DISSECTING THE FUNCTION OF *DROSOPHILA MELANOGASTER* INSULATORS USING RNA APTAMERS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Li Yao

May 2015

© 2015 Li Yao

ABSTRACT

Heat shock response in *Drosophila melanogaster* is a complex process causing changes of the cellular environment in reaction to the heat shock induction. RNA transcription is dramatically changed during heat shock response and many regulatory factors are required. Chromatin insulators are found to be involved in the transcriptional regulation process, but much remains to be understood on how insulator function. Described herein are multiplexed selections and characterization of inhibitory RNA aptamers against several *Drosophila* insulator proteins using an efficient selection method, RAPID (RNA APtamer Isolation via Dual-cycles), and high-throughput sequencing. Background binding sequences are identified through a series of comprehensive analyses on the sequencing dataset. I have also examined the binding affinity of the top aptamer candidates and designed assays to map the binding elements. Aptamers generated here can be utilized as a very powerful approach to dissect the insulator function *in vivo* in the well-characterized heat shock system.

BIOGRAPHICAL SKETCH

Li Yao was born in Changsha city in Hunan province of China, on October 22nd 1988. She spent the first 18 years in Changsha, where people like eating spicy food and have bright personalities like chili peppers. Like all the Chinese kids, Li was sent by her parents to all kinds of classes after school, such as mathematics, dancing, drawing, calligraphy, singing and musical instruments. However, Li didn't like any of them and she chose to join the athlete team in the elementary school because she liked doing out-door activities. Although Li didn't like attending the drawing class, she showed drawing talents since she was very young and she has been drawing for fun.

Before the last semester of middle school, Li never liked studying and she spent all the time hanging out with friends until her mom told her that she is going to send Li to a high school that Li didn't like unless she gets into the Olympiad Class. Li studied for two weeks before the entrance exam and with the best luck she has ever had she got into the Olympiad class. Li joined the Biology Olympiad team because no mentors in the math, physics and computational science team wanted her by looking at her terrible grades. But biology was the field she liked the most. Li started to study very hard and became the best student in the Biology Olympiad team in her school. Li entered Zhejiang University without taking the college entrance exam, to study biological science, where she had a lot of much fun learning biology and she did undergraduate research in a neuroscience lab.

Li began her graduate studies in the field of Genetics and Development at Cornell University. Ithaca, NY, in August 2011. Two years after Li started her graduate study, she realized that research is not what she wants to do for career. Li decided to combine two of her favorite fields, science and art, and joined the Medical Illustration program in Johns Hopkins Medical Institution.

To my family

ACKNOWLEDGMENTS

I would like to thank my thesis adviser, John T. Lis, for his support, advice and encouragement. I am very grateful to have the opportunity of working under such a great academic environment with my advisor. I would like to thank my committee members, Ailong Ke and Paul Soloway for all the suggestions and help on both research and other aspects of life. I would also thank our DGS Kelly Liu for helping me when I was having difficult times.

I would like to thank all the members in Lis lab for generously sharing their knowledge and experience with me. I have learnt a lot from them and enjoyed working with them. I would like to express my thanks especially to the aptamer group members in Lis lab, John Pagano, Abdulalh Ozer, Fabiana Duarte, Jacob Tome and Kylan Szeto. John Pagano, who was my mentor when I rotated in Lis lab, helped me to start my research in a completely new field. Abdullah is always very supportive for answering any questions I have for my experiments. Fabiana and Jacob, my lab partners and friends, helped me a lot in both research and getting adapted into the new culture and new language environment. Kylan, who is from the Applied Physics department in Cornell, provided me many insightful comments in my studies.

I would like to thank all my friends I have met in Ithaca who have treated me as a family member and made me feel secured. I would also like to thank my family for their endless love, support and understanding for everything I do, even when I am not able to do anything for them from so far away. None of my accomplishments can make without them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BEAF32-Boundary Element-Associated Factor of 32kD

CP190- Centrosomal protein 190kD

Su(Hw)-Suppressor of Hair wing

Mod(mdg4)- Modifier of mdg4

GAF-GAGA Factor

dCTCF-*Drosophila* CCCTC-binding facotr

CTCF- CCCTC-binding facotr

Zw5-Zest White 5

Scs-specialized chromatin structure

TAD-Topologically Associating Domain

SELEX-Systematic evolution of ligands by exponential enrichment

RAPID- RNA Aptamer Isolation via Dual-cycles

MEDUSA- Microplate-based Enrichment De- vice Used for the Selection of Aptamers

Pol II – RNA polymerase II

SDS-PAGE- sodium doceyl sulphate polyacrylamide gel electrophoresis

PRO-seq- Precision nuclear Run-On and sequencing assay

FP-Fluorescence Polarization

F-EMSA-Fluorescence-Electrophoretic Mobility Shift Assay

K_D-Disassociation constant

HiTS-RAP- High-Throughput Sequencing-RNA Affinity-Profiling

SD- standard deviation

MEME-Multiple Em for Motif Elicitation

ChIP-chip-Chromatin immunoprecipitation with DNA microarray

ChIP-seq-Chromatin immunoprecipitation with sequencing

MNase-Micrococcal Nuclease

qPCR-quantitative PCR

CHAPTER 1

INTRODUCTION

In eukaryote cells, genome DNA is associated with histones and form nucleosomes. With other nuclear protein and RNA components, the genome is packed in the nucleus as chromatin fibers, which can actively switch between loose and dense conformations during different stages in the cell cycle. Apart from being structurally involved in the chromatin formation, the protein and RNA components in the chromatin also act as regulatory factors in many essential nuclear processes, such as RNA transcription (Fuda *et al.*, 2009). RNA transcription is the key process of generating regulatory RNA and mRNA templates for proteins. The transcription process is sophisticatedly and tightly regulated at various steps to quickly react to the turbulence in the environment, maintain the homeostasis and trigger the important developmental events (Fuda *et al.*, 2009).

1.1 Chromatin insulators and RNA transcription.

When scientists didn't know much about how the genome works, they believed that in order to have all the transcription events smoothly regulated, the genome must be partitioned into higher-order physical domains, and each domain is mostly governed by the regulatory factors within it, thus the transcriptional activity of each domain is relatively independent from flanking domains (Benyajati and Worcel, 1976; Igo-Kemenes and Zachau, 1978; Weisbrod, 1982). One well-known example of regulation of transcription from the regulatory factors is the interactions between

promoters, the regions that initiate transcription, and enhancers, the elements that can activate transcription of certain promoters (Ho *et al.,* 2014; Arensbergen *et al*., 2014). It was believed that there are some factors acting as the barriers to dictate the boundaries of different domains, maintain the chromatin status within the domains and block the regulatory elements from acting across domains. Such factors were defined as "Chromatin Insulators" (Udvardy and Schedl, 1984; Udvardy *et al*, 1985). Based on the definition of insulators, if an insulator is between an enhancer and a promoter, the enhancer doesn't activate the promoter; and if a gene is within two flanking insulators, the transcription of this gene is not affected by regulatory elements outside of this domain. Some DNA-protein complexes were identified as potential insulators in different organisms, but the tests available to validate that they act as ideal insulators were very limited. Traditionally, several trans-genetic assays were used to test these features. However, the results of these assays are not always consistent (Kellum and Schedl, 1991; Kellum and Schedl, 1992).

1.2 Insulators mediate the formation of the higher-order chromatin structure.

Although it is known that chromatin has higher-order structures, such as forming chromosomes during cell division, for a long time, how chromatin is organized during interphase is not clear. Little was know about whether the three-dimensional structures of the nucleus have anything to do with the genome function until a method called Hi-C, an assay that comprehensively detects chromatin interactions, was performed in mammalian cells and *Drosophila* cells in 2009 and 2012(Lieberman-Aiden *et al*., 2009; Belton *et al*., 2012; Sexton *et al*., 2012).

Chromatin regions that are closely associated are fixed by cross-linking and the segments are linked with ligase and then sequenced by high-throughput sequencing. By mapping each segment of the cross-linked sequences back to the reference genome, they pictured a rough three-dimensional structure of the chromatin. Bing Ren's lab was able to partition the genome into mega base-sized Topologically Associating Domain (TADs), in which several chromatin segments from adjacent or distal genome regions are physical associated (Dixon *et al*., 2012). This suggests that the chromatin is not simply divided linearly into domains. With an optimized high-resolution Hi-C technique, Ren lab determined over one million long-range chromatin interactions at 5-10kb resolution in human genome (Jin *et al*., 2013).

Based on the Hi-C result performed in *Drosophila* embryo nuclei by Sexton *et al*., the chromosomal contacts are hierarchically organized into topologically associated domains as well. What is more interesting is that the earlier identified *Drosophila* insulators were found to reside at the boundaries of the topologically associated domains (Sexton *et al*., 2012). Previously, *Drosophila* insulator proteins have been shown to form interactions *in vivo* between insulators or with other sub-nuclear structures. Seven insulator proteins have been identified in *Drosophila* (Su(Hw), dCTCF, BEAF, GAF, Zw5, CP190 and Mod (mdg4)) (Gurudatta and Corces, 2009). A chromatin binding protein Chromator is a potential insulator protein due to its similar enrichment on the genome to other insulator proteins. Five subclasses of insulators were characterized, each containing a core DNA binding insulator protein (Su(Hw), dCTCF, BEAF, GAF and Zw5) and shared components CP190 and Mod (Mdg4). The shared components have common dimerization

domains, which can mediate homo- or heteromultimerization between insulators (Phillips-Cremins and Corces, 2013). Combining the Hi-C results, it suggests that insulators may be the key factors defining the topologically associated domain boundaries through these interactions (Sexton *et al*., 2012, Vogelmann *et al.,* 2014). Such information changed our traditional understanding of insulators. From this view, blocking of the regulatory factors from acting on certain promoters or defining linearly divided chromatin domains may just be part of the consequences of insulator interactions in some genomic contexts. How a gene reacts to adjacent or distal regulatory factors may be directed by its relative positions with regard to nearby insulators and the kinds of insulators around it. With this idea in mind, the inconsistent results shown in the canonical transgenic assay may now be explained since insulators' function is context-dependent (Kellum and Schedl, 1991; Kellum and Schedl, 1992).

Many studies about RNA transcription established their hypotheses based on the linear structures of the chromatin or genes of interest, which is somewhat inevitable due to the lack of knowledge of the higher-order structure of the genome. Insulators were found to be associated with many transcription events (Hou *et al*., 2012; Guertin and Lis, 2010; Guertin *et al*., 2012; Fuda *et al.,* 2015, and unpublished paper from Lis lab). However, the mechanism is poorly understood. Structures and functions are always closely integrated. If we can dissect how insulators are involved in the formation of the hierarchically organized topological associated domains, we may be able to explain what roles insulators play as transcriptional regulators.

1.3 Insulators are involved in the transcriptional regulation of heat shock genes.

The heat shock response in *Drosophila melanogaster* has been used for years as a powerful model to study transcription regulation. Exposure of *Drosophila* cells/tissues to 37°C (heat shock) induces genome wide transcriptional changes to respond to the stress caused by the temperature increase in the environment. Heat shock also induces dramatic chromatin decondensation at many genomic loci, which is observed optically as chromatin puffs on polytene chromosomes, and also results in a rapid induction of heat shock genes at those loci (Petesch and Lis, 2008; Petesch and Lis, 2012).

