrylamide:bis SDS-PAGE gel 120 V for ~110 min. The gels were coomassie stained, dried, and exposed to a phosphorimager screen for 1 hr before scanning.
Figure 2-7: Coomassie blue stained SDS-PAGE gel of P-TEFb preparation.

The input (soluble), flowthrough, wash, and fractions from a TALON metal affinity purification of *Drosophila* P-TEFb from Sf9 cells expressed using baculovirus. This single step purification is adequate to get relatively pure P-TEFb in adequate quantities for biochemical assays.
Promoter-proximal pausing assay using *Drosophila* nuclear extracts

The RNA capture transcription assay is a robust system amenable to testing NELF activity for a range of mutants. The benefit of this assay is that it shows transcripts from paused polymerase as opposed to earlier methods, which generally looked at larger products and runoff. This experiment uses *Drosophila* nuclear extract to initiate transcription on a plasmid containing an *hsp70* promoter using radioactive nucleoside triphosphates. To minimize non-specific background signal, much of which is from spurious labeling of nucleic acids independent of Pol II, the radiolabeled transcripts are then captured (fished out) from the reaction using an oligonucleotide complimentary to the upstream portion of the transcript. Immunodepletion of specific factors from the nuclear extract permits evaluation of that factor’s effect on transcription. Furthermore, purified factor may be added back to the depleted extract to both confirm that the immunodepletion was specific for that factor alone and to further examine the effect of mutations on the activity of the factor.
Preparing and processing a reaction can be broken into several steps:

1a. Preparation of reaction pre-mixture containing a DNA template.

1b. Allocation and adjustment of nuclear extract

1c. Preparation of limiting nucleoside triphosphates including radioactive nucleotides for transcript labeling.

2. Mixing and pre-incubation of nuclear extract with the reaction pre-mixture (1a with 1b) (PIC phase).

3. Timed mixing of limiting nucleoside triphosphates (1c) with the reaction (pulse phase).

4. Timed mixing of complete NTPs with the reaction (chase phase).

5. Stopping the reaction.

6. PCI extraction and incubation with the capture oligonucleotide (capture phase).

7. Binding captured transcripts to beads and washing away non-specific signal.

8. Loading and electrophoresis on denaturing sequencing polyacrylamide gel.

Reaction pre-mixtures are made by preparing a single pool of pre-mixture buffer containing 57 mM HEPES pH 7.5, 7 ng/μL plasmid template, 14 ng/μL HaeIII cut E. coli DNA, 2.86 mM DTT, ~1-2 U/μL RNasin (Promega), and 1.14 μM flavopyridol (See 1a from summary). An easy way to complete this buffer is to make a table of its components and the volume of each component to be added.
Table 2-3: Example of single sample contents before mixing with nuclear extract.

<table>
<thead>
<tr>
<th>Volume</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8 uL</td>
<td>1 M Hepes (pH 7.5)</td>
</tr>
<tr>
<td>0.1 uL</td>
<td>1000 ng/uL Hsp70 plasmid template</td>
</tr>
<tr>
<td>0.4 uL</td>
<td>0.5 µg/uL HaeIII cut E. coli DNA</td>
</tr>
<tr>
<td>0.4 uL</td>
<td>0.1 M DTT</td>
</tr>
<tr>
<td>0.8 uL</td>
<td>RNAse inhibitor</td>
</tr>
<tr>
<td>1.6 uL</td>
<td>10 uM Flavopyridol</td>
</tr>
<tr>
<td>9.9 uL</td>
<td>ddH2O</td>
</tr>
<tr>
<td>14 uL</td>
<td>Total Volume</td>
</tr>
</tbody>
</table>
Adjusted nuclear extracts contain enough transcriptionally active *Drosophila* nuclear extract [Biggin and Tjian, 1988] to achieve an A280 absorbance equivalent of 0.170 (for example, if the extract used has an A280 of 0.5, then use 6.8 µL of extract and 13.2 µL of nuclear extract buffer or add-back proteins). In this case, nuclear extract buffer refers to 150 mM HEMGK (25 mM HEPES pH 7.6, 12.5 mM MgCl₂, 0.1 mM EDTA pH 7.9, 10% glycerol, and 150 mM KCl). Add-back proteins must be in a similar buffer. If the salt concentration is too high, then transcription will be inhibited (See 1b from summary). Mix 14 µL of pre-mixture buffer to 20 µL of adjusted nuclear extract and incubate at room temperature for at least 10 min. This completes the PIC phase (See 2 from summary). The approximate solution component concentrations at this point are as follows: 38 mM HEPES pH 7.5, 2.9 ng/µL plasmid template, 5.9 ng/µL HaeIII cut *E. coli* DNA, 1.18 mM DTT, ~0.5-1 U/µL RNasIn (Promega), 470 nM flavopyridol, 7.35 mM MgCl₂, 60 µM EDTA, 6% glycerol, and 88 mM KCl.

A solution of limiting nucleotides consists of 1 mM ATP, 1 mM UTP, and 825 nM ³²P-αCTP. To initiate the radiolabeling pulse phase of transcription, 4 µL of this limiting nucleotide solution is mixed with 34 µL of PIC phase reaction and incubated for 10 min at 24°C, converting it to a pulse phase reaction. After 10 min time, 2 µL of a solution containing 2 mM ATP, 2 mM UTP, 2 mM CTP, and 2 mM GTP is mixed with the reaction and incubated for an additional 5 min at 24°C, converting it to a chase phase reaction. To complete multiple reactions, the timing may be staggered by 20 to 60 sec per tube depending on the number of reactions. It is not recommended to choose stagger intervals shorter than 20
sec as this is close to the fastest the nucleotide addition and mixing operation can be done with human hands without making mistakes or spreading radioactive contamination. Using aerosol barrier pipette tips makes it easier to avoid contamination of the micropipettes and may subsequently result in a small increase in the experimenter’s speed. The final concentrations at this point are approximately (NTPs are being depleted during the reaction) 200 µM ATP, 200 µM UTP, 100 µM GTP, 100 µM CTP, and 82.5 nM $^{32}$P-αCTP, 32.5 mM HEPES 7.5, 2.5 ng/uL plasmid template, 5 ng/uL HaeIII cut E. coli DNA, 1 mM DTT, 0.5-1 U/uL RNasIn (Promega), 400 nM flavopyridol, 6.25 mM MgCl$_2$, 50 µM EDTA, 5% glycerol, and 75 mM KCl.

After the 5 min chase phase, the reaction was stopped by adding 200 uL of stop solution (20 mM EDTA, 1% SDS, 200 mM NaCl, 0.1 mg/mL proteinase K, and 0.25 mg/mL Torula yeast RNA). The stopped reaction is then incubated for 20 min and extracted with 200 uL phenol : chloroform : isoamyl alcohol (25:24:1). Afterwards, 100 uL of 500 mM NaCl, 30 mM Tris-HCl pH 7.6 and 8 uL biotinylated transcript-complimentary DNA oligonucleotide is added and the reaction is allowed to incubate overnight at 4ºC. To make the biotinylated oligonucleotide, prepare a reaction containing 1x NEB buffer 4, 250 µM CoCl$_2$, 200 nM complementary oligonucleotide, 20 µM biotin-14-dATP, and 0.4 U/µL terminal transferase (NEB). Incubate the reaction for 30 min at 37ºC, then 10 min at 70ºC to deactivate, then pass the reaction through a desalting column (Bio Rad Micro-Biospin 6 columns are preferred). Incubation of the radiolabeled transcripts with the biotinylated oligonucleotides completes the capture phase.
The captured transcripts are bound to paramagnetic streptavidin-coated beads (30 uL resuspended beads are pre-washed as per company protocol, resuspended in 120 uL 2x binding and wash (B&W) buffer, then added to the captured transcripts). The bound transcripts are then washed twice with 300 uL wash buffer (10 mM Tris-HCl pH 7.5, 5 mM EDTA, 50 mM NaCl, and 0.05 mg/mL yeast tRNA). The bead-bound transcripts are then resuspended in 15 uL formamide sequencing gel loading buffer, heated to 95°C for 5 min, rapidly cooled, centrifuged at 16000x g for 1 min, then the supernatant is loaded on a 7 M Urea 10% 19:1 polyacrylamide Tris-borate-EDTA sequencing gel that has been pre-run at 45 W (approximately 2000 V) for 30 min. The gel is then run at 45 W for 90 min, then dried and exposed to a phosphorimager screen for at least 16 hr.