1.3.1 Insulators are potential barriers for nucleosome loss at *Hsp70* loci

Heat shock protein 70 (*Hsp70*) is one major induced gene upon heat shock (Ashburner and Bonner, 1979). There are two copies of *Hsp70* genes at the 87A locus about 1.7kb apart arranged in opposite orientation. Steven Petesch, a former graduate student in Lis lab, determined a high resolution profile of the changes in chromatin structure at the 87A locus upon heat shock using Micrococcal Nuclease (MNase) digestion followed by quantitative PCR (qPCR) (Petesch and Lis, 2008). He demonstrated that within 2 minutes after heat shock, the nucleosomes are depleted from chromatin rapidly and that the initial loss is faster than the movement of RNA Polymerase II (Pol II) and is transcription-independent. Steven also showed that the nucleosome loss is not confined to the *Hsp70* genes; instead, it proceeds beyond the polyadenylation sites and extends into flanking genes. However, the puffing at 87A locus observed in polytene chromosomes is enclosed to a certain area, suggesting that

there are barriers stopping the nucleosome loss to ensure precise spatial transcription regulation. Noticeably, two DNase I hypersensitive sites, scs and scs' (specialized chromatin structure), locate at either end of 87A heat shock locus (Udvardy and Schedl, 1984). Insulator proteins Zw5 and BEAF32 bind to their consensus DNA elements within scs and scs', respectively (Hart *et al*., 1997; Gaszner *et al*., 1999). Due to the barrier nature of insulators (Kellum and Schedl, 1992), they may be critical factors for the barrier function. Steven thus examined the chromatin structure flanking these two sites, either within or beyond the scs-scs' region. He showed that nucleosome loss proceeds through regions within but not beyond the scs and scs' sites (Petesch and Lis, 2008). These results defined that scs and scs' are the barriers of the heat shock induced nucleosome loss and suggested that Zw5 and BEAF32 insulators may enable the barrier function. At the 87A locus, a direct interaction between Zw5 at scs and BEAF32 at scs' has been observed *in vivo* (Blanton *et al*., 2003). Taken together, the interactions between Zw5 and BEAF32 may be essential for the barrier function.

1.3.2 Insulators are enriched in the promoter regions of heat-shock induced repressed genes.

Heat shock induction to *Drosophila* cells triggers fast and wide spread effects on RNA transcription on the genome. Transcriptions of some genes are dramatically enhanced, while many are repressed (unpublished data). In order to monitor the continuous genome wide change of transcription at base pair resolution, Michael Guertin, a former graduate student in Lis lab performed Precise nuclear Run-On and

sequencing assay (PRO-seq) in *Drosophila* S2 cells at non-heat shock condition, 30 seconds, 2 minutes, 5 minites, 10 minutes and 20 minutes after heat shock. PRO-seq is a global run-on assay developed by Hojoong Kwak, which can map the newly transcribed RNA, and transcriptionally-engaged RNA polymerase II (Pol II) at base-pair resolution (Kwak *et al.,* 2013). By mapping the newly transcribed RNA at different time points before and after heat shock, we are able to have a progressive view of how each gene is regulated by the heat shock induction. Fabiana Duarte, a current graduate student in Lis lab analyzed all the PRO-seq data generated by Michael Guertin and identified groups of genes whose transcription are activated, unaffected and depressed upon heat shock. Comparison between insulator occupation information collected by ChIP-chip or ChIP-seq and the promoters regions of repressed genes showed a high correlation. Fabiana Duarte found that insulator proteins are enriched in the promoter regions of the repressed genes, suggesting that insulators may be involved in the repression of these genes (unpublished data). This is possibly a new potential function of *Drosophila* insulators that has not been characterized previously. However, the mechanism behind this is still not clear.

1.3.3 Insulators may protect the paused Pol II in a subset of GAF-bound paused genes from GAF knock-down effects.

It was known that Pol II can pause at the promoter-proximal regions after initiation in many genes in *Drosophila* and the paused Pol II possibly provides fast regulation of Pol II transcription reacting to regulating signals (Core *et al*., 2012). Previous studies have shown that GAF enriches on promoters with the paused Pol II.

Nicholas Fuda, a former post-doctoral fellow in Lis lab studied how GAGA factor (GAF) regulate Pol II transcription and Pol II promoter-proximal pausing in *Drosophila* and he assayed the levels of transcriptionally-engaged Pol II using PRO-seq in control and GAF-RNAi *Drosophila* S2 cells (Fuda *et al.,* 2015). He found that the promoter-proximal polymerase was significantly reduced on a large subset of paused promoters where GAF occupancy was reduced by knock down. Surprisingly, some promoters, which lost the GAF occupation, are still able to maintain the paused Pol II. In addition, the insulator proteins BEAF and Chromator are strikingly enriched on those GAF-associated genes where pausing is not affected by knock down. Since GAF was also proposed to be acting as an insulator protein in many other studies (Negre *et al*., 2010), it is possible that other insulator proteins compensated the loss of GAF and somehow kept the paused Pol II at those promoters. Further experiments and techniques are required to fully understand whether these insulators are critical for this phenomenon and how they are able to regulate promoter-proximal paused Pol II.

1.4 Inhibitory RNA aptamers provide tools for specific interruption of insulators to dissect primary effects *in vivo*

Molecular mechanisms behind insulators are poorly understood. Because insulator function is genomic context dependent, making conclusions based on transgenic assays risky. Experiments examining endogenous insulators are needed to dissect detailed insulator mechanisms. Traditional disrupting methods are limited in studying insulator functions. Except for Su(Hw), null mutations of all the other insulator proteins in *Drosophila* result in lethality (Gurudatta and Corces, 2009).

RNAi knock-downs (10-fold reduction) failed to deplete insulator proteins from high affinity sites, which includes BEAF32 at scs' (Latulippe *et al*., 2013). This is possibly the reason why Steven did not observe any change of nucleosome loss in either Zw5-RNAi or BEAF32-RNAi cell lines (Petesch and Lis, 2008). Moreover, it is uncertain whether the consequences we observe are secondary effects due to the disruption of mRNA and genomic DNA. The time needed for RNAi effects to take place and permanent loss of the genes also can introduce compensatory mechanisms, which may mask the primary molecular mechanisms. Lis lab has been developing and experimenting Inhibitory RNA aptamers, as a new and powerful tool for biology research. Inhibitory RNA aptamers are RNA molecules that form intricate secondary structures and bind with high affinity and specificity to a variety of targets (Ellington and Szostak, 1990). When expressed *in vivo* in a temporally and spatially controlled manner, they can be delivered rapidly and specifically to their targets and block specific interactions, providing a powerful tool to dissect insulator functions (Shi *et al.,* 1999). For insulator studies, aptamers are especially useful because interactions between insulator components are complex and aptamers are able to target one specific interacting surface without disrupting others (Shi *et al*., 2007). Our lab has successfully generated dozens of high affinity inhibitory RNA aptamers (Shi *et al.*, 1997; Fan *et al.*, 2004; Huang *et al.,* 2007; Sevilimedu *et al.*, 2008; Pagano *et al.,* 2014) and developed *in vivo* aptamers expression systems in *Drosophila* (Shi *et al.,* 1999; Salamanca *et al.,* 2011). RNA aptamers can be generated through an iterative selecting process called SELEX (Systematic Evolution of Ligands by Exponential Enrichment) (Tuerk and Gold, 1990). During this process, iterative cycles of *in vitro* selection and

amplification allow rare high affinity RNAs to be enriched from large, sequence-diverse RNA libraries. In collaboration with Craighead's lab in the School of Applied and Engineering Physics at Cornell, we have developed a multiplexed micro-column based SELEX device (Latulippe *et al.,* 2013), microplate-based SELEX device (Szeto *et al.,* 2014) and a less time-consuming and effective selection method called RAPID (RNA APtamer Isolation via Dual-cycles), which systematically skips any unnecessary amplification steps (Szeto *et al*., 2013). Together with our well-characterized parameters (Latulippe *et al.,* 2013; Shi *et al.,* 2002; Ozer *et al.,* 2013; Ozer et al., 2014), generation of aptamers can be done very efficiently and effectively.

1.5 Thesis outline

Described in this thesis is the generation and characterization of inhibitory RNA aptamers for multiple *D. melanogaster* insulators for studies on heat shock genes. By having stringent selections and thorough detecting assays, inhibitory RNA with high affinity and specificity are generated to dissect the primary functions and molecular mechanisms of insulators in the well-characterized heat shock system.

Chapter 2 details the materials and methods employed for the research in this dissertation.

Chapter3 describes the generation of the insulator aptamers using newly developed multiplex aptamer selection device and high-efficiency protocol, RAPID.

Chapter 4 presents the sequencing results of the enriched insulator apatmers and selection of top aptamer candidates. Background Binding Sequences (BBS) are identified during the high-throughput aptamer selection and sequencing.

In Chapter 5, several assays characterizing the property and binding efficiency of the top aptamer candidates are described.

In Appendix A the generation of antibodies against several insulator proteins is discussed.

Finally, in Appendix B, protocols for RNA labeling with fluorescent dyes are described.

CHAPTER 2

MATERIALS AND METHODS

2.1 Insulator protein purification

To prepare for the selection, I've made GST tagged, His tagged and MBP tagged full length or truncated protein constructs containing known domains for protein purification. Full-length and certain identified domains on eight insulator proteins were included. These eight proteins are Chromator, CP190, BEAF32, Su(Hw), Mod(mdg4), dCTCF, GAF and Zw5. Several different protein purification protocols or systems were tried as well, including BL21 expression, Arctic expression and Balculovirus expression. Unfortunately the yield of the insulator proteins is very low in most conditions. In conclusion, MBP tagged protein expressed in BL21 gives the best yield for most insulator proteins.

Most of the proteins used in RAPID, shown in Table 2.1, have very high purity. Gel filtration Superdex75 10/300GL has been used to eliminate degraded chromo domain components from the protein prep, but it didn't improve the purity much.

Table 2.1 Vectors and expression systems used to purify the insulator proteins and affinity tags used in RAPID.

	Insulator protein target	Vector	Expression system
1	GST-CP190 zinc finger domain	pGST-Parallel 1	BL21
2	GST-Mod(Mdg4)	pGST-Parallel 1	BL21
3	MBP-chromo domain	pMBP-Parallel 1	BL21
4	MBP-Chromator	pMBP-Parallel 1	BL21
5	MBP-CP190	pMBP-Parallel 1	BL21
6	MBP-BEAF32B	pMBP-Parallel 1	BL21
7	His-CP190	pFastBac-HTa	Baculovirus/Sf9 cells
8	His-Su(Hw)		Gift from Pamela Geyer from U of Iowa
9	His-CP190 (Denatured)	pFastBac-HTa	Baculovirus/Sf9 cells
10	GST tag	pGST-Parallel 1	BL21
11	MBP tag	pMBP-Parallel 1	BL21
12	His tag		Gift from Kylan Szeto

2.2 Baculovirus protein preparation

Bac-to-Bac Baculovirus Expression System (Invitrogen Life Science Technologies) is used for Baculovirus protein preparation in Sf9 cells. Protocols of the vector construction and protein isolation are described in the Invitrogen Bac-to-Bac Baculovirus Expression System user guide, which can be found online. I strictly followed Invitrogen's user guide. pFastBac-HTa vector is used and it added a His tag at the N terminus of the protein. The His tag was kept for resin binding in the aptamer selection process.