**EMSAs on reconstituted elongation complexes containing NELF and DSIF**

DNA templates used to form stalled polymerase elongation complexes were prepared by annealing two synthetic DNA oligonucleotides (IDT) and purified by excising the DNA out of a native 0.5x TBE polyacrylamide gel and eluting the DNA from the gel by passive diffusion into 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl at 37ºC over 16 hr. Afterwards, the DNA was concentrated by ethanol precipitation. The complete DNA template contains an 11 nucleotide 3’ single-stranded tail on the template strand and a non-template strand G-less cassette 25 nucleotides long followed by a stretch of DNA containing all four nucleotides 24 bp long. The synthetic oligonucleotide sequences are as follows:
Non-template strand:
GATCTTCATTTCTCATTTCCACTCCCGGATCCTCTTAGAGTCGACCTGCA

Template strand:
CTGCAGGTCGACCTAGAGGATCAGGAGTGAATTGAAATGAAGATCAA
AAAAATTA

Generation of stalled elongation complexes was done by preparing 20 µL reactions containing 50 mM Hepes pH 7.6, 200 mM KCl, 1 mM MnCl₂, 12% glycerol, 0.5 mM DTT, 0.5 mM UpG, 20 U RNasin (Promega), ~125 ng of purified Drosophila Pol II (purified as previously described [Missra and Gilmour, 2010]) and 50 ng of DNA template. These reactions were allowed to incubate at room temperature for approximately 15 min, then 5 µL of either 150 mM HEMG or additional proteins (DSIF or NELF). The reactions were incubated at room temperature for another 5 min, then 2.5 µg HaeIII cut E. coli DNA and 5 µg yeast RNA was added to each reaction as nucleic acid competitors. These reactions were then loaded onto a 4% 39:1 acrylamide:bis, 1x running buffer, 2.5% glycerol, 0.5 mM DTT native gel. The running buffer used was 50 mM Tris-HCl pH 8.5, 0.38 M Glycine, 2 mM EDTA, and 5 mM MgCl₂. The native gel was pre-run for 60 min at 100 V at 4°C, run at 100 V for 30 min after loading the samples, and then run for an additional 4.5 hr at 200 V at 4°C.

NELF-E RNA binding EMSAs

The NELF-E RRM RNA-binding specificity reactions contained 3’-fluorescein-labeled Napt1min RNA, previously developed to bind with high
specificity for the NELF-E RRM [Pagano et al, 2014]. The 10 µL reactions contained 1 pmol Napt1min RNA, 10 mM Na$_2$HPO$_4$ pH 7.0, 12% Glycerol, 200 µg/mL BSA, 100 mM NaCl, 0.2% Nonidet P-40, 100 ng/µL tRNA, and a titration range of NELF-E. Reactions were incubated at room temperature for 15 min and loaded onto a native polyacrylamide gel. The gel formula was 4% 39:1 acrylamide:bis, 1x running buffer, 5% triethylene glycol, and 0.5 mM DTT. The running buffer used was 50 mM Tris pH 8.5, 0.38 M Glycine, 2 mM EDTA, and 5 mM MgCl$_2$. The gel was prerun at 4ºC for 1 hr at 100 V and, after loading the whole reactions, run for 4 hr at 160 V at 4ºC. After running, the gels were not dried, but transferred to plastic wrap and imaged using a Typhoon scanner with the emission filter set to 526 nm to detect fluorescein.

**Quantification of bands for EMSAs and sequencing gels**

Bands from EMSA or sequencing gels were quantified using the analyze/gel function in ImageJ 32bit for PC (program downloaded from http://wsr.imagej.net/distros/win/ij150-win-jre6-32-bit.zip). Image files in .tif format were dragged and dropped onto the running ImageJ program to open the image, which was then set to 32-bit by selecting Image/Type/32-bit from the top bar. The rectangular selection tool was used to draw a box around the first band(s) of interest in the first lane (Figure 2-8A). This box was duplicated for the next lane by selecting Analyze/Gels/Select_first_lane, dragging the box to the next lane. This was done for each subsequent lane by selecting Analyze/Gels/Select_next_lane and dragging the box to the next lane, and so on.
Once all of the relevant bands had boxes placed around them, the Analyze/Gels/Plot_lanes function was used to draw a graph of the signals in each box (Figure 2-8B). The area under each relevant peak was calculated by first closing off the base of the peak. This was done by selecting the line tool and drawing a line connecting the two ends of the base of the peak (Figure 2-8C). The area within the enclosed peak was quantified by selecting the wand tool and clicking on the peak area. This generates a value correlated to the peak area (Figure 2-8D). These values can be used to quantitatively compare lane signals.
Figure 2-8: Image quantification guide.  

**A:** Band land lane quantifications are done by selecting the rectangle tool and making a box around the lanes of interest.  

**B:** The Analyze/Gels/Plot_lanes function results in graphs of the contents of each box.  

**C:** Peaks need to be closed off with the line tool to accurately measure their area.  

**D:** Clicking on the peak area with the wand tool gives a numerical value correlated with the peak area.
Chapter 3

Biochemical investigation of NELF binding and activity

The goal of my project was to elucidate the mode of action of NELF – How it causes RNA polymerase II to pause, where and what it binds to, and to determine if NELF is indeed phosphorylated. Much of my focus was directed at NELF-E in particular, where I expressed individual NELF-E proteins in E. coli and constructed multiple viruses that expressed NELF containing mutant forms of NELF-E. The goal was to identify which regions in NELF-E were required for pausing, if any, and find the part of the protein necessary for association with the rest of the NELF complex. The crux of NELF activity assays utilized the cell-free RNA capture method, which uses nuclear extract to initiate transcription on a Drosophila promoter, captures and chromatographs the transcripts by gel electrophoresis. Binding and interaction tests were done using electrophoretic mobility shift assays or UV crosslinking followed by SDS-PAGE.

NELF-E mutant design strategy

The idea guiding the design of these mutants was twofold: make mutations in the RRM which were thought to no longer allow it to bind to RNA, and make a series of scanning deletions along NELF-E to identify specific regions important for pausing or for complex assembly. To address the first goal, the most obvious way to eliminate RRM activity was to delete the RRM.
I selected a region homologous to the human NELF-E RRM deleted in an early study of NELF [Yamaguchi et al, 2002]. In this study, residues 259 through 343 were deleted. This closely corresponds to residues 161-250 in Drosophila, so I designed the ΔRRM mutant based on those parameters (Figure 3-1).

There remained the possibility that eliminating the RRM altogether would have unintended consequences for the protein expression, folding, or NELF complex binding, so I also chose to make more conservative mutations in the RRM with the intent to only disrupt RNA binding activity. These mutations were a series of point mutations on the putative binding surface which I suspected were critical only for nucleic acid binding. My suspicions were informed by a predicted model of the Drosophila RRM using the Phyre 2 algorithm (Figure 3-1 A) [Kelley et al, 2015] and two studies. The first study produced a crystal structure of the human NELF-E RRM. They found that aromatic residues on the putative binding surface were important for HIV-TAR RNA binding [Rao et al, 2006]. Specifically, they argued that the bases in the RNA can stack with aromatic residues. While there are not as many aromatic residues on the surface of the Drosophila RRM, there were two phenylalanines roughly corresponding to two aromatics in the human structure, so I changed them to alanines. The other study found that if a certain stretch of amino acids in the Drosophila RRM was replaced with sequence found in the human RRM at that approximate location in the structure, then RNA binding specificity was diminished [Pagano et al, 2014]. The “humanization” of that stretch of sequence results in a charge reversal of two
**A**

*D. melanogaster* NELF-E RRM primary structure

GNTIFVSGNKVTEDFLKKTFTNDYGTIVNVSMIEKSRSFVSFAKPSADRAIAEIHGKNNVGINLQVQLARRQP

Drosophila RRM structure prediction (Phyre2)  Human NELF-E RRM crystal structure (Rao et. al. 2008)

**B**

<table>
<thead>
<tr>
<th>Mutations</th>
<th>6x His-tag</th>
<th>LZ</th>
<th>RRM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NELF-E (wild-type)</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>Δ161-250 (ΔRRM)</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>7-point mutation (7mut)</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>Δ1-44</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>Δ40-86</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>Δ83-121</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>Δ114-167</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>Δ247-277</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
</tbody>
</table>
Figure 3-1: Structural diagrams of NELF-E. A: The primary structure of *Drosophila* NELF-E RRM (top), a 3D model of the *Drosophila* RRM generated by Phyre2 (left) and NMR structure of the human NELF-E RRM (right) [Rao et al, 2006]. Bold black letters and residues highlighted in green are the putative RNA-binding surface based on human RRM structural information. Residues with red font and visible side chains are thought to be critical for RNA binding and have been modified in the 7mut mutant. B: To-scale primary structure diagrams of all NELF-E mutants used in this dissertation. Only two domains previously identified in human NELF-E also exist in *Drosophila* NELF: The leucine-zipper-like domain (LZ) and the RNA recognition motif (RRM). Amino acid deletions are denoted by dashed-line chevrons. All constructs have an N-terminal 6x modified His-tag (sequence: MGSSHHHHHHSSGS). Vertical black bars within the RRM denote mutations to alanine, the vertical white bar indicates the location of the charge reversal mutation from NK to ED.
positively charged residues, asparagine and lysine, to glutamate and aspartate, which would result in repulsion of nucleic acids. With this in mind, I opted to only incorporate the charge reversals into the mutant as the entire substitution described in the previous study mutated residues in the interior of the protein which did not appear necessary to disrupt RNA binding and increased the risk of improper folding [Pagano et al, 2014]. I also changed a few other positively charged or polar residues with side chains exposed to the solvent to further reduce the chances of electrostatic interactions that might still be sufficient to bind RNA. The resulting mutant that I called the seven-point mutation (referred to as 7mut), changed the following residues in NELF-E: F169 to alanine, S171 to alanine, N173 to glutamate, K174 to aspartate, K199 to alanine, R201 to alanine, and F203 to alanine (Figure 3-1A).

The scanning deletions were designed as a set of approximately 50 amino acid deletions, eliminating approximately 20% of the NELF-E protein for any given deletion. Together with the RRM deletion, these deletions covered the entire span of NELF-E with the exception of three residues, phenylalanine and two leucines, at the C-terminus. I left the C-terminus alone to minimize the chances of unintended stability issues since protein termini are a significant factor in protein stability [Berezovsky et al, 1999; Sati et al, 2002]. I also used Jpred 4 [Drozdetskiy et al, 2015] on the NELF-E primary sequence to predict structured parts of the protein and designed the breakpoints for these mutations to fall within relatively unstructured regions in an effort to delete only whole, discrete structured domains. I allowed slight overlap in designing the deletions to
minimize partial effects in activity in case one of the unstructured regions was critical for some function of the protein. The results of these predictions are shown in Figure 3-2. Diagrams of each of these mutants are shown in Figure 3-1 B.
Figure 3-2: Secondary structure predictions on NELF-E using Jpred4. The diagram shows the primary structure of NELF-E (jp_SDfye9Q/1-280), alpha helix (Green arrows) and beta sheet predictions (Red cylinders) as well as confidence of those predictions ranked 0 to 9 with higher numbers representing higher confidence (JNETCONF). Predicted buried residues at 25%, 5%, and 0% solvent exposure are listed under JNETSOL25, JNETSOL5, and JNETSOL0, respectively.
The RRM of NELF-E is required for binding to a specific sequence of RNA

Both as a proof of concept and as an attempt to replicate the results of a previous study [Pagano et al, 2014], NELF-E was titrated into reactions containing 1 pmol of fluorescent-labeled Napt1 RNA and analyzed on a native PAGE gel. Both the NELF-E bound and the free RNA bands were quantified as described in chapter 2. Signal intensities are shown as a percentage of the free RNA in lane 1 (Figure 3-3A). Using the disappearance of 50% of the unbound RNA band as a standard, the \( K_d \) of NELF-E for the Napt1 RNA is approximately 25 nM (Figure 3-3B). This is similar to, but slightly greater than the value previously reported using a similar assay, which was 19 +/- 1 nM [Pagano et al, 2014].

To determine if my NELF-E mutants disrupted RNA binding activity, a series of nucleic acid binding experiments were performed with RNA known to bind tightly to the RRM. Torula yeast RNA was used to test for nonspecific binding that might occur in regions outside the RRM of NELF-E. The proteins tested were wild-type NELF-E, NELF-E 7mut, and NELF-E ΔRRM. The preparations used for this experiment are shown in figure 2-2 on page 35 of this dissertation. Only wild-type NELF-E was able to shift the fluorescent RNA band, implying that if the RRM is removed or mutated in critical residues, NELF-E loses its RNA-binding specificity (Figure 3-4).
### A

<table>
<thead>
<tr>
<th>Lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Free</td>
<td>100</td>
<td>47</td>
<td>25</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% E/RNA</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>51</td>
<td>59</td>
<td>72</td>
<td>88</td>
<td>105</td>
<td>71</td>
</tr>
</tbody>
</table>

### B

**Npt1min to NELF-E binding as a function of NELF-E concentration**

![Graph showing the percentage of unbound RNA (%) as a function of NELF-E concentration (nM).](image)
Figure 3-3: NELF-E RNA binding EMSA using fluorescent-labeled Napt1min RNA.  

A: Samples containing 1 pmol of fluorescein-labeled Napt1 RNA were titrated with His-NELF-E protein to the concentrations specified. The samples were analyzed on a native PAGE gel and imaged using a Typhoon scanner set to detect fluorescein. Both the NELF-E bound RNA (notated NELF-E/RNA) and free RNA (notated RNA) were quantified using ImageJ. Signal intensities of quantified bands are shown as a percentage of the signal intensity of the free RNA band (% Free) in lane 1.  

B: A graph of the % free RNA as a function of NELF-E concentration with a best-fit line. NELF concentration is shown in nM on a log scale. The Napt1min-NELF-E binding $K_d$ is slightly less than 25nM, the concentration where half of the RNA molecules are unbound (dotted line).
Figure 3-4: The NELF-E RRM is required for RNA binding specificity. RNA binding reactions containing 100 nM 3’ fluorescein labeled Napt1min RNA aptamer were titrated with protein buffer (lane 1), wild type (WT), 7 point mutant (7mut), or RRM-lacking mutant (ΔRRM) forms of NELF-E to protein concentrations of 125, 250, and 500 nM (lanes 2-10). Coomassie blue stained gel images showing the NELF-E proteins used in this experiment can be found in figure 2-2. These reactions were run on a 4% native PAGE gel and then imaged using a Typhoon scanner set to detect fluorescein.
Identification of NELF-E domains required for integration into NELF complexes

In order to determine which regions in NELF-E are necessary for association with other NELF subunits, a series of scanning mutations were made in NELF-E. NELF complexes containing these NELF-E mutants were prepared as described in chapter 2 of this dissertation. Coomassie blue stained SDS-PAGE gels with each of the NELF types are shown in figure 3-5. Interestingly, deleting only amino acid residues 1-44 in NELF-E greatly diminishes its capability to form a complex with the other NELF subunits (Figure 3-5B lane 1). Since the Δ40-86 mutant can still integrate into the NELF complex, I have designated only the first 39 residues as critical for NELF complex association. If the previously discovered theme [Narita et al, 2003] for NELF subunit interactions holds true for Drosophila, this domain binds to NELF-B.
Figure 3-5: Partially purified NELF complexes containing NELF-E mutants.

A: NELF complexes purified from baculovirus-infected Sf9 cells containing wild-type (WT), RRM deletion (ΔRRM), or 7 point mutation (7mut) NELF-E, or lacking NELF-E altogether (E-less). The particular form of NELF-E is marked with an asterisk in each lane. B: Partially purified NELF containing the NELF-E mutants. Mutant Δ1-44 (Lane 1) failed to associate with the rest of the complex so only the NELF-E was retained after purification.
NELF-E is required, but the RRM is dispensable for pausing *in vitro*

To determine which parts of NELF-E are required for pausing, the NELF mutants were tested in an RNA capture assay. For the experiments shown in figure 3-7, the NELF-depleted nuclear extracts were prepared as previously described [Li et al, 2013] making sure not to let the extract conductivity drop below that of 150 mM HEMG so that NELF and DSIF did not precipitate from the extract. The pausing assay itself was performed as described in the methods section. The protein concentration of the nuclear extract was approximately 30 mg/mL and 10 uL of extract were used per reaction. The transcription template contained the wild-type *Drosophila* hsp70Bbb (CG5834) promoter. Post depletion, less than 1% of endogenous NELF remains in the extract (Figure 3-6 A, compare lanes 1 and 2, lower panel). Add-back amounts were calibrated by percentage of the amount of NELF originally in the extract.

Adding back just 15% of the NELF originally in the extract is enough to restore pausing similar to a mock depleted extract (Example: Figure 3-6B compare lanes 1 and 5). When NELF containing RRM mutations or deletions are added back to the extract, they restore pausing to levels indistinguishable from wild-type (Figure 3-6B). This is surprising as previous studies suggested that eliminating the RRM results in a decrease in pausing activity [Yamaguchi et al, 2002]. However, adding E-less NELF to the depleted extract did not restore any detectible pausing activity, implying that NELF-E is necessary for pausing (Figure 3-6C).
Figure 3-6: The NELF-E RRM is dispensable for pausing *in vitro*. **A:** Nuclear extract immunodepleted of NELF. Depletion was verified by western blotting with antibodies for *Drosophila* NELF-D (top) and NELF-E (bottom). **B** and **C:** Captured in vitro reaction transcripts were run on a 10% 19:1 acryl 7M urea denaturing gel. Reactions used either mock depleted (M) or NELF depleted nuclear extract. NELF of the type specified was added back as a percent of NELF originally in the extract, calibrations shown in panel A. Quantifications for % pausing are based on signals from the section of the images labeled “paused transcripts”. 100% pausing is defined as the paused transcript signal from mock depleted extract samples (lane 1). Quantifications for % runoff are based on signals from the section of the images labeled “runoff transcripts”. 100% runoff is defined as the runoff transcript signal from NELF depleted extract samples. Two replicates of each experiment in panels B and C were performed.
NELF-E has RNA binding activity outside of the RRM.