2.3 RAPID selection

Details of RAPID selection procedures are described in Szeto *et al.,* 2013 and the selection device is previously described in Latulippe *et al.,* 2013. In the insulator RAPID selection, 10uM micro-columns were used. The immobilized protein concentration is 0.1ug/ul. The RNA library was injected at 10ul/min and the total binding step lasts 100 minutes for each cycle. Two cycles take about one day.

The enriched RNA pools were reverse-transcribed, barcoded and gel purified for high-throughput sequencing. Hi-seq 2000 (Illumina) platform was used for sequencing.

2.4 Background binding sequence identification and specificity tests

Background sequences are identified by comparing the specificity of each sequence among the nine selection pools. Table 2.2 shows the top 15 BBS. Sequence. No. 1 is BBS1 and sequence No. 2 is BBS2 described in chapter 4.

In the MEDUSA test, 50pM of BBS1 and 50pM of BBS2 are mixed into the N70 library and the library was subjected to a one cycle selection to empty column, affinity tags, His-GFP, His-NELF-E, GST-hHSF1 and His-UBLCP1 on the MEDUSA device. Details of this test and the device are available in Szeto *et al.,* 2014.

Table 2.2 Background binding Sequences (BBS)

No.	Sequence
1	ATTCCGCAGGGCTAGCCGCATGCTCAGGCCTGGCGGGTAGGGAGTTAGGGTAGGGAGACCAGGAGAGCTGGC
2	ATTCCGAAGCTCGTGACGGTACCTCCTAAAATGTCCATGGGGAAGGGAGGGAATGGGAAGGACAATCGGACACCG
3	ATTCGCACCAAGTCCTGGATTGACCCGGAAGTCCAACCAGGGAGGGGTAGGGTTAGGGTGAGTGAGGACCCGGCG
4	ATTCCGTACCACCTCTGGACCCACGTCATGATCTCGCTTAGAAGGCCGGGAGAAGGGGAGGAGGCGGGCGTGCC
5	ATTCACCCACCTGTGGAGCTGGCCCTAGGGAAGGGTAAGGGCGTGCAGACGTGCACCATGGGAAGGGGCCCGTA
6	ATTCACTCCCAGGTCAGGCCTGCGAGGAGCGTCTGTGAACTACAATGGAGTCGATAACGCGACAGTCGCTCGC
7	ATTCGTCTCGCTCGGTTAGAAGACCGGGGAGGGAGGGGAAGAACTTGCGCCACGTACGCAATTCGGGGAAAGTC
8	ATTCACAACCTTTGAATATATGACAAGTGACGGCGTTCTGGGAGTGGGAAGTGGGAAGCCGGGGTGCACGCCG
9	ATTCGCATACCGCCGCGAATAAGATCGTCCAGCCACGAATGGAGAACGCCGGGAGGGCAGAGGAAGGGTGTGCG
10	ATTCCTGTCGAACGTGAACCCCACGGGCACGTTAGGGAAAAGGGAATCGGGAAGGGAACACGTATCGTGGGGA
11	ATTCACGAAGAGCGCTCTGCAATGTATTTCGTAACCCAACCGGGAAGGGGAATGGGTAGAGGGCGTGTTGGG
12	ATTCAGAGGCAAGCCGACCTGATATTGAGATGTCAGCTGGCAGTGACCCGGGGAAAGGGCAAGGGAAAGGGAGCG
13	ATTCCGCCAAGCATCCTGCAGATGACTGGGAACAGTGGCGTGCGATATGACCACTTGGGAGGGAATGGGAAAG
14	ATTCTAGGAGCAACCCACTCAAGTACAGGAACAGAGGGATAGGGACGGGATGTGTGGCCAACCACACCGGGATG
15	ATTCGCGTGAACCCGTGCATAAAAAGCTCATGGGACATCAGGGATGGGAATAGGGGAGCTGTAATTAGCCCGG

2.5 RNA labeling with fluorescent tags

The RNAs were 3' end fluorescently labeled following protocols from John Pagano (Pagano *et al.,* 2011). I used fluorescein 5-thiosemicarbazide to label the 3'-end of the RNA and Alexa 647 to label the 5'-end of the RNA. Both of the protocols are included in Appendix B.

2.6 F-EMSA

F-EMSA was performed with protocols from John Pagano and the details about the assays are previously described in Latulippe *et al.,* 2013. Images were acquired at the fluorescein scan settings on a Typhoon 9400 imager (GE Healthcare Life Sciences).

2.7 Fluorescence Polarization

The same RNA and protein mixture that used for F-EMSA was scanned under the Synergy H1Hybrid Multi-Mode Microplate Reader (BioTek) to acquire the fluorescence polarization data. The polarization data was fit to the Hill equation by use of Igor (Wavemetrics) to estimate the equilibrium dissociation constant (K_D).

2.8 HiTS-RAP for insulator proteins

mOrange tagged insulator proteins are made using the mOranged fused vectors constructed by Jacob Tome. HiTS-RAP was run following protocols described in Tome *et al.,* 2014.

2.9 Cloning of the aptamer candidates

For each RNA aptamer, two single stranded DNA molecules that have 20nt overlap in the random region are designed and the double stranded DNA template is produced by primer extension using Klenow fragment (3'-5' exo-) as followed protocol. Single stranded DNA molecules were synthesized by Integrated DNA Technologies (IDT).

Table 2.3 Primer extension reaction

Component	Stock	Volume
Oligo 1	10uM	2ul
Oligo 2	10uM	2ul
NEB buffer 2	10x	2ul
dNTPs	2mM each	0.3ul
Klenow fragment (3'-5' exo-)	5000units/ml	0.2ul
ddH2O		13.5ul
Total		20ul

The oligonucleotides and water were mixed and heat to 90C for 2 minutes. Cool the sample at room temperature to denature and anneal the strands for extension. Add buffer and dNTPs, mix and add enzyme lastly. Mix the mixture gently and incubate at 37C for 10 minutes to 1 hour. Add 2ul of the primer extension product to 100ul of PCR reaction.

2.10 Alkaline hydrolysis assay

Alkaline hydrolysis is modified from Robert Batey *et al* 2001. The *in vitro* transcribed RNAs were fluorescently labeled on the 5' end with Alexa 647 or on the 3' end with fluorescein in separate reactions (Pagano *et al.,* 2011). Two end-labeling

products are combined together for each candidate. The labeled RNAs (50nmole) were uniformly cleaved in the alkaline hydrolysis buffer (50mM NaHCO3/Na2CO3 (pH 9.2), 1mM DETA) and heated to 95 °C for 0, 2, 5, 8 and 15 minutes. The variously cleaved RNA pools were combined and ethanol precipitated. The cleaved RNA pool was incubated with MBP-BEAF32 (100nM) at room temperature for 1 hour. The mixture was then subjected to amylose bead (NEB) binding for 1 hour at room temperature. The amylose beads were centrifuged and washed twice with MBP protein binding buffer (50mM Tris pH 8.8, 20mM NaCl, 2mM DTT) and eluted with 10mM maltose in the binding buffer. RNase T1 (Life Technologies) digested RNA was used as RNA ladder to show the guanine position in the RNA backbone. Untreated RNA is also used as a control.

2.11 RNA sequencing gel

The large RNA sequencing gel was made using the reagents and protocols from Sequagel Ureagel System (National diagnostics). Different concentrations can be made for different needs. The sequencing gel I have made was 12% acrylamide gel. The glass plates were washed once using soap, two times with ddH_2O and two times with alcohol. After the glass plates were completely dried, the Gel Repel liquid solution from Sigma-Aldaich was spread on both of the glass plates. Before pouring the gel, the glass plates were aligned and completely sealed with Gel Sealing Tape (C.B.S. Scientific, 3.8cm x65.8cm, GBGT-72-15). Other tapes have been tested and this gel sealing tape is the only tape that completely seals the glass plates. The gel mixture was made on the ice to prevent it from polymerization. As soon as the gel

mixture was mixed, it was poured from the edge of the top space between the two glass plates slowly with the plated slightly tilted. If there are bubbles within the plates, tap the plates to exclude the bubbles. After the plates were filled, a compatible comb was added into the space. Keep the gel tilted until it is fully polymerized. If the gel is not used immediately, it can be kept in room temperature with for a few weeks after wetting with by water and sealing with a plastic warp (Saran Classic Wrap for food was used). In my experiment, all the RNA samples recovered from the alkaline hydrolysis and affinity binding were loaded on the gel, and it was run at 2500V for 3 to 4 hours.

2.12 Insulator antibody production

MBP-tagged or GST-tagged proteins were purified from E.coli and sent for antibody production in Pocono Rabbit Farm. 1mg of each protein is used for each animal. I used the 91-day antibody production protocol. Pre-serum, 2 small bleeds, 2 large bleeds and a final bleed were collected, which is about 100ml serum in total from a rabbit and about 50ml serum in total from a guinea pig. The serum aliquots are stored in the -80 freezers. For the antibody test, S2 cell lysate is produced and run on a SDS-PAGE gel. The antibodies were diluted by 1:500, 1:1000, 1:2000 and 1:5000 for efficiency tests on the western blot. HRP conjugated secondary antibodies were used to visualize the gels. Pre-serum was used as the control.

CHAPTER 3

SELECTION AND ENRICHMENT OF RNA APTAMERS FOR D. MELANOGASTER INSULATOR PROTEINS

3.1 Introduction

Insulators have been found in many eukaryotes, from yeast to human (Gurudatta and Corces, 2009). In vertebrates, the most widely studied insulator protein is CCCTC-binding factor (CTCF). Another nuclear protein, Cohesin, associates with CTCF and it is required for CTCF function. CTCF is possibly dependent on Cohesin's capacity to form insulator-insulator interaction and distal chromatin association (Zuin *et al*., 2014). RNA polymerase III is also found to act as both heterochromatin barriers and chromatin insulators in S. cerevisiae (D'Ambrosio et al., 2008) and S. pombe (Iwasaki *et al.,* 2010; Iwasaki *et al*., 2012). Different from other eukaryotes, many insulator proteins have been identified based on the canonical insulator transgenic assays. Validated *Drosophila insulator proteins include* Su (Hw), dCTCF, BEAF, GAF, Zw5, CP190 and Mod (mdg4) (Gurudatta and Corces, 2009). Multiple insulator proteins and correspondent DNA binding elements form insulator complexes in *Drosophila*. Among the complex, one or several insulator protein interact with the DNA binding element(s) and other insulator proteins form interactions within insulator proteins. Interactions between insulators in different complexes bridge multiple insulators together. Noticeably, CP190 and Mod(mdg4)2.2 are found in most of the insulators and they both have a BTB/POZ domain which can dimerize with the same domains. Another nuclear protein, Chromator (also named Chriz) has been

included in most of the *Drosophila* insulator studies recently because of its similarity to the validated insulator proteins in the Hi-C results (Sexton *et al*., 2012). Chromator was shown to bind the putative spindle matrix protein skeletor and plays a role in proper spindle dynamics during mitosis (Yao *et al.,* 2014). In a recent publication (Bohla et al., 2014), Dorte Bohla et al. performed an RNAi screening for insulator proteins using the well-studied Fab-8 insulator of the bithorax locus in *Drosophila*. They demonstrated that all four components of the NURF complex and several subunits of the dREAM complex are required for the insulator function. More and more factors are being related to insulator functions. They may be core components of the insulator complexes or some factors that are directed by insulators to help insulating functions. However, because of the complex nature of the *Drosophila* insulators, the study of the insulator function is very difficult. Here I choose to develop inhibitory RNA aptamers to specifically disrupt single or multiple interactions among single insulator or between insulators, which allows us to examine primary disruption effects *in vivo*.