Mutating or deleting the RRM disrupted RNA binding to Napt1min RNA. However, this alone did not address whether residual NELF-E RNA binding activity was eliminated. For example, a stretch of 5 lysines resides in the N-terminus of NELF-E that might bind RNA through electrostatic interactions. To determine if NELF-E had RNA binding activity beyond what was provided by the RRM, I tested its ability to bind to any of a complex milieu of RNA. NELF-E non-specific RNA-binding EMSAs were done the same as the high specificity binding reactions except that no competitor RNA was used and instead of fluorescent Napt1min, 10ng/µL radiolabeled Torula yeast RNA was used as a probe.

All forms of NELF-E tested, wild-type and the RRM mutants, were able to shift the radiolabeled RNA (Figure 3-7 B). If this binding were due to electrostatic attraction, this RNA binding activity may be disrupted by increasing salt concentrations. However, this RNA binding activity still persists at 300 mM NaCl, well beyond the physiological range. Two controls were done to establish that the binding to torula yeast RNA was due to the recombinant NELF-E. First, an unrelated protein, BSA, was tested. I also did a mock purification of NELF-E in the same manner as the NELF-E preparations but from cells that contained an empty vector. Neither BSA nor the mock preparation caused any significant shift in the RNA (Figure 3-7 A). To determine if it was DNA or RNA that was being shifted, I also set up a binding reaction using wild-type NELF-E, but also included either DNase I or a combination of RNase T1 and RNase A. Only the RNase treatment eliminated the shifted RNA (Figure 3-7 B compare lane 3 to lanes 13
and 14). Taken together, these results indicate that even if the RRM is mutated or eliminated, NELF-E retains some RNA binding activity.
### A

[Image of gel electrophoresis with samples labeled as WT NELF-E, Mock, BSA, and a sample well containing RNA.]

### B

<table>
<thead>
<tr>
<th>NELF-E</th>
<th>NaCl</th>
<th>BSA</th>
<th>wild-type</th>
<th>7mut</th>
<th>ΔRRM</th>
<th>wild-type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

[Image of gel electrophoresis with lanes labeled 1 to 14, showing the effect of different conditions on RNA migration, with a free RNA marker.]
Figure 3-7: NELF-E has RNA binding activity outside of the RRM. A: Reactions containing 100 nM radiolabeled Torula yeast RNA were prepared with or without 200 μg/mL BSA, 500 nM wild type NELF-E, or a comparable amount of mock-purified E. coli protein prepared the same way as the NELF-E proteins (Mock). B: As with panel B, reactions containing 100 nM radiolabeled Torula yeast RNA were prepared with 500 nM wild-type, 7 point mutant (7mut), or RRM-lacking mutant (ΔRRM) NELF-E proteins. Salt concentrations were titrated for each protein at 100 mM, 200 mM and 300 mM NaCl. Reactions containing wild-type NELF-E and 100 mM NaCl were also subjected to either DNase I or RNase treatment for 15 min at 30ºC. Images for both panels B and C were obtained by autoradiograph of dried 4% native PAGE gels of the reactions stated.
A previously undefined region of NELF-E is required for pausing

Since NELF-E is required for pausing and the RRM is not, this prompted me to identify regions of NELF-E that are required for pausing. Using the same assay as the previous section, I tested the remaining NELF-E mutants with the exception of the Δ1-44 mutant since it does not form a complex with NELF. NELF-E Δ114-167 and Δ247-277 complexes restored pausing activity to NELF-depleted nuclear extract just as well as wild-type (Figure 3-8 B). Complexes with NELF-E deletions of residues 83-121 do not restore any detectible amount of pausing activity. Complexes with NELF-E deletions of residues 40-86 do not restore any detectible amount of pausing unless at least 15% of the NELF is added back. At 15% restoration of NELF Δ40-86, pausing levels appear to be restored to levels similar to a 3% restoration of wild-type NELF (Figure 3-8 A). These stretches of residues are fairly well conserved among metazoans but do not have any known secondary structure models. Structure predictions using Jpred failed to detect possible helices, coiled coils, beta turns or sheets, or buried residues with any substantial confidence (Refer back to Figure 3-2). In short, while this region of the protein is critical for pausing activity, the nature of the domain or how it causes pausing activity is unknown.
Figure 3-8: NELF-E residues within amino acids 40 through 113 are required for pausing. Captured in vitro reaction transcripts were run on a 10% 19:1 acryl 7 M urea denaturing gel. Reactions used either mock depleted (M) or NELF depleted nuclear extract. NELF of the type specified was added back as a percentage (3%, 9%, or 15%) of what was originally in the extract. Gel image from panel A is the result of a replicate performed by Yijun Qiu. A replicate of the experiment shown in panel B was performed by Yijun Qiu. Two replicates of each experiment were performed.
NELF-E is required for binding to the elongation complex and the RRM is dispensable for this binding.

To determine which parts of NELF-E, if any, are required to bind to the elongation complex, I titrated NELF-E mutant-containing NELF complexes into reactions containing either transcriptionally engaged Pol II alone or transcriptionally engaged Pol II and DSIF. As shown in figure 3-5 panel A, NELF complexes lacking NELF-E fail to bind to elongation complexes. Furthermore, RRM mutants or deletions bind to the elongation complex comparably to wild type (Figure 3-9 A). This also holds true for deletions of regions immediately upstream (114-167) and immediately downstream (247-277) of the RRM (Figure 3-9 B). Interestingly, mutations deleting regions previously undefined in any study, residues 40-86 and 83-121, resulted in a loss of elongation complex binding activity (Figure 3-9 C).
Figure 3-9: NELF-E is required for binding to the elongation complex, but the RRM and residues 114-167 or 247-277 are not. Pol II elongation complexes with or without 200 ng DSIF were titrated with 10, 20, or 40 ng NELF containing the NELF-E type indicated. The addition of increasing amounts of NELF to Pol II/DSIF complexes coincides with the reduction of the Pol II/DSIF band and the appearance of a supershifted band corresponding to a Pol II/DSIF/NELF complex. In panel A, lanes 15-17 were from a different gel than lanes 1-14.
Mutating the RRM results in more runoff products in G-less cassette assay

The results from the RNA capture assay seem to contradict previous findings that the RRM is required for pausing. To determine whether our system behaves the same as previously reported, I chose to first replicate previously published results [Yamaguchi et al, 2002]. The assay in the previous study used a template with a guanosine-less (G-less) cassette downstream of the promoter. With the addition of RNase T1, any RNA present in the reaction is cleaved at any single-stranded guanosine residue. Therefore, only transcripts from the G-less cassette are present. This assay produces clearly defined runoff products.

I used transcription conditions identical to the RNA capture assay as the basis for this RNase T1 assay. The assay was done as described earlier in this chapter, but did not include the capture steps. Following transcription, samples were ethanol precipitated after the phenol chloroform step, then dissolved in sequencing gel loading buffer and run on a sequencing gel. A plasmid template was used with a partial Hsp70Bbb promoter. The region downstream of the transcription start site lacked guanosines in the non-template strand for approximately 400 nucleotides. The reaction premixtures contained 250 U RNase T1 per reaction. The nucleotides in both the pulse phase and the chase phase included 125 µM 3’-O-methyl GTP, but did not include GTP.

Like previously published results [Yamaguchi et al, 2002], mutating or deleting the RRM of NELF-E resulted in restoration of runoff transcripts in NELF-depleted nuclear extracts (Figure 3-10 A). It is interesting to note that this effect is relatively weak and can be overcome by simply adding more NELF to the
reaction beyond the minimum necessary to completely restore pausing for an RNA capture assay (Figure 3-10 B).
Figure 3-10: Mutations in the RRM allow for more runoff. **A**: Nuclear extract-based *in vitro* transcription assays on a G-less cassette in the presence of RNase T1 generates transcripts specific only for the G-less transcribed region. The longest possible transcript, defined as “runoff” for this assay, is shown in the images from a 6% acrylamide sequencing gel. 2.5 ng per reaction of NELF was added back, which reduces runoff. The greater amount of runoff for the two NELF-E mutants compared to wild-type (WT) indicates a slight defect in NELF activity. **B**: The defect in NELF activity in the mutants can be overcome by adding more (12 ng instead of 2.5 ng) of the complex to the reactions.
RRM mutants retain pausing activity in G-less cassette assay

An important thing to distinguish between measuring runoff transcripts and directly measuring paused transcripts is that, unlike direct measurements, we can only presume that more runoff means a loss of pausing activity. To test if there is substantial correlation between pausing activity and runoff transcripts, I used the same assay as in the previous section, but fished-out the transcripts as done for RNA capture assays. In this case, the amount of paused transcripts remained indistinguishable between wild-type and the RRM mutants (Figure 3-11) indicating that the NELF complexes containing wild-type or RRM mutant forms of NELF-E induce promoter proximal pausing on this G-less cassette template.

There appears to be a discrepancy between the levels of run-off transcripts detected in Figure 3-10 and 3-11: Figure 3-10 shows that when limiting amounts of NELF are added, wild-type inhibits production of run-off transcripts better than the mutants whereas this difference is not apparent in Figure 3-11. The discrepancy in the signal of runoff transcripts could be explained by the resolution of the gels used for the two assays. It was necessary to use a 6% polyacrylamide sequencing gel to observe the runoff transcripts in figure 3-10 while a 10% polyacrylamide sequencing gel was needed to see the smaller transcripts in figure 3-11.

The simplest explanation for why paused transcripts remain unaffected and runoff transcripts change is that only a small portion of polymerases are affected by the mutations in the RRM, while the dominant trend is that NELF
containing RRM mutants can still pause. Another interesting observation is that even wild type NELF does not inhibit runoff transcripts quite as well as non-depleted nuclear extract (Figure 3-11 lanes 1 and 5). It is important to bear in mind that only a maximum of 15% of the NELF was restored to these reactions. It may be that while only 15% is necessary for pausing, adding more NELF ensures that fewer RNA polymerase molecules escape the pause, and therefore a lower abundance of runoff products. Indeed, adding more NELF to the reactions diminishes differences in runoff products between wild-type and mutant NELF (Figure 3-10B). Therefore, it may be that minor differences in activity can be detected by doing titrations of the NELF and then measuring the runoff transcripts.
Figure 3-11: Capture of RNase T1-cut transcripts from a G-less cassette.

Capture of the transcripts from the G-less cassette is possible using the same biotinylated oligonucleotide used to capture transcripts from the Hsp70 plasmid template. These transcripts were cleaved with RNase T1, which cleaves at each single-stranded guanosine residue. These transcripts show that there is virtually no difference between the levels of paused transcripts or runoff products for any of the NELF samples tested.
P-TEFb kinase reactions on DSIF and NELF-E proteins

It was proposed that NELF is phosphorylated by P-TEFb along with DSIF and Pol II immediately before disengaging from the elongation complex [Kwak and Lis, 2013]. The basis for this idea rests on data of human NELF, in which NELF-A and NELF-E are phosphorylated by P-TEFb [Lu et al, 2016]. Mutation of the phosphorylation sites to alanines prevents phosphorylation of NELF-E and suppresses HIV Tat transactivation [Fujinaga et al, 2004], implying that phosphorylation of NELF-E relieves NELF-mediated pausing. Inspection of the sequence of *Drosophila* NELF-E indicates that it lacks the serine residues corresponding to those that are phosphorylated in human NELF-E. To determine if there might be other residues in *Drosophila* NELF-E that might be phosphorylated by P-TEFb, I performed an *in vitro* kinase assay using purified *Drosophila* P-TEFb, human and *Drosophila* NELF-E, and used DSIF as a positive control.

Both *Drosophila* and human NELF-E were tested as targets for P-TEFb, but only the human version of NELF-E could be phosphorylated (Figure 3-12B, compare lanes 6 and 9). Flavopyridol was added to some of the reactions to verify that the kinase activity was specifically from P-TEFb and not a contaminating kinase (Figure 3-12A and B). The additional band identified in lanes containing P-TEFb was likely radiolabeled Cdk9, which runs at approximately that location. A band corresponding to Cyclin T was also present in lanes containing P-TEFb, but was much fainter.
Figure 3-12: Human NELF-E, but not *Drosophila* NELF-E is phosphorylated by P-TEFb. A: Coomassie blue stained SDS-PAGE gel showing proteins specified +/- *Drosophila* P-TEFb. Molecular weight markers in kDa are displayed to the left. B: An autoradiograph of the same gel. Proteins were phosphorylated using $^{32}$P ATP. Cdk9 (red) and Cyclin T are prominent at 42 kDa and 190 kDa, respectively. Human NELF-E is marked in green.
NELF-E is required for pausing but the RRM of NELF-E is not.

To date, no other studies have implicated the importance of NELF-E outside of its RRM. I have found that the first 113 amino acids are important for pausing, either by providing stable association of NELF-E to the rest of the NELF complex, or by providing residues necessary to stably bind NELF to the Pol II elongation complex. While seemingly in stark contrast to the published data, it is important to note that the study which evaluated the role of the RRM in pausing only examined runoff transcripts (about 300 bases in length) and not paused transcripts [Yamaguchi et al, 2002]. The most parsimonious explanation is that a small minority of Pol II is affected by the loss of the RRM through an as of yet unidentified mechanism which results in pausing relief, while the majority of the Pol II remains paused. It is only through the elimination of a critical domain, NELF-E residues 40-113, or by preventing NELF-E from forming a complex with NELF that pausing may be relieved to near completion. Since residues 40-113 in NELF-E is a relatively small part of the smallest subunit in NELF, a series of mutations or cross-linkable residues incorporated into this region of the protein could reveal much about how NELF works.
RNA binding specificity of the RRM is not required for pausing

As the RRM is not required for pausing, the RNA-binding specificity provided by the RRM is not required either. This is not to say that RNA binding serves no significant role in pausing. I found that even if the RRM is eliminated from NELF-E, the remaining residues of NELF-E still have RNA-binding activity. Furthermore, a recent study has identified numerous interactions between single-stranded nucleic acids and positively charged patches of the NELF complex. These positively charged patches are located on other NELF subunits with loosely sequence-dependent binding strength independent of Pol II. Notably, these subunits can interact with the single-stranded non-template strand of DNA or with the sequences of RNAs that would protrude from the RNA exit channel of Pol II at known pause sites [Vos et al, 2016]. It may be that nucleic acid interactions with NELF do not strictly determine if pausing happens on any given gene, but influences where on that gene a paused polymerase comes to rest.

Implications for structural dependency on NELF-E

According to a recent study, residues near the N-terminus of NELF-E, mostly between residues 80 and 110, were found to crosslink with all of the other NELF subunits [Vos et al, 2016]. This is within the region I found to be important for pausing (Figure 4-1A). To understand what this could mean, I propose an alternate interpretation of the data from the structural study. Residues within amino acids 80-110 of NELF-E crosslink with the middle domain of NELF-B as well, not just the N and C-terminal domains. Furthermore, some crosslinking was detected between these three domains of NELF-B and near the LZ domain of
NELF-E. A closer examination of the crosslinking data between the NELF-A/D complex and NELF-B and NELF-E reveals a few interesting details: That there are NELF-B crosslinking sites on both the B/E association side and the RNA-binding side of the A/D complex, and that NELF-E crosslinks near a "groove" in NELF-D with residues 518 and 522 (Figure 4-1B). The model proposed in the study did not provide a model which reconciles these facts, instead opting to present the complex with the A/D RNA binding face away from NELF-B and E.

In my alternate interpretation of the data, I adjust the location of NELF-E relative to NELF-B such that the N-terminal domain of NELF-E interacts with the middle, N, and C-terminal domains. Next, I placed the "groove" of NELF-D in close proximity with NELF-E. With this model, all crosslinking interactions are taken into account while still ensuring that the RNA binding surface and Pol II interacting domains are exposed to the solvent (Figure4-1C). This model shows that the N-terminal domain of NELF-E is in the central part of the NELF complex. It is reasonable to suggest that disrupting or deleting this central domain in NELF not only prevents NELF-E from associating with the complex, but could change the internal structure of NELF. These structural changes could easily result in a loss of binding activity of NELF to the Pol II elongation complex.
**Figure 4-1: An alternative structural model of NELF.** A: A comparison between recent crosslinking data and regions of NELF-E required for EC binding. I found that residues within 40-113 are required for pausing. In human NELF-E, this region crosslinks with parts of NELF-A and NELF-D and crosslinks extensively with NELF-B. B: The crystallized portion of NELF-A/D complex. Protein-protein crosslinking sites are highlighted in orange for NELF-B and purple for NELF-E, with specific residues indicated. RNA binding patches 1 through 4 are circled. C: A cartoon structure of NELF-B and NELF-E with the crystallized portion of NELF-A/D overlaid. The NELF-A/D subcomplex is hidden on the left side of the panel to clearly show NELF-B and E. RNA binding patches 1, 2, and 4 are indicated on NELF-D on the right side of the panel. Figures adapted from a previous study [Vos et al, 2016].
Human NELF-E RRM phosphorylation sites are only conserved in mammals