Inhibitory RNA aptamers are RNA molecules that form intricate secondary structures and bind with high affinity and specificity to a variety of targets. When *expressed in vivo* in a temporally and spatially controlled manner, they can be delivered rapidly and specifically to their targets and block specific interactions. Traditionally, RNA aptamers are generated by incubating a sequence-diverse ligand pool, in this case synthesized random RNA pool, iteratively to the target molecules, in this case the insulator proteins. This repetitive target binding selection process called SELEX (Systematic Evolution of Ligands by Exponential enrichment) enriches the

RNA aptamer candidates with high affinity (Tuerk and Gold, 1990). Several different methods have been used for the *in vitro* selection processes, including nitrocellulose membrane, magnetic bead, affinity chromatography and capillary electrophoresis based selection methods (Song *et al*., 2012). In the selection, RNAs that are bound to the target are retained, reverse transcribed and amplified during each selection round, which is very time-consuming. In collaboration with Harold Craighead's lab in Cornell, Lis lab has developed multiplexed microcolumn-based selection platform that allows both serial and parallel selection of many targets or ligands (Latulippe *et al.,* 2013). Up to 10 selections can be performed in parallel in the first version of multiplexed device, and up to 96 selections can be performed at the same time in the microplate-based multiplexed device, which was named as MEDUSA (Microplate-based Enrichment De- vice Used for the Selection of Aptamers) (Szeto *et al*., 2014). I used the first version to select RNA aptamers for the insulator proteins and used MEDUSA to validate some of the enriched candidates (discussed in Chapter 4). In addition, we also creatively modified the traditional selection protocol to a more efficient and affective version, RAPID (RNA Aptamer Isolation via Dual-cycles), where we skip pool amplification step in all the even selection rounds (Szeto *et al.,* 2013). In other words, two consecutive selections are done before the pool is amplified. The consecutive selection makes the RNA screening very stringent and timesaving.

This chapter describes the multiplexed RAPID selection done for 9 different insulator proteins/domains.

3.2 Results

Full-length and identified domains on eight insulator proteins, 28 constructs in total, were all included in the selection targets of interests. However, insulator protein purification doesn't give soluble proteins all the time. After exploring different affinity tags, purification systems and protocols (discussed in Chapter 2), I was able to obtain 8 different proteins/domains (Table 3.1). I also denatured His-CP190 protein by incubating it at 95 °C for 5 minutes as the ninth target in the selection, for possible usage of RNA aptamers in denaturing experiments such as western Blot.

RAPID selection was performed as described in the RAPID method paper (Szeto *et al.,* 2013). Each protein was immobilized to the agarose resins through the fused affinity tags and the selections were done in microcolumns. The RNA library generate by Abdulla Ozer contains synthesized random RNA molecules with 70 nucleotides variable region flanked by two constant regions (120nt total length) with a complexity on the order of 10^{15} unique sequences (called N70 library). Parameters of this particular RAPID experiment are described in Chapter 2. Three rounds of RAPID were done for 9 targets. Two cycles of selections were performed consecutively in each RAPID round, while the RNA pool amplification was only done after cycle 2, 4 and 6. In order to eliminate the possible affinity tag (GST-tag, His-tag and MBP-tag) binding RNA aptamers, but not over-eliminate them, I also included negative screening of the affinity tag in cycle 1, 3 and 5. A selection column that contains the corresponding affinity tag was attached serially ahead of the positive selection column of the insulator proteins. For the first selection, all 9 columns and 3 negative columns are connected serially starting from No. 1 to No. 9 as listed in table 3.1. After the first cycle, all the following selections were done in parallel.

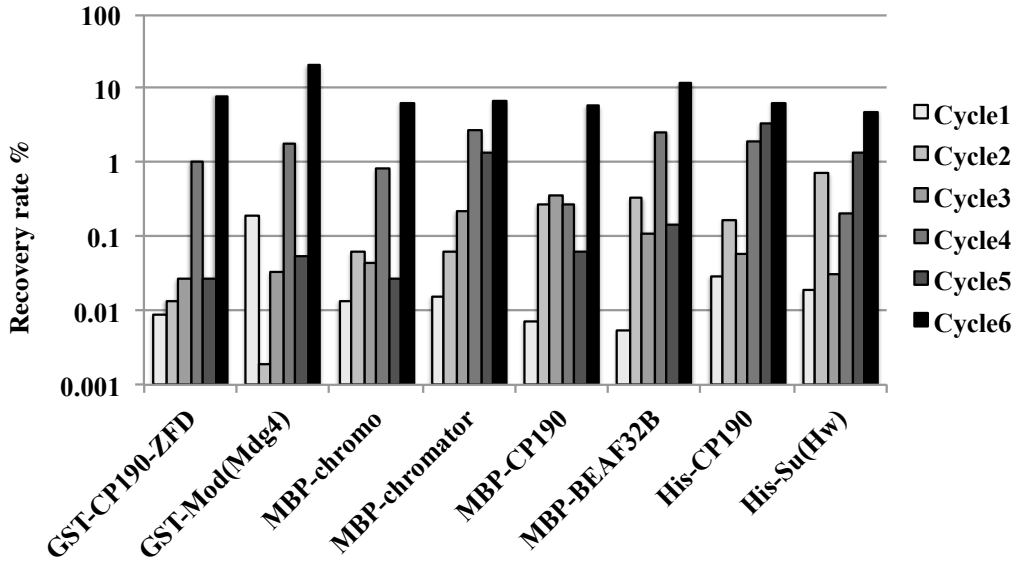
To monitor the enrichment and recovery rate of each selection cycle, qPCR was done using Roche qPCR reagents. And the recovery rate, which is the recovered RNA comparing the RNA pool injected before each selection cycle, is calculated. Although the recovery rate cannot directly tell us whether the selection is giving us the candidates we are interested, it can provide us a very straightforward view about whether the selection is retaining high-affinity candidates and excluding low affinity sequences. As shown in Figure 3.1, the recovery rate in each cycle is increasing along with the selection, from 0.01% to around 10% in all pools (Figure 3.1), indicates that the RAPID selection is enriching candidates with a higher affinity on average.

Table 3.1 Insulators proteins and affinity tags used in RAPID.

No.	Insulator protein target	Mw (kDalton)
1	GST tag	35.76
2	MBP tag	42.5
3	His tag	1.828
4	GST-CP190 zinc finger domain	35.76
5	GST-Mod(Mdg4)	94.48
6	MBP-chromo domain	51.35
7	MBP-Chromator	145
8	MBP-CP190	166.95
9	MBP-BEAF32B	76.94
10	His-CP190	126.68
11	His-Su(Hw)	>109.19 *
12	His-CP190 (Denatured)	126.68

* His-Su(Hw) is a gift from Dr. Pamela Geyer from University of Iowa. The size of the protein is estimated from the construct and the migration of the band on the SDS-PAGE gel.

Figure 3.1 Recovery rates of the RNA pools throughout the RAPID selection. The recovery rate is calculated as follow: Recovery rate=RNA recovered from cycle N/RNA injected into cycle N*100%. Recovery rate for all six cycles for each target are plotted together. From the left to right are cycle1 to cycle6. His-CP190 (denatured) is not shown here, but it looks very similar to His-CP190 selection.



3.3 Discussion and Perspectives

Traditionally, 10 to 20 rounds of traditional SELEX will be done before testing the aptamer candidates to ensure that the high affinity bound RNAs are very enriched in the pool. So many rounds are needed, because of the low efficiency of the traditional SELEX, also because of the slow way of obtaining the enriched sequences in the pool (discussed in Chapter4). Although the big trend of recovery rate was increasing along with the selection cycles, cycle 1, 3 and 5 are increasing slower comparing to cycle 2, 4 and 6 separately. This is expected because the even number cycles are selections that were done without amplifying the RNA pools. In this situation, the RNA that started with in the even number cycles are total enriched RNA pools directly recovered from the previous cycle. Without the amplification, the low affinity RNAs are not amplified, and they should be eliminated more efficiently comparing to the high affinity RNAs, which makes the selection more stringent as we expected. In addition, unlike traditional SELEX, nine selections were done in parallel with the RAPID protocol, which dramatically decreased the processing time and the lab work. Three rounds of RAPID uses approximately the same time that is needed for three rounds of traditional SELEX, while the recovery percentage is comparable to five to six rounds of traditional SELEX.

One more parameter that I changed in the insulator protein RAPID is the immobilized protein concentration on the agarose resins. Previous multiplexed SELEX performed in the same platform by other colleagues in Lis lab set the immobilized protein concentrations at 1ug/ul and they have shown that such concentration works very well for RNA aptamer enrichment (Szeto *et al.,* 2013). Here

in this thesis, I have used 0.1ug/ul instead of ug/ul. The basis for such adjustment is that the RNA candidates with higher affinity should be mostly enriched when the immobilized concentration is low. If the immobilized protein concentration is set at a certain level, then theoretically only RNAs with comparable or lower K_D will be retained on the resins. However, the microcolumn based selection platform is not recapitulating what is happening in the real cell environment, since the protein is immobilized. In addition, the surface of the resins, columns and other instruments used to process the RNA and proteins may as well be potential selection targets or cause the lost of the samples. So it is still hard to make a conclusion that a lower concentration of an immobilized target protein makes the selection more efficient or not. Further tests should be done to address this hypothesis.

CHAPTER 4

HIGH-THOUGHPUT SEQUENCING RESULTS PROVIDE EXTENSIVE INFORMATION OF RNA APTAMER CANDIDATES AND BACKGROUND BINDING SEQUENCES

4.1 Introduction

The next step after enriching the RNA aptamers is obtaining the sequence information from the pools. Traditionally, 10 to 20 rounds of SELEX will be done before the sequence obtaining process, because the sequence is separated using PCR amplification and it will be cloned into a vector and get amplified and purified in E. coli cells (Shi *et al.,* 1997). In that way, only limited amount of sequences can be tested due to the time-consuming process and the pool has to be highly enriched to ensure that the PCR amplified sequences can really represent the enriched RNA sequences. The emerging of the high-throughput sequencing techniques gave us the possibility of accessing every single sequence in the enriched RNA pools. Such information is very valuable because with all the enriched pools from every cycle sequenced, we are able to monitor how each sequence is enriched or eliminated during the selections and we can pick any of the sequence of interest for binding assays to test its affinity. The high-throughput sequence can allow us to interpret a lot of information from a single aptamer selection. Furthermore, when multiple selections are done in parallel, comparisons between all the selections can provide more information that cannot be read from a single selection. First of all, by comparing the sequences between different selections, I am able to identify the RNA sequences that

bind to the device or resins instead of the protein itself. Secondly, since I have included proteins with three different affinity tags, by looking across the similar sequences enriched in selections with the same affinity tags, I may be able to find RNA aptamers that associate with affinity tags. By doing this, we won't waste the time to test such sequences for its affinity to the insulator protein portion.

I selected the enriched RNA pools for all the proteins after cycle 2, 4 and 6 for high-throughput sequencing (27 pools in total). I used Illumina Hi-seq service provided by the Cornell sequencing center. This chapter presents the sequencing results and the background binding sequences identified through the multiplexed RAPID.