A recurrent statement in the current literature is that the RRM binds to RNA and can be prevented from binding to the RNA through phosphorylation of NELF-E by P-TEFb. This was found to modulate NELF mediated pausing activity. This is based solely on one study that involved measuring the effects of NELF-E mutants using a transient transfection assay in HeLa cells, in vitro kinase assays and RNA binding assays [Fujinaga et al, 2004]. However, there are no residues in *Drosophila* NELF that are homologous to the ones phosphorylated in humans. Indeed, most of the residues, T288 and S374 in NELF-A and S181 along with the serines in RD of NELF-E are in regions that have no homology to *Drosophila* NELF subunits whatsoever (Figure 4-2). Furthermore, the putative phosphorylation sites found in humans only seem to be present in mammalian lineages. They are not readily detected in other amniotes (reptiles and amphibians) nor are they detected in any other lineage.
**Figure 4-2. An examination of NELF-E primary structure.** Primary sequence alignment of NELF-E between representatives of four phyla of metazoans: echinoderms (*S. purpuratus* - sea urchin), mollusks (*C. gigas* - oyster), arthropods (*T. castaneum* - beetle, *D. melanogaster* - fly), and chordates (*D. rerio* - fish, *X. laevis* - amphibian, *A. carolinensis* - reptile, *M. musculus* and *H. sapiens* - mammal). Serine residue thought to be phosphorylated only exists in mammals and is marked with an asterisk. Alignment was performed using the MUSCLE algorithm [Edgar, 2004]. Conserved consensus residues are highlighted in black, similar residues are highlighted in gray.
**Drosophila P-TEFb does not phosphorylate Drosophila NELF-E**

To determine if *Drosophila* NELF-E could be phosphorylated as it could be in humans, I attempted to phosphorylate NELF-E from either *Drosophila* or humans with *Drosophila* P-TEFb. I replicated the result that human NELF-E could be phosphorylated, but was unable to do so with *Drosophila* NELF-E. This is not surprising as there is no phosphorylation site in *Drosophila* analogous to the one identified in humans; while RD is common only in vertebrates, the domain upstream of RD containing the suspected phosphorylation site appears to only exist in mammals. While it is still possible that mammals have adapted phosphorylation of NELF-E by P-TEFb as a regulator of RNA binding, it is not the case with *Drosophila*.

**Implications of the role of NELF in metazoans**

NELF-mediated promoter proximal pausing occurs in metazoans from *Drosophila* to humans. This phenomenon rarely occurs outside of metazoans. In other words, metazoans have a unique dependence on NELF that other living things typically do not. Since NELF has been shown numerous times to coordinate development and synchronized gene expression, I propose that NELF has become a necessary factor for cell synchronization in early development. Synchronized cell division is a process upon which much of metazoan life is dependent with a few exceptions. One such exception to this is *C. elegans*. In *Drosophila* and vertebrates, early embryo cell division is synchronized [Lee et al, 2014], development of *C. elegans* involves many asynchronous cellular divisions [Ho et al, 2015]. Testing this hypothesis makes studying the exceptions to this
idea an attractive venture. Determining why some nematodes have discarded their dependence on NELF may bring new understanding of this protein complex’s function in nature.

**Closing remarks**

Since the RRM is not required for pausing, this eliminates the possibility of pausing being determined solely by RNA-binding specificity for the RRM. It remains possible that pausing is influenced by single stranded nucleic acid binding specificity, but it appears that binding of NELF to the elongation complex is the first priority in determining whether or not polymerase pauses on any given gene. Determining if NELF binds to the elongation complex necessitates the presence of DSIF, itself requiring at least 18 nucleotides to be transcribed [Missra and Gilmour, 2010]. It may be as simple as an opportunistic phenomenon – NELF binds to the elongation complex at the nearest opportunity (that is, Pol II is transcribing and DSIF is present and unobstructed) which decides where polymerase pauses on any given gene. Nucleic acid sequence may play a role in determining the exact pause locations – those being thermodynamically favorable locations for polymerase to do so. Indeed, the stability of the RNA-DNA duplex correlates with the location of engaged Pol II with short transcripts [Nechaev et al, 2010]. However, a large subset of genes with paused polymerase do not have pause sites at the local energetic minima. Furthermore, the pausing sites can be changed without changing the nucleic acid sequence [Li et al, 2013]. This lends confidence to the opportunistic phenomenon as the dominant factor.
Since the RRM of NELF-E is so well conserved and persistent in organisms that have NELF, it is highly probable that the RRM serves some other function, likely in transcription or RNA processing. While the RRM is not necessary for pausing in the strictest sense, it could still play a role in the recruitment of NELF to genes or the removal of NELF in paused polymerase. For example, the results of my experiments do not contradict the finding that NELF-E binds to enhancer RNAs, which lures NELF away from paused polymerase, attenuating pausing activity [Schaukowitch et al, 2014]. Also, the RRM is required for interaction with the cap binding complex, where NELF plays a role in 3’ RNA processing on histone genes [Narita et al, 2007].

We have yet to discern the nature of the pausing activity NELF provides, though binding of NELF to the elongation complex could be the critical step in pausing, it may impart its activity through allosteric regulation of the elongation complex. If that is the case, then identifying how NELF works may be solved through the use of structure determination technologies. My development of the baculovirus expression system for isolating NELF is an important step towards obtaining sufficient quantities of material for structural studies. With recent advancements in cryo-EM and the possibility of solving ever more complex structures at higher resolutions, solving the structure of an elongation complex bound by NELF is becoming more feasible. With that, we may finally learn how NELF imparts its pausing activity.
Appendix A

Protein – Nucleic Acid Crosslinking Assays

Introduction

Previous studies have shown that NusG, the prokaryote analog of Spt5 [Yakhnin et al, 2016] and DSIF, a complex necessary for the binding of NELF to the elongation complex, interacts with the non-template strand of DNA [Crickard et al, 2016]. This series of experiments used crosslinking strategies based on those done in that previous study in order to determine if NELF interacted with nucleic acids in the context of the elongation complex. How NELF interacts with proteins and DNA in the context of the elongation complex remains unclear and the mapping of crosslinking sites on the DNA non-template strand could provide valuable insight on the structure of a paused elongation complex and how NELF sustains polymerase pausing.

Protein to DNA crosslinking experiments

A series of DNA templates with photoreactive cross-linkable nucleotides incorporated into the non-template strand were prepared similarly as described in previous work [Crickard et al, 2016]. The general strategy was to synthesize the templates by first annealing two oligonucleotides, one serving as the template strand, and the other a variable length non-template strand primer. The primer was then labeled by Klenow exo- polymerase with 5-iodo-dCTP and α$^{32}$P-dATP,
then extended by chasing with all four dNTPs. The templates were then purified to eliminate incomplete products and unincorporated nucleotides. A list of primers used to make these templates can be found in table A-1.
Table A-1: DNA oligonucleotides for assembling crosslinkable templates.

<table>
<thead>
<tr>
<th>Oligo Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>IodoNTS-21/23/25</td>
<td>5’GATCTCTCCTCTCATCTTT 3’</td>
</tr>
<tr>
<td>IodoNTS-19</td>
<td>5’GATCTCTCCTCTATCTTCGACT 3’</td>
</tr>
<tr>
<td>IodoNTS-12</td>
<td>5’GATCTCTCCTCTATCTTCGACTCATT 3’</td>
</tr>
<tr>
<td>IodoNTS-5/7</td>
<td>5’GATCTCTCCTCTATCTTCGACTCATTTCATTCTCATT 3’</td>
</tr>
<tr>
<td>IodoNTS+7</td>
<td>5’ GATCTCTCCTCTATCTTCGACTCATTTCATTTCATTTCAGGATCTCCTCGGG 3’</td>
</tr>
<tr>
<td>IodoNTS+15/16</td>
<td>5’GATCTCTCCTCTATCTTCGACTCATTTCATTTCAGGATCTCCTCGGG 3’</td>
</tr>
<tr>
<td>EC42not1BOTTOM</td>
<td>5’GGCCGCCACCGCGGATCTAGAGGATCCCGGGGA TGGGAATGAGAAATGAGTGTGAAGATAGAGGAGAG ATCAAAAAAATTA 3’</td>
</tr>
</tbody>
</table>
There are a few key changes I made to the original protocol. First, the amount of DNA I used was approximately 150 ng for the bottom strand and 2 fold molar excess of the top strand. This was necessary to increase the radioactive signal per unit template. Second, since there is much less DNA than the original protocol uses, I excised the DNA from the native gel after locating it using autoradiography film instead of UV shadowing. Lastly, I calibrated UV exposure time to 2 min instead of 10. Ten minutes of UV exposure gave a very strong crosslinking signal for Rpb2, which tended to obscure other signals from proteins of a similar size, namely Spt5 and NELF-A. A smaller version of DSIF lacking the C-terminal region (CTR) was prepared similarly to wild-type DSIF (See DSIF purification in this dissertation for reference). DSIF lacking the CTR could be effectively resolved from the Rpb2 band, but trace impurities within the ΔCTR preparation crosslink strongly around 85 kDa and weakly from 80 to 110 kDa (Figure A-2 A). Fortunately, this fell outside the range of NELF-B, NELF-D, and NELF-E.

For any one template, crosslinkable nucleotides were incorporated at residues 21/23/25, -19, -12, -5/7, +7/8, and +15/16 relative to the first guanosine residue. Pol II was allowed to transcribe these templates under limiting NTPs without GTP, which caused the Pol II to stall at the first G in the non-template strand. DSIF or DSIF + NELF was added to the reaction and allowed to bind before UV crosslinking. After UV crosslinking, the samples were digested with DNase I and RNase A, then run on a 7% 37:1 Tris-acetate PAGE (TA-PAGE) gel. The gel was then dried and exposed overnight to a phosphorimager screen.
Polymerase, NELF, and DSIF were confirmed to form a complex on these templates by EMSA (Figure A-1).
Figure A-1: Elongation complex binding EMSA using crosslinkable templates. The templates used for the crosslinking experiment were used in an EMSA to verify that paused, complete Pol II/DSIF/NELF complex could be stably made on those templates. The templates used are designated by their crosslinkable residue location.
Figure A-2: Protein-DNA crosslinking experiments on elongation complexes. A: Autoradiograph of TA-PAGE gel on samples containing Pol II paused on a crosslinkable template (crosslink site -12). UV exposure time was titrated to give a strong signal, but not so much that excessive background signal from polymerase subunits is produced which could obscure the signal from more weakly crosslinking proteins. A 2 min crosslinking time was chosen as appropriate for this purpose. Proteins that crosslinked were identified on the left. Spt5-T is suspected to be a truncation of Spt5-CTR. B: The templates, using the same reaction conditions as in the EMSA, were treated with UV for 2 min, treated with nucleases and analyzed on a large size TA-PAGE gel. Addition of DSIF to the samples increases Rpb2 reactivity and shows that Spt5 crosslinks. There is no trace of crosslinking by NELF-B, NELF-D, or NELF-E. The expected location of the NELF-A band is obscured at least in part by Rpb2, so it is unknown if NELF-A crosslinks.
Protein to RNA crosslinking experiments

Our lab was previously able to see trace amounts of crosslinking between the nascent RNA and NELF-E provided the transcript was over 32 bases long [Missra and Gilmour, 2010]. I attempted to replicate this result using synthetic EC70 and EC42 templates and my baculovirus NELF. Reactions were set up the same way as for EMSAs (See “EMSAs on reconstituted elongation complexes containing NELF and DSIF” in this chapter). For the NTPs during the pulse phase, UTP was substituted with 5-bromouridine triphosphate. After incubation of the stalled complexes with DSIF and NELF and the addition of competitor RNA, the reactions were then transferred to a parafilm-wrapped metal block in an ice bath. The samples were then irradiated with 300 nm wavelength UV-light 2 cm from the source for 10 min. The reactions were then put back into their respective tubes and 2 µL of 2% sarcosyl was added. The tubes were incubated at RT for 5 min and 1 µL of 10% NP-40 was added to quench the sarcosyl. 3 µL of 5 mM MgCl₂ and CaCl₂, 1 µL RNase A, and 4 U DNasel were added to each sample. The reactions were incubated for 60 min at 37°C to digest the nucleic acids. SDS-PAGE loading buffer was added to each sample, heated to 75°C for 5 min, then loaded on a 10% 37:1 tris-acetate PAGE gel. After the dye front ran off the bottom of the gel, the gel was dried and exposed to a phosphorimager screen overnight.

Discussion

In the DNA-protein crosslinking experiments, there is a distinct lack of signal from anything below the minor DSIF contaminant band (Spt5-T in Figure
A-2) for any of the templates tested with the exception of some incompletely digested nucleic acids, which present as a diffuse signal. This result was repeatable for both wild-type forms of DSIF and for the ΔCTR mutation. There remains the possibility that NELF-A crosslinked at one of these sites, but any such signal is obscured by the Rpb2 band (Figure A-2B). The remaining subunits of NELF were probably not contacting the non-template strand of DNA at any of the sites tested.

For the RNA-Protein crosslinking experiments, Rpb1 and Spt5 were readily detected in all applicable lanes. Despite repeating this experiment several times, no trace of a band corresponding to NELF-E could be detected in either the EC42 or the EC70 template for WT NELF or for NELF with mutations in the RRM (Figure A-3). An important distinction between my experiments and work previously done in our lab are the source and purity of the NELF used. NELF used in previous work was purified by FLAG epitope tag from Drosophila embryo nuclear extract and had multiple contaminants [Missra and Gilmour, 2010]. The NELF I used was purified from Sf9 cells and went through additional purification steps (figure 2-6, Mono S 24% lane). Basically, the contaminants were both fewer and different in my preparation. There is a chance that whatever was detected previously was from an RNA-binding contaminant in the preparation.
Figure A-3: Protein-RNA crosslinking assay. Autoradiograph of a TA-PAGE gel with samples from paused elongation complexes on an EC70 template with UV-crosslinkable nascent RNA. Molecular weight markers, in kDa, are notated on the left. Each sample has strong Rpb1 reactivity, which is expected based on previous work [Missra et al, 2010]. Samples containing DSIF also crosslinked to Spt5. Samples containing NELF were expected to have NELF-E bands between 25 and 46 kDa. Unlike the previous study’s results, no crosslinking by NELF-E was detected.
Appendix B

Addback of NELF-E into Incomplete NELF Complexes

Introduction

A potentially worthwhile, yet difficult to execute, experiment would be to incorporate crosslinkable residues into the NELF-complex such that it becomes possible to identify specific binding partners of NELF in the context of a paused elongation complex. However, current technology did not lend itself to easily producing testable quantities of complete NELF complex that has been modified for this purpose. Therefore, I attempted to establish a method to produce crosslinking-ready NELF-E in *E. coli* and incorporate it into partial NELF complexes expressed using baculovirus in Sf9 cells.

Production of Mini-NELF-A

A recombinant virus expressing NELF-B, NELF-D_{His}, and NELF-A or a version of NELF-A missing the 70kDa Diptera-specific insert, but lacking the subunit NELF-E, was prepared using methods similar to those described in chapter 2 of this dissertation. This truncated version of NELF-A, designated Mini-NELF-A, was engineered to remove a large section of the primary sequence, allowing what remained to somewhat resemble NELF-A from the red flour beetle *Tribolium castaneum*. This was done with the goal of increasing
expression yields without causing a loss in ability to associate with other NELF subunits or a loss of pausing activity. This mutant was produced by amplifying DNA from the plasmid pST53-NELF-A using the following primers:

NELF-A_MutN
5’phos – AGGAGTAGTTTTGACGAGGCTAGGGATATCGCATTCCATATTC
GCTTT – 3’

NELF-A_MutC
5’phos – TCCACTCAACGGACGATCATAACTGGACAGCAGGGAAATCCGAAT
AAT – 3’

The PCR amplified product was circularized with DNA ligase and transformed into E. coli. The new plasmid, pST53_Mini-A, was used as the source of the Mini-NELF-A coding sequence for subsequent cloning operations. A comparison between Drosophila NELF, Tribolium NELF, and Mini-NELF-A can be found in figure B-1.
Figure B-1: Sequence comparison for Mini-NELF-A. The primary amino acid sequences of NELF-A from *Tribolium castaneum*, *Drosophila melanogaster*, and Mini-NELF-A are shown in alignment best fit using the Multiple Sequence Comparison by Log-Expectation algorithm (MUSCLE) [Edgar, 2004]. Identical residues are highlighted in black.
**Expression and purification of E-less NELF**

Using the viruses which expressed E-less NELF, proteins were expressed and purified using the NELF purifications techniques described in chapter 2 of this dissertation. A construct containing Mini-NELF-A was chosen as well because it was found to express at much higher levels than E-less constructs containing wild-type NELF-A (Figure B-2A). NELF-E was absent from either preparation, but only Mini-NELF-A could be readily detected. NELF-E expressed in *E. coli* was purified and added (at least 100 µg) to a 40 mL crude cell lysate of cells infected with the E-less NELF expressing virus. Adding the NELF-E this way was critical as adding the NELF-E after completely purifying the E-less complex failed to integrate the NELF-E back into the complex (Figure B-2B). NELF-E may be integrated into Mini-NELF-A complexes, but not full length NELF-A complexes (Figure B-3). Whether this is related to the ejection of full length NELF-A on binding of NELF-B-FLAG to anti-FLAG resin is unknown (See Appendix C of this dissertation).