4.2 Results

The Illumina Hi-seq 2000 sequencing platform gave rise to 2.3 million to 4.6 million reads in each sequenced pool. Taking the sequencing error into account, sequences with low multiplicity and very high similarity to high multiplicity sequences might be the high multiplicity sequences with sequencing errors. Thus, we decided to cluster the sequences with 85% sequence similarity and the top sequence (sequence with the highest multiplicity) was used to represent the entire cluster. Collin Waters, a former technician in Lis lab performed data filtering and sequence clustering and returned top 3000 clusters or sequences (referred as sequences in later text) for each pool. In each pool, sequences are ranked by the multiplicity in the pool. Based on the sequencing results, RNA pools after 2nd cycle are not significant enriched and they are comparable to the N70 library (sequenced by Abdullah Ozer).

Significant enrichment of RNA can be seen since cycle 4. Top sequences enriched in cycle 6 are mostly enriched as well in cycle 4. Such sequencing dataset provides us a guide of choosing RNA candidates for future characterization. Sequences with higher ranking may have higher affinity to the target proteins.

Before I progressed to characterize the candidates, I compared the ranked sequences in all the enriched pools and noticed that there are about 15 sequences enriched with high multiplicity in six or more selections (Table 4.1). Surprisingly, there are two sequenced enriched in all nine selections with ranking ranging from 1 to 3. Several possible explanations were proposed for this situation. These two sequences may be sequences that bind to all the insulator proteins that I used as targets. Another possibility is that these sequences bind to the resins or the device; In other words, these sequences may be background binding sequences (BBS). The possibility of these sequences being highly associated with the affinity tags can be eliminated since the nine targets were fused with three different affinity tags.

Since my colleagues in Lis lab have performed many other SELEX for different protein targets with exactly the same multiplexed device and resins, I can make use of their datasets and explore whether these two sequences are enriched in their pools as well. Datasets that I have compared with include some pausing factor SELEX performed by John Pagano, Heat shock factor SELEX performed by Fabiana Duarte and histone related protein SELEX performed by Kylan Szeto. Noticeably, Kylan Szeto has performed SELEX for empty columns as well. These two sequences have shown up in almost all the other SELEX pools that have been done independently, including the empty column selections. The ranking in those pools range from 1 to 610.

Although they are not as dominant as my selections in those pools, the rankings are still high enough to conclude that these sequences are enriched in most of those pools.

To test whether these two sequences are really Background Binding Sequences (BBS), they were included in the first experiment in our newly developed 96 well plate based MEDUSA selection device. In this experiment, empty columns, several nuclear proteins and all three affinity tags (GST, MBP and His) are included as targets, and the two potential BBS are spiked into the N70 library. One cycle of selection was performed and the recovery rate is quantified by qPCR. As we expected, BBS2 has enriched in all the selections at 10 fold or more, especially in the Nickel resins containing columns. This indicates that BBS2 has a very high affinity to the background, especially to the Nickel resins. BBS1 is also enriched in all the selections but the enrichment is not as significant as BBS2. More details about this experiment are described in Chapter 2 and in Szeto *et al.,* 2014.

Table 4.1 Ranking of the background binding sequences among all the insulator selections. The sequences are numbered from 1 to 15. The number in the table represents the ranking by multiplicity of each sequence in each selection. BBS1 is No. 1 sequence and BBS2 is No. 2 sequence. Slash means that the sequence is no in top 3000 sequences in that pool.

	Ranking based on multiplicity								
Sequence No.	GST-CP190ZFD	GST-Mod(mdg4)	MBP-chromo	MBP-chromator	MBP-CP190	MBP-BEAF32B	His-CP190	His-Su(Hw)	His-CP190DN
1	1	1	1	1	2	1	2	3	2
2	2	2	2	2	1	3	1	1	1
3	3	3	3	5	11	1550	7	18	5
4	4	36	6	3	3	2	3	2	3
5	5	211	11	102	15	137	4	23	8
6	6	4	5	36	60	/	817	/	/
7	7	12	10	10	18	9	13	56	10
8	9	5	7	85	5	/	14	146	11
9	11	22	4	618	29	1549	6	25	6
10	16	25	20	18	14	894	5	4	4
11	17	10	8	151	25	382	8	16	7
12	35	29	18	/	131	/	35	75	18
13	/	83	9	/	166	12	21	/	16
14	61	/	84	74	/	39	19	/	36
15	73	148	45	58	27	33	520	68	68

4.3 Discussion and Perspectives

High-throughput sequencing allowed us to interpret the selection results in depth. The finding of BBS was not totally unexpected based on Shi's past selections (Shi *et al*., 2002), and identifying them is a very important step in optimizing RNA aptamer selections. The device and resins are always present in the selection process and they inevitably become the potential targets for the RNA aptamers. With the known sequences of the BBS, they can be taken out from the sequencing dataset directly for any new selections. As mentioned above in the results, BBS1 and BBS2 are the top two sequences in almost every pool in my selections, but not that dominant in other selections. The possible explanation is that the immobilized protein concentrations on the resins in the insulator protein selections are about 10 times lower than the other selections. This change was designed to select for higher affinity aptamers and avoid non-specific binding aptamers. However, if there is less protein on the resins, or in the device, more surfaces of the resins and device are exposed for RNA aptamer binding. One thing that I also noticed from other selections is that if a protein target has very high affinity to RNAs, the aptamer candidates that associate tightly with the target always dominate the pool and the BBS is barely enriched in the pool. Thus, a dominance of BBS of these enriched pools may suggest that the sinsulator election didn't work very well. It could also be that the insulators don't have high affinity to RNAs. This also indicates that the low protein concentration that I started with did not necessarily improve the selections.

CHAPTER 5

AFFINITY ASSAYS AND BINDING-ELEMENT MAPPING OF THE INSULATOR APTAMER CANDIDATES

5.1 Introduction

From the high throughput sequencing datasets, I was able to identify which sequences are enriched the most in each pool. After eliminating the background binding sequences, theoretically, the rest of the sequences ranked high should be potential RNA aptamers that bind to the target proteins. However, experience from previous selections done in our lab suggests that the multiplicity is not necessarily correlated with the affinity (Pagano *et al.,* 2014). In addition, some sequences may be able to interact with the target protein, but that doesn't ensure that this aptamer will interrupt any enzymatic activities of this protein, or disrupt any interactions between this protein and other insulator components. Whether the aptamer candidate can non-specifically interact with other proteins is not known either. Thus, additional assays determining the affinity and specificity of aptamer candidate and functional assays are required to select the best inhibitory RNA aptamers that I need. In this chapter, several assays I performed are discussed.

5.2 Results

The two basic RNA-protein interaction assays that I used to test the binding affinity are Fluorescence Polarization (FP) and fluorescence electrophoretic mobility shift assays (F-EMSA) (Pagano *et al*., 2011). In both of the assays, the RNA is

fluorescently labeled at one end. Proteins with different concentrations are incubated with the fluorescently labeled RNA. FP takes advantage of the change in the tumbling properties of the labeled RNA upon binding to the target protein. A free RNA molecule in the solution tumbling rapidly will depolarize the emitted light, where as protein bound RNA is tumbling less and this reduces the depolarization of the emitted light. In the F-EMSA, RNA and protein mixture will be run on the gel, where free RNA migrates faster and protein bound RNA migrates slower and shows a shifted band on the gel. K_D (disacsociation constant), which equals the concentration of the target protein when half of the total RNA molecules are associated with the target protein, can be calculated based on FP and F-EMSA. Radioactive labeling can also be used in these assays, but the half-life of these reagents is much shorter than fluorescently labeled RNA. Once the RNA is labeled by fluorophores, it can be stored in the freezer for a few months and still have the integrity and signal.

However, the disadvantage of these two assays is that they can only test few candidates at once and the time it takes to perform the experiments is tremendous. To solve this problem, Abdullah Ozer and Jacob Tome from Lis lab have developed a High-Throughput Sequencing-RNA Affinity-Profiling (HiTS-RAP) assay (Tome *et al.,* 2014). It takes advantage of the programmable feature of the Illumina GAIIx platform to assay the RNA-protein binding affinity at a high-throughput scale. Millions of K_D for all the sequences in the enriched RNA pool can be generated in one week. Details about this assay are available in Jacob tome *el al*. 2014. I have followed all the processes and performed HiTS-RAP for several insulator proteins, but the sequencing results contain mostly sequencing primer dimers, so I was not able to get any K_D from

this assay. This is probably due to the low protein concentration I used for the selection, which made the pools not very highly enriched. Due to the failure of the HiTS-RAP assay, I decided to go back to the traditional F-EMSA and FP assays. Among all the insulator proteins, I choose to investigate BEAF32, Chromator and CP190 first due to the importance of these three proteins and the availability of the purified proteins.

Based on the multiplicity ranking and the specificity of the aptamer candidates, I selected 6 candidates in each of the enriched RNA pools from selections against Chromator, BEAF32 and CP190 (28 candiates in total shown in Table 5.1). These RNAs are *in vitro* transcribed, acrylamide gel purified and 3' end labeled by fluorescein. N70 library is also included as a control. Proteins from 0 to 2uM are incubated with 2nM fluorescently labeled RNA for two hours and then scanned under the plate reader for Fluorescence Polarization. The same samples are loaded in the electrophoretic mobility shift assay (Figure 5.1 and Figure 5.2). The gel shift assays for Chromator are shown in Figure 5.1 where 6 candidates and N70 library are tested in parallel. The N70 library is able to shift the RNA indicates that Chromator may have non-specific RNA binding capacity. Comparing to the N70 library control, the candidates shifted the RNA at a lower concentration, suggesting a lower K_D of the top candidates. Noticeably, there are two shifted bands in some samples. One of these bands may be presenting complexes containing more than one RNA molecule and one protein molecule. BEAF32 protein also associated with the N70 library, with a higher K_D comparing to Chromator-RNA interaction (Figure 5.2). The BEAF32 aptamer candidates bind slightly tighter than the N70 library. Gel shift assays for both of the

proteins have very high background and the signals of the fluorescently labeled RNA are not prominent on the gel, so it is not reliable to quantify the K_D based on the gels. I've also used the samples for FP assay, thus I was able to calculate the K_D using the FP data. As shown in Table 5.2, both Chromator and BEAF32 aptamer candidates gave very low estimated K_D as expected. The Chromator candidates showed sub nano molar K_D. However, the SD for the Hill equation fitting is dramatically high. The K_D derived form FP data is reasonable, given that the RNA molecules are shifted at very low concentration in F-EMSA. Some shifted at the second or third data points above 0nM in Chromator gel shift assays. I believe that the high SD is caused by the lack of data points at low protein concentrations. Repetition of these assays using lower protein concentration may solve the high SD. Aptamer candidates for CP190 were cloned and fluorescently labeled, but the F-EMSA and FP were not performed yet.

Table 5.1 Aptamer candidate sequences selected for affinity tests. C1 to C6 are aptamers against Chromator. B1 to B6 are aptamers against BEAF32. P1 to P6 are aptamers against CP190.

Name	Sequence
C1	GCCGTAGATGTCGTCCACTGGATGAGGGACGCTCATGAACGATTC TGATCACGTTCTGAGCGAAGGGAGA
C2	CCACACCGGGACGGCATGCGGGGGGGAGTACTCGACAAGCACCG TGGAATGGGACGACGTGAGTCCACT
C3	CCACTGGAACCCCCGACGACGTTTGCAAAAAGTGGCCGCTTTGAA GAGGAAGGGATGGGAGGGCAAGCGG
C4	TTTGACAAGCGGCAGCCAGCGCACACGAGGAGGGACCGGTTACA GACCGACTATGGGATAGGGGGTGCGC
C5	TGGCCCGTGCTGGATCCAACGGACGGGGGACCGCAGCACTAGCAG TGCGCGAAGGGAGGTAAAGGGCCGA
C6	GCCCCGGGACGGATTCGCCGTAGTGTGCGAAACGATGTTCCCACA GGGTGGTATGAGAGGAGGGAGGC
B1	GGGATCCTGCGGGGATGCGTCGGTACTCCGTCCTTACACGGACCA GGAGCACGCATAAGGGTTTGGGAGC
B2	CATGACGTAGAACCTTAGAGTACGATCCTGGGTTGGGAGACCCTC ATCATCTGAGGGCGGGCCTCTGCAC
B3	ACCGGAGGATCCAACTGGAGGAGGGTAGGGCTGGTATAAATTGAT ACCAATTGCCCGGGGGTTGTAGGC
B4	GTAGAGTCGACGATCCGGGAAGGGGCGCCACTAATAAAGAGAGT GAGCGCGTGGGATAGCTCAGCAGCAC
B5	GGATACTTCGCGAGGACTCAAGTCAGTTCAACGGCGAACCGAATG GATCCGAGGGAGGGAATGGGCGC
B6	CCGTAGCTCTGCGAAATGGAACGTCTCCAGAGGAGGGAGGGAGAT AACACGACGGTGTTGCGGGCAGACG
P1	CGCGCATCTAAGGCGCCGTCGCAGTTGGGAAAATGGGTATTGGGC AAGGGGTCTGCGACGACGCGTGCGC
P2	ATCCACGGACGGAAGACATGTGTCGGCGACCTGGAAATGCAGTGA TGGGCAAGGATGGGTCGGGAGGGAA
P3	CGAGTAAATAACACTAACTGCGTAAACCGGGAAGGGAAAGGGAA TAAAGAGGCCCCTGCGGGAAGGGTGG
P4	CGTGCAGTGAGAAACCTGAGCCGTCCCGGGACGGGTCAGGGGTGG GAGAGGACGGCAAGCAGCTAGTGAC
P5	CCCAAGGTCGCCGGAACGCTTGACTCCAGCAGGGAATGGGATTAG GGATAAGGGAATGCTGGACCTAGGG
P6	ACACACCCCGGTGGATCCATAATTGAAACCTGCAGACTGAATCTG GGGAGGGAACAGGGATAGGGTTACA

Figure 5.1 F-EMSA of Chromator protein with Chromator aptamer candidates. Chromator proteins with concentration from 0uM to 2uM were incubated with 2nM 3' end fluorescently labeled RNA candidates for two hours. Seven binding assays were performed for the 6 candiates and N70 library. All the binding assays are separated on 10% Acrylamide gel. The bottom band is free RNA and the shifted bands that migrate slower are RNA molecules associated with Chromator.

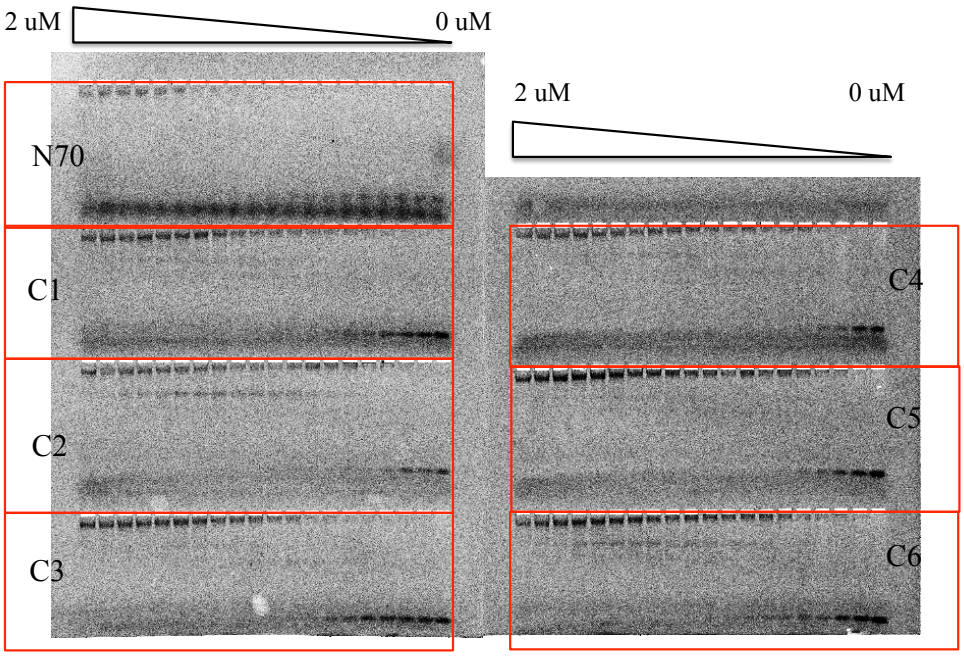


Figure 5.2 F-EMSA of BEAF32 protein with BEAF32 aptamer candidates.

BEAF32 proteins with concentration from 0uM to 2uM were incubated with 2nM 3' end fluorescently labeled RNA candidates for two hours. Six binding assays were performed for the 5 candiates and N70 library. All the binding assays are separated on 10% Acrylamide gel. The bottom band is free RNA and the shifted bands that migrate slower are RNA molecules associated with BEAF32.

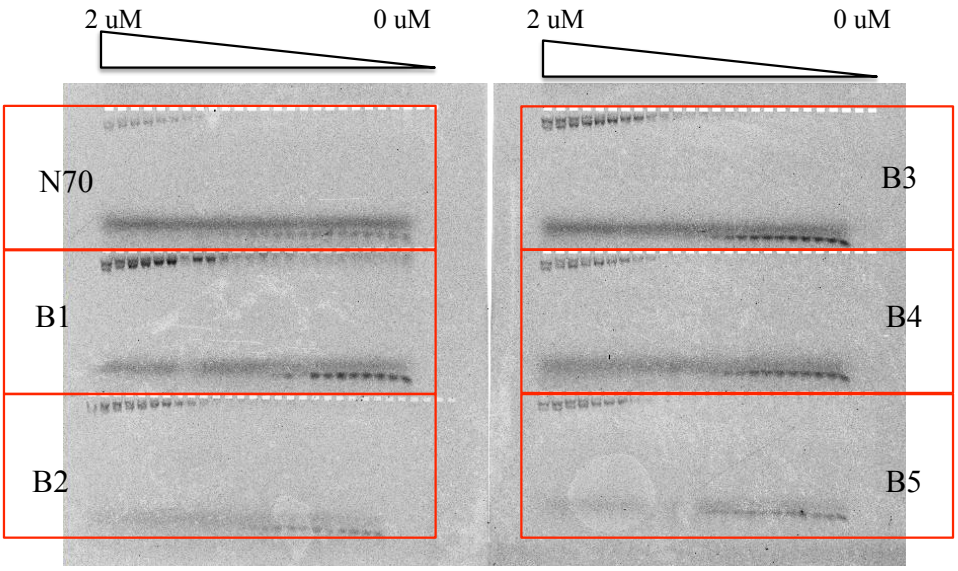


Table 5.2 Estimated K_D of the apatmer candidates. The disassociation constant (K_D) is estimated fitting the FP data to Hill equation. Standard Derivation (SD) is from the fitting.

Chromator			BEAF32		
RNA	K_D (nM)	SD(nM)	RNA	K_D (nM)	SD(nM)
N70	13.2	2.4	N70	93.8	86.1
Apt1	2.2	1.2	Apt1	12.3	6.0
Apt2	0.6	1890	Apt2	7.4	5.8
Apt3	5.5	0.4	Apt3	7.3	0.6
Apt4	0.5	83000	Apt4	2.0	1.0
Apt5	0.6	34700	Apt5	1.5	6.7
Apt6	0.6	61500			

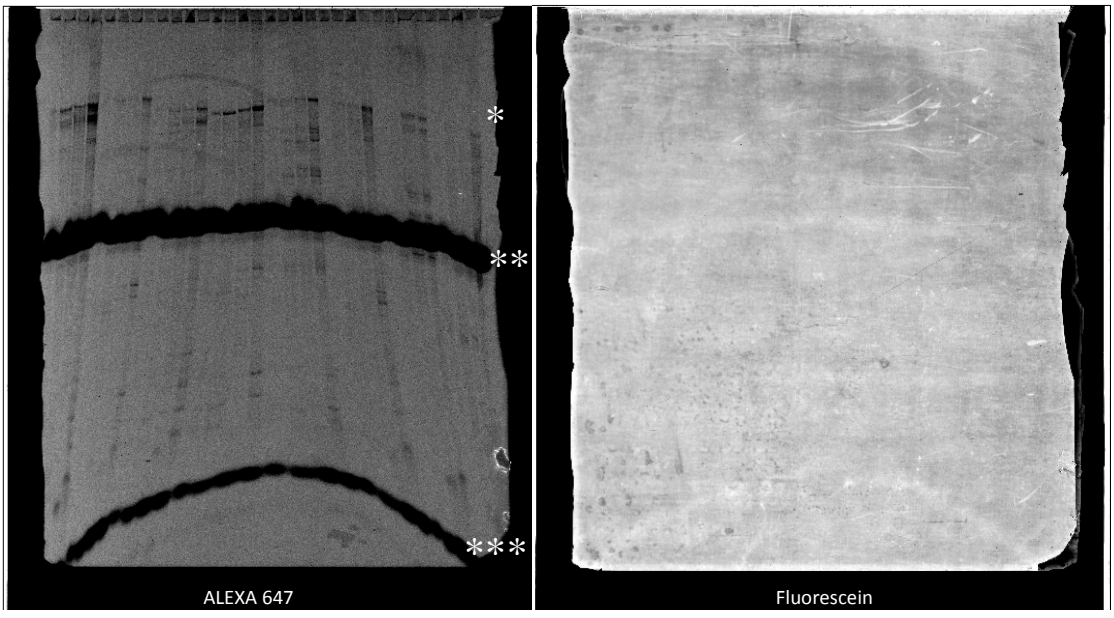
Apart from the affinity assays, I was also curious about which parts of the RNA aptamer are required to associate with the target protein. To investigate that, I firstly tried to search for RNA elements on the top candidates using MEME. No significant elements are enriched in CP190 and Chromator pools. In the BEAF pools, I was able to obtain a GGATCC element in 39 candidates of the top 80 candidates. However, when I closely look at the predicted secondary structure of these candidates, I found GATCC in the forward constant region, which forms complimentary strands with the GGATCC element within the random region. I do not exclude the possibility that the secondary structure formed between the constant region and the random region is required for the protein RNA-protein interaction.

Searching on MEME didn't give me much information about which region on the RNA are necessary for protein binding, so I researched about many other possible mapping assays. One powerful assay that I tried was the alkaline hydrolysis assay (Batey et al., 2001). Under alkaline condition, the 2' hydroxyl acts as a nucleophile and breaks the phosphate diester bound at random sites in the RNA. After the alkaline treatment, the single RNA sequence will become a RNA pool with various lengths. Such a pool will then be incubated with the target protein and the segments that bind to protein will be recovered and visualized on a base pair resolution sequencing gel. The aptamer candidate is either lableled at the 5' end with Alexa or at the 3' end by fluorescein to map from both ends. These two fluorophores were visualized from different detector channels.

In the first trial, six BEAF candidates and the two BBS were included. In Figure 5.3, the same sequencing gels with the enriched RNA segments are imaged to

show either Alexa 647 (on the left) or fluorescein (on the right). Unfortunately, the mapping didn't work successfully. On the Alexa gel, some bands can be seen but they are very unclear. The two thick dark bands in the middle and at the bottom of the gel are the loading dyes that I used to load the gel (Xylene cyanol and bromophenol blue). They are fluorescent under the Alexa 647 exciting channel. On the fluorescein gel, no signal is readable unless the gel is highly contrasted. To improve the assay, first of all, the loading dyes should be changed to some dyes that are not fluorescent in these two channels. Orange G, which runs lower than 10nt, can be tested for this purpose. In addition, the amount of RNA used in this assay should be boosted since the bands are barely visible on the gel, especially on the fluorescein channel.

Figure 5.3 Alkaline hydrolysis and affinity binding assay using the BEAF32 aptamer candidates and BEAF32 protein. Alkaline hydrolysis was done for the 6 BEAF32 aptamer candidates (B1-B6) and 2 background binding sequences (BBS1 and BBS2). The cleaved pools were bound to MBP-BEAF32 protein and recovered RNA is run on the sequencing gel. Left image is showing the Alexa 647 signal and the right image is showing the fluorescein signal. From right to left on the image are C1 to C6 and lastly the two BBS pools. Each candidate has four lanes on the gel. For each candidate, from right to left are untreated RNA, cleaved RNA before protein binding, recovered RNA from protein binding and RNase T1 ladder showing the G position on the RNA backbone. Full length RNAs and the loading dyes are labeled in the picture: * indicates full length RNAs, ** indicates xylene cyanol and ***indicates bromophenol blue.



5.3 Discussion and Perspectives

Our experience on aptamer selection and the new techniques and protocols that we developed have made the aptamer selection very straightforward and efficient. However, identifying the real useful inhibitory RNA aptamers for functional-disrupting experiments requires characterization of RNA aptamer affinities for target proteins. HiTS-RAP is a very powerful assay working for this need, but it may require the RNA pools to have high quality and be very enriched. The alkaline hydrolysis assay provides an efficient and accurate mapping technique for RNA candidates as well and the preliminary results look very promising. Future work on characterization of the enriched RNA pools should be done to continue this project.

APPENDIX A

GENERATION OF ANTIBODIES AGAINST INSULATOR PROTEINS

When the inhibitory RNA aptamers are generated for insulator disruption, insulator antibodies will be needed for detecting the positions of the insulator proteins. In addition, insulator antibodies are also required to investigate the interactions between different insulator components. For these reasons, I purified three proteins (MBP-CP190, GST-BEAF32 and MBP-Chromator) from E. coli for antibody production. The affinity tags are maintained in the proteins because they are immunogenic and they can promote the immune system to react to the insulator proteins. Future experiments using the antibodies will be mostly performed in *Drosophila* cells or animals, so antibodies generated against the affinity tags will not cause background problems.

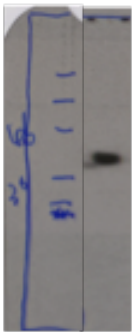
The proteins are purified and sent to Pocono Rabbit Farm, a company that provides antibody production service. BEAF32 antibodies were generated in two rabbits. Chromator antibodies are generated in two Guinea pigs. CP190 antibodies were generated in two Rabbits. A pre-immune sample, 2 small bleed, 2 large bleed and a final bleed were produced in each animal. I've received about 200ml serum of BEAF32 or CP190 immunized Rabbits and about 100ml of Chromator immunized Guinea pigs.

To test the specificity of the antibody, I ran S2 cell lysate on the SDS-PAGE gel and then stained the gel with the insulator antibodies. A Secondary antibody conjugated with HRP (Horseradish peroxidase) was then used for film exposing.

Dilutions from 1:500 to 1:5000 have been tested for each serum. Pre-serum was used as the control.

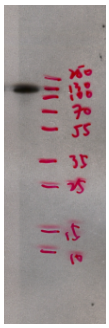
Figure A1 shows the western blot test of the 2nd bleed of BEAF32 antibody and 2nd bleed of Chromator antibody. Both of the antibodies stain a unique band on the gel, indicating a high specificity of the antibodies. There is no band on the control western blots (not shown). The first bleeds for BEAF32 have less antibodies comparing to the following bleeds, but a dilution at 1:1000 worked well on western blot (not shown). The first bleeds for Chromator showed very high background (not shown). Later bleeds of BEAF32 and Chromator were not tested. The serum generated for CP190 is not tested either. More tests will needed to access the specificity and quality of these insulator antibodies.

Figure A1 Western blot tests for insulator antibodies. The serum from the second bleed is tested here in Western blot. The left panel is the BEAF32 antibody and the unique band migrates as expected for BEAF32. The right Panel shows the Chromator band stained by the Chromator antibody and the unique band migrates as expected for Chromator. S2 whole cell lysate was collected and frozen at 100,000 cells/ul in the -80 freezer. 10 ul of the S2 whole cell lysate was loaded in each well in the Western blot.



BEAF32
42kDa

Rabbit
2nd bleed
1:2000



Chromator
110kDa
Shown as 130kDa on SDS-PAGE

Guinea Pig
2nd bleed
1:2000

APPENDIX B

PROTOCOLS FOR RNA LABELING WITH FLUORESCENT DYES

B.1 3'-end labeling of RNA with fluorescein 5-thiosemicarbazide

Reaction #1:

(0.5 nmoles RNA, 100 mM NaOAc, pH 5.1, 5 nmoles $NaIO_4$ in a 50 μL reaction)

(10 equivalents of $NaIO_4$ to 1 equivalent of RNA)

Example:

146 μM RNA oligo:	3.5 μL
0.5 M NaOAc, pH 5.1	10 μL
0.5 mM $NaIO_4$	10 μL
H_2O	26.5 μL
	50 μL

- React at room temperature (~25 deg. C) for 90 minutes or more.
- Add 1 μL pure RNase free glycogen (Invitrogen 20 μg/μL)
- Add 2.5 μL of NaCl ($1/20^{th}$ volume)
- Add 100 μL of ice cold 100% ethanol (2 volumes)
- Put in -20 deg. Freezer for 20 minutes.
- Spin at max speed in microcentrifuge for 10 minutes
- Dry by pipetting of residual ethanol. Save ethanol, the pellet is slippery.

Reaction #2:

(0.5 nmoles oxidized RNA, 100 mM NaOAc, pH 5.1, 1 mM fluorescein 5-thiosemicarbazide in a 50 μL reaction)

(100 equivalents of fluorescein 5-thiosemicarbazide to 1 equivalent of RNA)

- Dissolve pellet from Reaction #1 in 50 μL of 100 mM NaOAc, pH 5.1, 1 mM fluorescein 5'-thiosemicarbazide. Let it react at 4 deg. C in the dark overnight (may be able to get away with shorter time, I'm not sure)

Purification:

- Pack a Handee 1.5 mL spin column with pre-swollen G-25 or G-50 resin. (I usually prepare this as I'm preparing the RNA on the previous day, as it needs to swell overnight. I swell 4 grams of resin in 25 mL of TE, and this usually gives a 50% slurry of resin). To pack the column, add 2 mL of G-25 slurry to a column, and spin for two minutes at 1100xg. Discard the flow through, and wash the column with 50 uL of TE by adding it to the top and spinning the column for another 2 minutes at 1100xg.
- Apply 50 μL of reaction #2 (no more) to the top of the column.
- 1100 x G, 3 minutes.
- Remove flow through to a new tube, this contains your labeled RNA in TE. Unincorporated reactant (bright yellow) and other salts are left behind in the column.
- Measure the absorbance from 220 – 600 nm (1:5 dilution should work)
- The molar extinction coefficient of fluorescein at basic pH at 491 nM is 70,820, and 14,518 at 260. This changes drastically at acidic pH, so this is just an approximation.

- Determine the RNA concentration from absorbance measurement at 260 nm including extinction coefficient of RNA plus fluorescein. Determine the fluorescein concentration from the absorbance measurement at A491. The ratio of concentrations is a loose approximation of labeling efficiency. Determine purity on a gel. If unreacted label remains, run a second spin column. Typical efficiency is 70-90%.

Stock Recipes:

0.5 M NaOAc

Sodium Acetate (FW = 82.03 Da)	4.1 g

Add water to 80 mLs

Adjust pH with glacial acetic acid

Bring volume to 100 mL with water, filter sterilize

Store at room temperature

50 mM $NaIO_4$

Sodium periodate (FW 213.89 Da)	53 mg

Dissolve in 5 mL of water, filter sterilize

Store at room temperature

0.5 mM $NaIO_4$

50 mM Sodium Periodate	10 μL
H_2O	990 μL

200 mM fluorescein 5-thiosemicarbazide in DMF

Fluorescein 5-thiosemicarbazide (FW 421.43 Da)	100 mg (entire container)

Add 1.186 mLs of anhydrous DMF. Do not vortex or filter. Lightly stir until in solution.

Store in the dark, behind aluminum foil, at -20 deg. C

100 mM NaOAc, 1 mM fluorescein 5-thiosemicarbazide

0.5 M NaOAc, pH 5.1	80 μL
200 mM fluorescein 5-thiosemicarbazide in DMF	3ul
H_2O	318 μL

Make right before using. Discard when done.

Protocol adapted from Wu et al., Nucleic Acids Res., 1996, 24, 3472-3 and from Kruse et al., Biochem. J., 2000, 346, 107-115.

Based on the protocol of Reines and Cantor, 1974, 1, 767-786

Note: This protocol should work with any hydrazide, semicarbazide, or other aldehyde-reactive dyes. We've done biotinamidohexanoic acid hydrazide (Sigma), fluorescein 5-thio-semicarbazide (Invitrogen/Molecular Probes), Alexa Fluor 488 hydrazide, and Oregon Green cadaverine.

B.2 5'-end labeling of RNA with Alexa 647

Note that PNK kinase reaction is inhibited by ammonium ions. Therefore, ammonium acetate should not be used to ethanol precipitate the oligo nucleotide-molecular probes protocol

To ensure that the oligonucleotide is free of interfering compounds, especially amines, such as triethylamine or Tris and amoonium salts, we strongly recommend extracting and precipitating the sample prior to initiating the kinase reaction and labeling reaction.

1.1 Dissolve the oligonucleotide in 100ul dH2O and extract three times with an equal volume of chloroform.

1.2 Precipitate the oligonucleotide by adding one-tenth volume (10ul) of 3M NaCl and two and a half volumes (250ul) of cold absolute ethanol. Mix well and place at -20C for 30minutes.

1.3 Centrifuge the solution in a microcentrifuge at -12000xg for 30minutes.

1.4 Carefully remove the supernatant, rinse the pellet once or twice with cold 70% ethanol and dry. Resuspend and measure concentration

1. Resuspend non-phosphorylated oligos at 1 µg/µl for a 22-24-mer.
2. **Reaction #1**: Assemble phosphorylation reaction with ATPgS.

0.03 mM oligonucleotide	11 µl of 1 µg/µl
0.5 mM ATPgS	1 µl of 25 mM
1x buffer	5 µl of 10x
5 mM DTT	2.5 µl of 0.1 M

0.4 U/µl T4 PNK	2 µl of 10 U/µl
Ribonuclease inhibitor	2ul (40 unites, superase-in, ambion)
H_2O	to 50 µl

10x buffer for T4 PNK

	10 ml
700mM Tris-Cl pH 7.6	7 ml of 1M
100 mM MgCl2	1 ml of 1 M
H_2O	to 10 ml

3. 37°C overnight in an air incubator or pcr machine with a hot bonnet.
4. Add 1ul of pure RNase free glycogen (Invitrogen, 20ug/ul)
5. Add 2.5ul of NaCl (1/20th volume)
6. Add 100ul of 100% ethanol (2 volumes)
7. -20°C/20 minutes or more.
8. Spin, max, 4°C, 25 minutes. (10 minutes?)
9. Wash pellet with 1 ml 70% EtOH. Briefly air dry pellet.
10. Resuspend in 42.5 µl 25 mM HEPES pH 7.4.

11. **Reaction #2:** Immediately before use, prepare 10 mM stock of fluorescein iodoacetamide in DMSO making as little excess as possible. Discard unused portion.

Eg. use 0.65 mg fluorescein iodoacetamide in 125 µl DMSO

12. Add 7.5 µl of the 10 mM Alexa 647. (Molecular probes-10-20moles of reagent for each mole of protein) to the phosphorothioate oligo.
13. Mix well. Incubate 2-3 hours at RT in dark. (or overnight- molecular probes)
14. Purify the DNA away from untreated fluorescein by passing over a G-25 or G-50 spin column
15. Determine labeling efficiency by measuring absorbance at 260 (DNA) and 490 (Fluorescein) and calculating the concentration of each.

REFERENCES

Arensbergen, J. V., Steensel, B. V. and Bussemaker, H. J. (2014) In search of the determinants of enhancer-promoter interaction specificity. *Trends in cell biology* **24**, 695-702.

Ashburner, M. and Bonner, J. (1979) The induction of gene activity in Drosophila by heat shock. *Cell* **17**, 241-254.

Batey, R. T., Sagar, M. B. and Doudna, J. A. (2001) Structural and energetic analysis of RNA recognition by a university conserved protein from the signal recognition particle. *J. Mol. Biol.* **307**, 229-246 .

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y. and Dekker, J. (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58** (3), 268-276.

Benyajati, C. and Worcel, A. (1976) Isolation, characterization, and structure of the folded interphase genome of *Drosophila melanogaster*. *Cell* **9**, 393-407.

Blanton, J., Gaszner, M. and Schedl, P. (2003) Protein: protein interactions and the pairing of boundary elements *in vivo*. *Genes & development* **17**, 664-675.

Bohla, D., Heroild, M., Panzer, I., Buxa, M. K., Ali, T., Demmers, J., Kruger, M., Scharfe, M., Jarek, M., Bartkuhn, M. and Renkawitz, R. (2014) A functional Insulator Screen Identifies NURF and dDEAM components to be required for enhancer-blocking. *PLoS One* **9**.

Core, L. J., Waterfall, J. J., Gilchrist, D. A., Fargo, D. C., Kwak, H., Adelman, K. and Lis, J. T. (2012) Defining the status of RNA polymerase at promtoers. *Cell Rep*. **2**, 1025-1035

D'Ambrosio, C., Schmidt, C. K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K., and Uhlmann, F. (2008) Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes Dev*, **22**, 2215–2227.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Yin, Hu, M., Liu, J. R. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380.

Ellington, A. D. and Szostak, J. W. (1990) In vitro selectin of RNA molecules that bind specific ligands, *Nature* **346**, 818-22.

Fan, X., Shi, H., Adelman, K. and Lis, J. (2004) Probing TBP interactions in transcription initiation and reinitiation with RNA aptamers that act in distinct modes. *Proc Natl Acad Sci USA* **101**, 6931-6939.

Fuda, N. J., Guertin, M. J., Sharma, S., Danko, C. G., Martins, A. L., Siepel, A. and Lis, J. T. (2015) GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. *PLoS Genet.* **11**.

Fuda, N. J., Ardehali, M. B. and Lis, J.T. (2009) Defining mechanisms that regulate RNA polymerase transcription *in vivo*. *Nature* **461**, 186-190.

Gaszner, M., Vazquez, J. and Schedl, P. (1999) The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer–promoter interaction. *Genes & development* **13**, 2098-2107.

Gurudatta, B. V. and Corces, V. G. (2009) Chromatin insulators lessons from the fly. *Briefings in functional genomics and proteomics* **8**, 276-82

Guertin, M. J., Martins, A. L., Siepel, A. and Lis, J. T. (2012) Accurate prediction of inducible transcription factor binding intensities in vivo. *PLoS Genet*. **8**.

Guertin, M. J. and Lis, J. T. (2010) Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet*. **6**.

Hart, C., Zhao, K. and Laemmli, U. (1997) The scs' boundary element: characterization of boundary element-associated factors. *Molecular and cellular biology* **17**, 999-1009.

Ho, J. W. *et al*. (2014) Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449-452.

Hou, C., Li, L., Qin, Z. S. and Corces V. G. (2012) Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* **48**, 471-484.

Huang, Z., Pei, W., Jayaseelan, S., Shi, H. and Niu, L. (2007) RNA aptamers selected against the GluR2 glutamate receptor channel. *Biochemistry* **46**, 12648-12655.

Igo-Kemenes, T. and Zachau, J. (1978) Domains in chromatin structure. Cold *Spring Harbor Symp. Quant. Biol*. **42**, 109-118.

Iwasaki O, Tanaka A, Tanizawa H, Grewal SI, Noma K. (2010) Centromeric localization of dispersed Pol III genes in fission yeast. *Mol Biol Cell* **21**, 254–265.

Iwasaki O. and Noma K. (2012) Global genome organization mediated by RNA

polymerase III-transcribed genes in fission yeast. *Gene* **493**,195–200.

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C., Schmitt, A. D., Espinoza, C. A. and Ren, B. (2013) A high-resolution map of the three-dimensional chromatin interactiome in human cells. *Nature* **503**, 290-294.

Kellum, R. and Schedl, P. (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941-950

Kellum, R. and Schedl, P. (1992) A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell. Biol*. **12(5)**, 2424-2431.

Kwak, H., Fuda, N. J., Core, L. J. and Lis, J. T. (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339,** 950-953.

Latulippe, D., Szeto, K. and Ozer, A., Duarte F. M., Kelly. C. V., Pagano, J. M., White, B. S., Shalloway, D., Lis, J. T., and Craighead, H. G. (2013) Multiplexed microcolumn-based process for efficient selection of RNA aptamers. *Anal Chem*. **85**, 3417-3424

Lieberman-Aiden, E. et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293.

Negre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., Feng, X., Ahmad, K., Russel, S., White, R. A., Stein, L., Henikoff, S., Kellis, M. and White, K. P. (2010) A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet*. **6**.

Ozer, A., Pagano, J. M., Lis, T. M. (2014) New technologies provide quantum changes in the scale, speed, and success of SELEX methods and aptamer characterization. *Mol Ther Nucleic Acids* **5.**

Ozer, A., White, B., Lis, J. and Shalloway, D. (2013) Density-dependent cooperative non specific binding in solid-phase SELEX affinity selection. *Nucleic acids research* **41**, 7167-7175.

Pagano, J. M., Clingman, C. C. and Ryder, S. P. (2011) Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. *RNA* **17**, 14-20.

Pagano, J. M., Kwak, H., Waters, C. T., Sprouse, R. O., White, B. S., Ozer, A., Szeto, K., Shalloway, D., Craighead, H. G. and Lis, J. T. (2014)Defining NELF-E RNA binding in HIV-1 and promoter-proximal pause regions. *PLoS Genet.* **10**.

Petesch, S. J. and Lis, J. T. (2008) Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at *Hsp70* loci. *Cell* **134**, 74078.

Petesch, S. J. and Lis, J. T. (2012) Activator-induced spread of poly (ADP-ribose) polymerase promoters nucleosome loss at *Hsp70*. *Mol Cell*. **45**, 64-74.

Phillips-Cremins, J. E. and Corces, V. G. (2013) Chromatin insulators: linking genome organization to cellular function. *Molecular cell* **50**, 461-471.
Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Gavalli, G. (2012) Three-Dimensional Folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458-472

Salamanca, H., Fuda, N., Shi, H. and Lis, J. (2011) An RNA aptamer perturbs heat shock transcription factor activity in Drosophila melanogaster. *Nucleic acids research* **39**, 6729-6740.

Shi, H., Fan, X., Sevilimedu, A. and Lis, J. T. (2007) RNA aptamers directed to discrete functional sites on a single protein structural domain. *Proc Natl Acad Sci USA* **104**, 3472-3476.

Shi, H., Fan, X., Ni, Z and Lis, J. T. (2002) Evolutionary dynamics and population control during in vitro selection and amplification with multiple targets. *RNA* **8**, 1461-1470.

Shi, H., Hoffman, B. and Lis, J. T. (1997) A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. *Molecular and cellular biology* **17**, 2649-2657.

Shi, H., Hoffman, B. and Lis, J. T. (1999) RNA aptamers as effective protein antagonists in a multicellular organism. *Proc Natl Acad Sci USA* **96**, 10033-10038.

Song, K., Lee, S. and Ban, C. (2012) Aptamers and their biological applications. *Sensors* **12**, 612-631.

Szeto, K., Latulippe, D. R., Ozer, A., Pagano, J. M., White, B. S., Shalloway, D., Lis, J. T. and Craighead, H. G. (2013) RAPID-SELEX for RNA aptamers. *PLoS One* **8**.

Szeto, K., Reinholt, S. J., Duarte, F. M., Pagano, J. M., Ozer, A., Yao, L., Lis, J. T. and Craighead, H. G. (2014) High-throughput binding characterization of RNA aptamer selections using a microplate-based multiplex mucrocolumn device. *Anal Bioanal Chem*. **406**, 2727-2732.

Tome, J. M., Ozer, A., Pagano, J. M., Gheba, D., Schroth, G. P. and Lis, J. T. (2014) Comprehensive analysis of RNA-protein interactions by high-throughput sequencing-RNA affinity profiling. *Nat Methods* **11**, 683-688.

Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)* **249,** 505–10.

Udvardy, A. and Schedl, P. (1984) Chromatin organization of the 87A7 heat shock locus of Drosophila melanogaster. *J. Mol. Biol* **172**, 385-403.

Udvardy, A., Maine, E. and Schedl, P. (1985) Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J. Mol. Biol*. **185**, 341-358

Vogelmann, J., Le Gall, A., Dejardin, S., Allemand, F., Gamot, A., Labesse, G., Cuvier, O., Negre, N., Cohen-Gonsaud, M., Margeat, E. and Nollmann, M. (2014) Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the *Drosophila* genome. *PLoS Genet*. **10**.

Weisbrod, S. (1982) Active chromatin. *Nature* (London) **297**, 289-295.

Yao, C., Wang, C., Li, Y., Ding, Y., Rath, U., Sengupta, S., Girton, J., Jahansen, K. M. and Johansen, J. (2014) The spindle matrix protein, Chromator, is a novel tubulin binding protein that can interact with both microtubules and free tubulin. *PLoS One* **9**.

Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IJcken WF, Grosveld FG, Ren B, Wendt KS. (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A*, **111**, 996–1001.