**Discussion**

Under certain conditions, readily modifiable NELF-E may be integrated into a mostly complete NELF complex. Specifically, the addition of *E. coli* expressed NELF-E to a Sf9 cell lysate containing an E-less and Mini-A NELF complex expressed in Sf9 cells using baculovirus permits integration of NELF-E. Given that it is now known that utilizing the FLAG-tag on the C-terminus of NELF-B results in the ejection of NELF-A from the complex, it may be worth re-trying this method using a His-tag on NELF-D and a different tag to purify NELF-E to
avoid this problem. If the ejection of wild-type NELF-A is related to the apparent failure of NELF-E to integrate into the complex, then this may eliminate the need to use a Mini-NELF-A containing complex as well.
Figure B-2: Expression level comparison of E-less NELF constructs and purified complex integration test.  A: A western blot composite image of E-less NELF containing either wild-type NELF-A or Mini-NELF-A. The blot was probed with antibodies against all four subunits, the secondary antibody for NELF-B was detectible in a different channel than the other three.  B: NELF-E from *E. coli* was incubated with anti-FLAG M2 resin with (+) or without (-) purified E-less, Mini-A NELF, then washed and eluted from the resin with FLAG peptide. There is no significant enrichment of NELF-E from adding purified E-less complex, suggesting there was no capture and integration of the subunit.
Figure B-3: NELF-E may be successfully integrated into E-less NELF complexes containing Mini-NELF-A. A: Coomassie blue stained gel of two preparations of E-less Mini-A NELF, one with *E. coli* derived NELF-E added back to the crude lysate and one without. The preparations were otherwise purified using identical methods, first using TALON metal affinity resin and then using anti-FLAG M2 resin. Though much of the NELF-E flowed through the anti-FLAG column, some was captured and eluted with the other subunits at near stoichiometric levels. B: E-less NELF lysates that contained wild-type NELF-A were subjected to the same treatment as the samples in panel A. However, unlike the Mini-NELF-A samples, very little NELF-E was co-purified with the other subunits after using the FLAG column.
Appendix C

Testing of NELF construct affinity tags

Introduction

Though the 6x His tag on the N-terminus of NELF-E served a major role in the general NELF purification scheme, that tag alone was not adequate to obtain protein preparations that were greater than 90% pure or at 1:1 stoichiometry for all subunits. What was especially important was the removal of excess NELF-E, which was a recurring problem from expression. This led me to try producing a series of viruses expressing NELF which had different tags in different places in addition to the 6x His on NELF-E. Of all the other tags tested, only one tag, a 6x His tag on the C-terminus of NELF-D, was found to be useable. This tag, of course, was redundant with the tag used for NELF-E, and only has utility in complexes that either lack NELF-E or, hypothetically, has a different tag on NELF-E. Table C-1 is a chart showing all of the constructs tested.

NELF affinity tags

Of the tags tested which did not work, the failures can be classified into three categories: disruption of protein expression or folding, inability for the tag to bind to the resin, or, surprisingly, ejection of one or more subunits from the complex on binding to the affinity resin. One of the first alternative tags tested was a 2xFLAG-tag on the C-terminus of NELF-D. This seemed a natural choice
at the time since a previous study [Missra and Gilmour, 2010] found that the FLAG-tag in that position was useable for purification of NELF from Drosophila embryos. Additionally, a 2x FLAG-tag was readily detectible by western blot and bound more tightly to the affinity resin, allowing stringent washes. After initial purification attempts, it became apparent that NELF-D_{FLAG}, while expressed, was expressed at extremely low levels (Figure C-1). Subsequent attempts to improve expression of NELF-D_{FLAG} by increasing the initial cell culture density and multiplicity of infection yielded no significant changes.

The next thing I tried was to put a strep-tag on either terminus of NELF-A. The amino acid sequence was WSHPQFEK appended to the C-terminus or MASWSHPQFEK for the N-terminus. Both of the viruses were constructed starting with pACEBac1_NELF. The construct that has the strep-tag on the C-terminus expressed at levels similar to the original construct, but the tagged NELF-A would not bind to the strep resin unless it was released from the rest of the complex (Figure C-2). Putting the strep-tag on the N-terminus of NELF-A resulted in a highly unusual expression pattern; NELF-A did not show up as a clear band, NELF-D was highly over-expressed, and, contrary to all previous work with NELF expression, NELF-E had only a slight presence in soluble fractions (Figure C-3). Instead, the majority of NELF-E was rendered insoluble. It is unclear why this is, but my best guess is that when NELF-A is expressed at relatively equivalent levels to NELF-E, the N-terminal strep-tag on NELF-A is insoluble in any complex that contains it, and so any NELF-E integrated into a NELF-A-containing complex was precipitated.
Table C-1: NELF constructs tested with brief description of results.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, B, C</td>
<td>Working. Required Mono S column to purify.</td>
</tr>
<tr>
<td>A, D, B, E</td>
<td>Working. Cannot be purified with Mono S.</td>
</tr>
<tr>
<td>D, B, E</td>
<td>C-terminal Flag-tag on NELF-D disrupts protein expression and assembly.</td>
</tr>
<tr>
<td>α, D, B</td>
<td>Utilization of the FLAG-tag on NELF-B ejects NELF-A from the complex.</td>
</tr>
<tr>
<td>α, D, E</td>
<td>Str tag on C-term of NELF-A obstructed by NELF-D.</td>
</tr>
<tr>
<td>A, D, E</td>
<td>Str tag on N-terminus causes expression problems</td>
</tr>
</tbody>
</table>

Figure C-1: NELF-D_{FLAG} expresses at very low levels. This western blot was probed with antibodies against all four NELF subunits. Molecular weight markers are indicated on the left, expected NELF product sizes are indicated on the right. NELF-D_{FLAG} was barely visible even in the soluble lysate (lane 3 marked with arrow). This was not due to lack of antibody sensitivity as NELF-D is readily detected in samples from the wild-type preparation using the same antibody (lane 13). Low expression levels of NELF-D resulted in no detectible capture of NELF-A by the TALON resin (lane 6), and no detectible intact complex after purification by anti-FLAG resin (8-12).
Figure C-2: NELF-A-Str constructs cannot be purified using the strep-tag.

Coomassie blue stained SDS-PAGE showing samples from NELF-A-Str complex protein preparation. Molecular weight markers in kDa are shown on the left and expected product sizes are shown on the right. NELF complex purified by metal affinity chromatography was loaded onto Strep-tactin resin (IBA). Almost all of the protein flowed through (Str FT). The protein which did bind was only NELF-A (Str F2 and Str F3), implying that NELF-A needed to separate from the rest of the complex before binding to the strep resin. This suggests that full complexes cannot be purified using a C-terminal strep-tag on NELF-A.
Figure C-3: N-terminal strep tag on NELF-A causes expression problems.

This western blot of samples taken during an attempted purification of NELF complex containing an N-terminal strep-tag on NELF-A was probed with antibodies against NELF-A, NELF-D, and NELF-E. In this preparation, the Str-NELF-A was not clearly identifiable by western blot (Suspected product indicated on the right) and the majority of NELF-E was rendered insoluble (Compare lanes 1 and 2. NELF-E marked with asterisk). As the majority of NELF-E was insoluble, very little could be captured during the metal affinity purification.
Concurrently with my attempts to produce strep-tagged NELF complex, I prepared a virus which expressed NELF-B with a 2x FLAG-tag on the C-terminus of NELF-B. This construct expressed all subunits similarly to the original NELF virus, except that NELF-B was no longer a doublet on an SDS-PAGE gel. The complex assembled correctly – capturing His-NELF-E on metal affinity resin captured the other three subunits as well. However, when attempting to bind the complex to anti-FLAG M2 resin, the NELF-A subunit was not captured (Figure C-4, lanes 6 and 7). NELF-A must have been in complex with the other subunits in order to have shown up in the nickel eluate (lane 3) and all of the NELF-B was captured by the anti-FLAG resin (lane 5). It follows that at least some of the NELF-B in the FLAG eluate was at one time in complex with NELF-A, ergo, when the complex was bound to the anti-FLAG resin, NELF-A dissociated.

**Conclusion**

Attempts at finding additional tag sites and types were unsuccessful. This highlights the sensitivity of the NELF complex for added affinity tags and should be considered for future attempts at tagging NELF subunits. To date, the His-tag on the N-terminus of NELF-E and the His-tag on the C-terminus of NELF-D remain the only tags tested that do not cause expression or purification problems. It is possible that a FLAG-tag on the N-terminus of NELF-E may be tolerated by the complex and can be used to successfully purify NELF as this has been used to purify NELF subunits from *Drosophila* embryos [Missra and Gilmour, 2010].
Figure C-4: Utilizing the FLAG tag on the C-terminus of NELF-B causes NELF-A to dissociate from the complex. This western blot of samples from an attempted NELF purification containing NELF-B with a FLAG-tag on the C-terminus was cut into three sections and probed with anti-NELF-A antibody, anti-NELF-B and anti-NELF-D antibody, and anti-NELF-E antibody. The lower panel is the section of the top panel between markers 80 and 58 viewed in a different emission channel. All subunits are present in the soluble fraction of the cell lysate and the eluate from the nickel resin. Flag FT1 is flowthrough from an earlier attempt at purifying this complex where I did not give the nickel eluate enough time to bind. Flag FT2 is the flowthrough of the preparation from which the two subsequent fractions were purified. The two fractions purified using anti-FLAG resin are shown in lanes 6 and 7, where NELF-A is conspicuously absent.
REFERENCES


Gilmour DS (2009) Promoter proximal pausing on genes in metazoans. Chromosoma 118(1), 1-10


VITA
Michael James Fisher

EDUCATION:

Ph.D., Biochemistry and Molecular Biology, Penn State University, May 2017

B.S., Biochemistry, John Carroll University, Dec 2008

PUBLICATIONS:


