SEARCHES FOR PROTEIN INTERACTIONS INVOLVED IN PRIMARY SEX DETERMINATION OF *Drosophila*

A Thesis

by

YI SUN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	James Erickson
Committee Members,	Bruce Riley
	Jerome Menet
	Jun-yuan Ji
Head of Department,	Thomas McKnight

December 2019

Major Subject: Biology

ABSTRACT

Sex determination in *Drosophila melanogaster* depends on the X chromosome number. Embryos with two dose of X chromosomes develop into females while embryos with one dose of X develop into males. Four X-linked signal elements (XSEs), *sisterless A* (*sisA*), *scute* (*sc*), unpaired (*upd*) and *runt* are responsible of conveying X dose to master regulatory gene, *Sex-lethal* (*Sxl*). As a result, two dose of X can activate the female specific establishment promotor, *SxlPe*, and set up an autoregulatory loop that maintain the continuous production of SXL while one dose of X lead to no *Sxl* activity and default male development. Hence, it is of great importance to understand how XSEs regulate *Sxl* activity.

The mechanism of how sisA, a strong XSE, regulate *SxlPe* activity is not known. Predicted to be a bZip transcription factor, SisA protein is believed to heterodimerize with a dimerization partner. To find out the potential partners, I generated transgenic flies carrying tagged-*sisA* constructs, which allow purification of SisA and pull down of SisA and potential partners from embryonic nuclear extracts using high-affinity antibodies against the tags. However, mass spectrometry revealed no candidates of significant abundance. Possible reasons and alternative approaches are discussed in this thesis. Meanwhile, transgenic flies carrying tagged-*sc* constructs were generated with similar approach to search for any protein interactors apart from Da that are involved in Sc's regulation of *Sxl* activity and genetic tests were performed on transgene lines

carrying Zelda binding site mutations to better understand the role of Zelda in regulation of *Sxl*.

DEDICATION

This dissertation is dedicated to my parents and friends, who have always believed and supported me throughout the years.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. James Erickson as I would not be where I am today without his guidance during my graduate studies. I would also like to thank my committee members, Dr. Jun-yuan Ji, Dr. Bruce Riley, Dr. Jerome Menet, and former member Dr. Arne Lekven, for their valuable suggestions and support throughout the course of this research. Thanks also go to my friends and former colleagues for making lab a wonderful place to work and graduate school a great experience. Finally, thanks to my mother and father for their encouragement, trust and unconditional love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor James Erickson, Bruce Riley and Jerome Menet of the Department of Biology and Professor Ji Jun-yuan from Texas A&M University Health Science Center.

The transgenic lines carrying TAGteam site mutations were designed and constructed by Jayashre Rajendren, a previous PhD student in our lab. All other work conducted for the thesis was completed by the student independently.

Funding Sources

This work was also made possible in part by NSF under Grant Number 1052310 and startup funding from Texas A&M University. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NSF and Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Cell fates and developmental choices are commonly defined by small concentration differences of regulatory molecules. In some cases, the decisions are made in early development during a narrow time window in response to subtle dose differences of regulatory proteins. It is, therefore, of great importance to understand the mechanism of how the regulators' doses are interpreted at the molecular level to precisely trigger the distinct cell fates.

Sexual fate in *Drosophila melanogaster* depends on the number of X chromosomes [1, 2]. A special promoter of *Sex lethal* (*Sxl*), the sex determination switch gene, senses the two-fold differences of X chromosomes via the collective dose of four X-linked signal elements (XSEs) [3-5]. Hence *Drosophila* somatic sex determination has been served as an ideal system to study how two-fold dose differences of regulatory proteins can set promoters into ON or OFF state and eventually lead to alternative cell fates, male or female.

Overview of *Drosophila* somatic sex determination and dosage compensation

The primary determinant of *Drosophila* somatic sex is the number of X chromosomes. *Sex-lethal* (*Sxl*), discovered as a direct target of X chromosome counting mechanism, acts as a master regulatory gene of sex determination pathway [3, 6]. In XX embryos,

Sxl is activated, which initiates female development. In XY embryos, *Sxl* remains to be inactive, leading to male development.

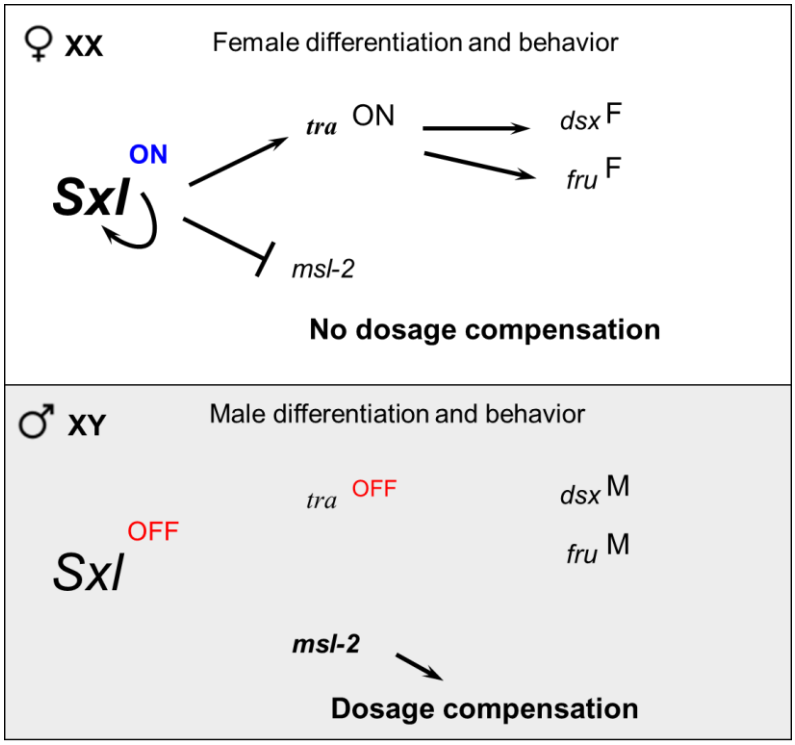

Figure 1. Overview of *Drosophila* somatic sex determination.
(Top Panel) In female XX embryos, SXL expression leads to the production of functional TRA protein and TRA directs the splicing of *dsx* and *fru* into female specific form. SXL also blocks *msl-2* translation so there is no dosage compensation. SXL is capable to maintain itself via an auto regulatory feedback loop. (Lower Panel) In male XY embryos, SXL is not produced so its downstream targets are spliced into male specific form and *msl-2* is translated, resulting in dosage compensation.

SXL protein, with its RNA binding property, is known to direct the sex-specific splicing of its downstream target gene, *transformer* (*tra*) [7-9]. In the absence of SXL in males,

tra transcript is spliced by default, resulting in a non-functional, truncated protein. This allows the transcripts of *doublesex* (*dsx*) and *fruitless* (*fru*), two downstream targets of TRA, to undergo a default splicing mechanism and produce male specific isoforms, DSX^M and FRU^M [10-12]. DSX^M works as a transcription factor to control male development and differentiation while FRU^M is involved in controlling male behavior and sexual orientation[10, 13]. In females, SXL protein directs female specific splicing of TRA by binding to its pre-mRNA, leading to the production of functional TRA protein. TRA, also an RNA binding protein, directs the alternative splicing of *dsx* and *fru* transcripts into female specific isoforms, DSX^F and FRU^F, which promote female differentiation and behavior [10]. Therefore, SXL lies on the top of somatic sex determination pathway through a cascade of alternative splicing events (Figure 1).

In addition, SXL exerts it effects on dosage compensation via another downstream target, *male-specific-lethal 2* (*msl2*) [14]. In flies, dosage compensation is achieved in males by upregulating the expression of the single X chromosome to equal the expression level of the two X chromosomes in females. This requires the assembly of dosage compensation complex (DCC), a ribonucleoprotein complex that is able to associate with male X chromosome and ultimately cause elevated transcription of male X. MSL2, a key component of DCC, is essential for dosage compensation to take place [15-17]. As active SXL protein inhibits the translation of *msl2* transcripts, it is ensured that no DCC is assembled and no dosage compensation occur in females [18-20].

The establishment and maintenance of *Sxl*

The establishment and maintenance of *Sxl* activity relies on a dual promoter system and an auto-regulatory feedback loop. *Sxl* has a female-specific, early acting establishment promoter *SxlPe* and a late acting maintenance promoter *SxlPm* that turns on in both sexes [21].

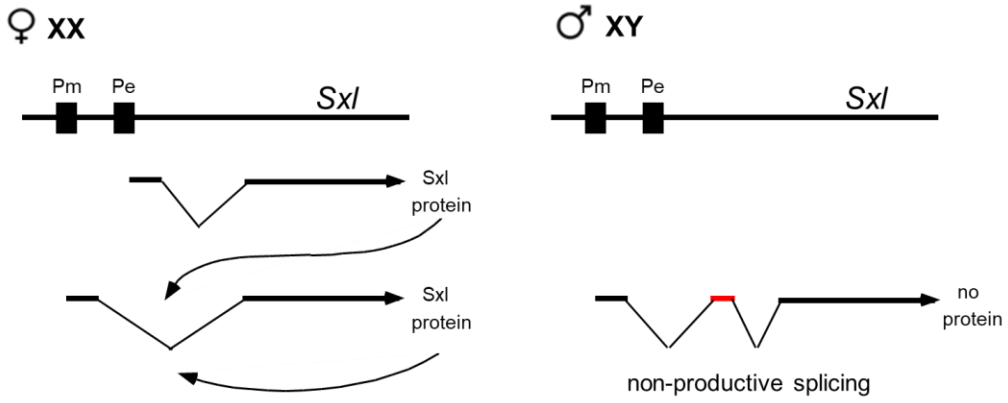

Figure 2. Establishment and maintenance of SXL via auto regulatory feedback loop.
In females, *SxlPe* is activated and produce an initial burst of SXL protein. This early SXL protein directs the splicing of mRNA from *SxlPm* to maintain the continuous production of functional SXL, setting up an auto regulatory loop. In males, *SxlPe* remains off, leading to no functional SXL production as *SxlPm* RNA is spliced by default and produces nonfunctional truncated protein.

When *SxlPm* is transcribed in males, the transcripts are spliced in a default mode, causing the mRNA to contain a premature stop codon within the male specific third exon and generate truncated, non-functional SXL [12, 22-24]. In females, however, *SxlPe*

responds to two dose of X chromosomes in early development, which in turn generates female-specific *Sxl* mRNA and initiates the production of functional early SXL protein [22-24]. SXL, as stated earlier, is an RNA binding protein and is capable of directing the splicing of its own transcripts. As a result, when *SxlPm* comes on later in females, the early SXL protein binds with *SxlPm* transcripts by recognizing the poly(U) rich sites flanking the male specific third exon, enabling the skipping of this exon and female specific splicing [24, 25]. The subsequent functional SXL protein then continues to regulate the splicing of *SxlPm* transcripts (Figure 2). Thus, even though *SxlPe* activity is transient and shuts off soon after *SxlPm* turns on, the brief burst of the initial SXL protein is sufficient to establish a positive autoregulatory splicing loop that let the females to maintain continuous generation of SXL in the rest of their life [6, 22, 23].

Expression and regulation of *SxlPe*

SxlPe expression starts very early during *Drosophila* female development, even before the occurrence of maternal to zygotic transition (MZT), while most of zygotic genome remains silenced. Nascent *SxlPe* transcripts can be detected by *in situ* hybridization in most nuclei at nuclear cycle 12. The expression continues to increase in cycle 13 and reaches the peak in early cycle 14, then its activity declines and completely shuts down by mid cycle 14 [2, 26].

How does X chromosome dose signal the early activation of *SxlPe* in females? The expression levels of four known X-linked signal elements (XSEs), *sisterless A* (*sisA*),

scute (*sc* or *sisB*), *runt* and unpaired (*upd* or *sisC*), are responsible for conveying the X dose and regulating *SxlPe* activity [26-31]. Among them, *sisA* and *sisB* are the two strongest XSEs. Both sisA and sisB are uniformly expressed in the early embryo and each is required for *SxlPe* activation. The other two XSEs, *runt* and *upd*, are weaker activators and exert less effects on *SxlPe* activity due in part to their later expression times and spatial limited expression patterns.

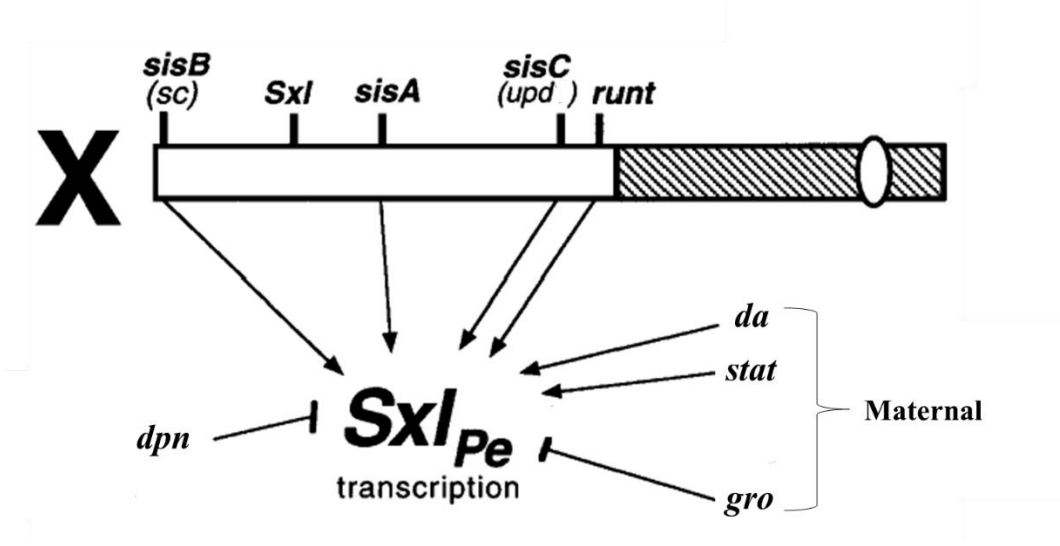

Figure 3. XSEs and other *SxlPe* regulators.
SisA, *sc*, *upd* and *runt* are the four XSEs that activate *SxlPe*. Maternal *da* and *stat* are positive regulators of *SxlPe* while *dpn* and *gro* act as repressors.

SisA was identified as a key XSE due to its sex-specific lethal characteristics. Mutations of *sisA* had female lethal effect and duplication of $sisA^{+}$ decreased male viability [26, 27, 32]. The sequence of *sisA* suggests that the gene is likely to encode a non-canonical

basic leucine zipper (b-zip) transcription factor. The structure prediction of SisA protein indicates that SisA is very likely to heterodimerize with a partner protein to carry out its function as *SxlPe* activator [26, 33]. However, little is known about the detailed mechanism of how SisA regulate *SxlPe* activity.

Sc, also known as *sisB*, is another XSE that is indispensable for the female-specific activation of *SxlPe*. Homozygous *sc* mutants are female-lethal due to their failure to activate *SxlPe* and combination of sc^+ and $sisA^+$ duplications were male-lethal because they mimic the female dose of X chromosomes [27]. *Sc* also shares a similar temporal expression pattern with sisA. Transcription of *sc* was first detected in nuclear cycle 9, reached peak in late cycle 12, experienced a sudden decrease and finally end in cycle 14 [2, 26]. Both the expression of *sisA* and *sc* appeared before the onset of *SxlPe* and peaked when *SxlPe* transcripts were first detected, befitting their role as essential XSE for *SxlPe* initiation. *Sc* encodes a basic helix-loop-helix (bHLH) transcription factor. Previous research revealed that Sc forms heterodimers with maternally deposited Daughtless (Da) and the heterodimer probably activate *SxlPe* via direct binding [34-36]. 11 SC/Da binding sites located within the 1.4kb region of *SxlPe* were identified, with 6 of them located in the proximal 400bp region, which was recently verified to be the sufficient unit to activate *SxlPe* [36] (Erickson, unpublished)*Runt*, a pair-rule gene, is well-known because of its role in the regulation of segmentation genes. As a DNA binding protein, RUNT also act as an XSE in sex determination. In *runt* null mutants, *SxlPe* expression was almost abolished in the central region of the embryo in cycles 13

and 14, suggesting *runt* is needed for maintaining the uniform expression of *SxlPe* [28]. Although it is initially believed that RUNT could activate *SxlPe* as a conventional heterodimeric activator, research in our lab strongly suggests that RUNT activated *SxlPe* by antagonizing co-repressor Groucho (Gro) via its C-terminal WRPW domain [37] (Mahadaveraju and Erickson in revision for G3).

Unlike the other three XSEs that encode transcription factors, *upd* encodes a ligand for the Jak-Stat pathway [29, 31]. Upd signals via Jak kinase, Hopscotch (Hop), which in turn phosphorylated and activated maternally supplied transcription factor Stat92E [38, 39]. Stat92E is believed to be able to function as direct activator and three Stat binding sites are identified in the 1.4kb region of *SxlPe* [38, 39]. Upd is shown only to have temporal and spatial limited effect on *SxlPe* activity. The expression of *upd* appears only in cycle 13, which is after the onset of *SxlPe*. Loss of *upd* or its downstream target, *hop* or *stat* only cause reduction of *SxlPe* expression in the central region of the embryo in cycle 14 [29, 31, 38]. Hence, it is believed that *upd*, along with *runt*, are not required for the initiation of *SxlPe* activity, but function to reinforce *SxlPe* activity and ensure uniform expression.

In addition to the four XSEs, the zinc finger transcription factor Zelda may also play an important role in the regulation of *SxlPe* activity. Maternally deposited Zelda works as a global zygotic genome activator by binding to TAGteam sequence motifs, which are abundant in the promotor regions of many early expressed genes [40-42]. A cluster of

four TAGteam sites were found within the 400bp enhancer region of *SxlPe* and mutating them resulted in reduced *lacZ* expression in embryos carrying 1.4kb *SxlPe-lacZ* transgene [42]. A great amount of research suggests that Zelda binding to TAGteam sites could increase chromatin accessibility, leading to the early transcription of those zygotic genes [43-45].

Negative regulators also work in conjunction with the above activators to better define the activation threshold of *SxlPe*, making sure that *SxlPe* remains to be off in males. One key repressor is Deadpan (Dpn), a bHLH transcription factor of Hairy-Enhancer of Split (HES) family. Transcription of *dpn* is detected as early as cycle 12 and reaches peak during cycle 13, with higher level in the central region comparing to the two poles [46, 47]. Mutation in *dpn* can induce ectopic expression of *SxlPe* in males, which is consistent with the role of *dpn* as a repressor [46-48]. Dpn has four binding sites on *SxlPe*, all located within the 400bp proximal enhancer region [48]. It is believed that once Dpn binds to those sites, it is able to recruit maternally deposited co-repressor Gro via its C-terminal WRPW motif to repress *SxlPe* activity [48-50]. Loss of maternal Gro result in ectopic *SxlPe* expression in males and early onset of *SxlPe* expression in females [48]. One possible model suggest that Gro is recruited by Dpn to ensure the repression of *SxlPe* in males while Runt functions to activate *SxlPe* by antagonizing Gro-mediated repression in females.

Hence, all these positive and negative regulators together build a complex regulatory network to precisely and accurately regulate the on and off state of *SxlPe*. However, there are still a lot of missing pieces, like what is the protein partner of SisA and how these two together regulate *SxlPe* activity; does Sc has protein interactors apart from Da that are involved in activating *SxlPe*; how does Zelda function as an activator in *SxlPe* scenario, etc. In this thesis, I presented the efforts and progress I made in answering the above questions.

CHAPTER II

MATERIALS AND METHODS

Generation of tagged *sisA* and *sc* transgenic lines

The tags used were eGFP-FlAsH-StrepII-TEV-3xFlag, RFP-3xHA and mCherry (simplified as eGFP, RFP and mCherry). The sequence of the plasmids used to amplify the tags can be found in the supplements of Venken et al [51]. These tags were introduced into P[acman] vectors carrying *sisA* or *sc* and flanking sequences via *galK* recombineering. The *galK* recombineering technique used was developed by Warming et al [52] and further standardized by Jayashre Rajendren from our lab. The protocol contained two rounds of selections. First, *E.coli galK* was introduced into the P[acman] in the specific region for tag insertion. Minimal medium containing galactose as carbon source were used to select for recombinants. *GalK* positive recombinants were confirmed as pink colonies when grown on MacConkey indicator plates. Then, PCR fragments containing the tags and 51 bp flanking sequences were introduced to replace *galK*. Minimum glycerol medium containing 2-deoxy-galactose was used to select for $galK^-$ recombinants and recombinants were confirmed as white colonies when grown on MacConkey indicator plates. The recombinant plasmids were then purified by miniprep and retransformed into epi300 cells. To increase the copy number of the plasmids, epi300 cells with the plasmids were induced for 5 hours at 37°C using 0.1% L-arabinose. Then recombinant plasmids were purified either by miniprep or using Invitrogen PureLink HiPure Plasmid Midiprep Kit. The sequence of the 3xFLAG tag was

GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTA CAAGATGACGATGACAAG. 3xFLAG was cloned into the N terminal of *sisA* in pBluescriptKS(+) vector via NcoI restriction site. All constructs were confirmed by sequencing.

In the first trial, eGFP, RFP and mCherry tags were introduced between the 101st and 102nd amino acid of SisA. In the second trial, eGFP, RFP, mCherry and 3xFLAG were introduced into the N'-terminal of SisA. For Sc, only eGFP constructs were successfully integrated into fly genome in the first trial. The insertion site was between the 95th and 96th amino acid of Sc. eGFP, RFP and mCherry constructs were all successfully introduced between the 85th and 86th amino acid of Sc in the second trial. The purified plasmids were sent to BestGene, Inc. or Rainbow transgenic flies, Inc. for transgenic injection services. The estimated integration site in the genome for the all SisA constructs was ZH-86Fb. The estimated integration site of Sc constructs was ZH-68E.

Genetic complementation tests

All genetic tests were carried out at 25°C on fly medium containing yeast, cornmeal and molasses.

To test if tagged-*sisA* or tagged-*sc* transgene (simplified as TG) still have wild type SisA or Sc function, complementation tests were performed to see the viability of *sisA*$^{-}$ or *sc*$^{-}$ flies carrying the corresponding transgenes. *+/Y; TG/+* flies were first crossed with *sisA*$^{-}$

*/F*M7 or *sc⁻/FM7* virgin females. Male progeny *sisA⁻/ Y; TG/+* or *sc⁻/ Y; TG/+* from the above cross were crossed again with *sisA⁻/FM7* or *sc⁻/FM7* virgins. The progeny of this second-generation cross was counted and *sisA⁻/ sisA⁻; TG/+* or *sc⁻/ sc⁻; TG/+* were compared to their sister siblings *sisA⁻/ FM7; TG/+* or *sc⁻/ FM7⁻; TG/+* to calculate the viability.

For TAGteam site mutation transgenes, they were brought into *Sxl⁻* background to see how the mutations of the TAGteam sites affect *SxlPe* activity. *+/Y; TG/+* flies were first crossed with *Sxl⁻/Binsincy* virgin females. Male progeny from this cross *Sxl⁻/Y; TG/+* were then crossed again with *Sxl⁻/Binsincy* virgins. The progeny of the cross was counted and the viability of *Sxl⁻/ Sxl⁻; TG/+*was calculated by comparing to *Sxl⁻ /Binsincy; TG/+* siblings.

In situ hybridization and antibody staining

In situ hybridization were performed on whole embryo collection of 3 hour and 30 minutes. Embryos were dechorionated in 50% bleach and then fixed in heptane/PBS-10% formaldehyde for 50 minutes. *In situ* hybridizations were as described in previous publications [26, 53].

Antibody staining were performed on whole embryo collection of 7 hours. Embryos were dechorionated in 50% bleach and then fixed in heptane/PEM-4% formaldehyde for 30 minutes. Anti-SisA antisera raised in guinea pig (produced by Cocalico Biologicals,

Inc. #GP15) was used as primary antibody at 1:200 to 1:2000 dilution. Anti-guinea pig IgG-AP was used as secondary antibody at 1:300 dilution. Detailed antibody staining protocol was described in previous publication [32].

Protein nuclear extraction

Protein nuclear extraction protocol was modified from Strubbe et al [54], Pazin [55] and Lewis et al [56]. 1-5 hours embryos were collected at 21˚ C and dechlorinated in 50% bleach. Embryos were manually homogenized in Buffer I (15mM HEPES, 10mM KCl, 5mM $MgCl_2$, 0.5mM EGTA, 0.1mM EDTA, 350mM sucrose, 1mM DTT, 1mM sodium metabisulfite, 0.2mM PMSF and 1mM benzamidine) using Wheaton Dounce homogenizer and then filtered through 70 μm cell strainer. Nuclei were pelleted by centrifuge. Supernatant was discarded and the loose white nuclei was resuspended in Lysis Buffer (15mM HEPES, 100mM KCl, 3mM $MgCl_2$, 0.1mM EDTA, 10% glycerol 1mM DTT, 1mM sodium metabisulfite, 0.2mM PMSF and 1mM benzamidine). One tenth volume of saturated $(NH_4)_2SO_4$ were added to the nuclei to extract proteins by rotating for 20 mins. Insoluble materials were discarded after centrifuge and extracted proteins in the supernatants were precipitate by adding saturated $(NH_4)_2SO_4$ to a final 60% saturation. Precipitated proteins were dissolved in PBS or TBS buffer with protease inhibitor (Thermo Scientific™ Pierce Protease Inhibitor Mini Tablets) and dialyzed against the same buffer twice. All the above steps were performed in 4℃ cold room and on ice to avoid protein contamination and degradation.

Immunoprecipitation (IP)

ChromoTek RFP-Trap® Magnetic Agarose beads were used for IP of RFP tagged SisA. The protocol can be found on the company website (https://www.chromotek.com/fileadmin/content/PDFs/Protocols/RFP-Trap_MA_Kit_manual_171004 .pdf). Anti-FLAG® M2 Magnetic Beads (Sigma M8823) were used for IP of 3xFLAG tagged SisA. The protocol can be found on sigma website (https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Bulletin /m8823bul.pdf). All IP steps were performed in 4°C cold room or on ice.

Protease inhibitor (Thermo Scientific™ Pierce Protease Inhibitor Mini Tablets) were added into all the buffers used in IP. 50 μg/ml ethidium bromide was added to wash buffer. Protein elution was conducted either by using 0.2M glycine pH2.5 with RFP-Trap beads/0.1M glycine pH3.0 with anti-FLAG beads or by boiling at 95°C for 3 minutes in 2xSDS sample buffer.

Immobilizing *E.coli* expressed 3xFLAG-SisA with anti-FLAG M2 magnetic beads

sisA with N' terminal 3xFLAG tag were cloned into PET28a and transformed into NovaBlue cells (Novagen). Overnight cultures were used to inoculate fresh LB culture containing 50 μg/ml kanamycin in a 1:20 proportion. IPTG was added to a final concentration of 0.5mM for induction for 4 hours at 37°C after OD reached 0.2-0.4. Induced NovaBlue cells were pelleted by centrifuge and then lyzed with buffer containing 100mM NaH_2PO_4. 10mM Tris HCl and 8M urea, pH 6.3 by incubating on

rotator for 1 hour at room temperature and followed by centrifuge. Supernatant was diluted with same volume of TBS. Refolding of the proteins were performed by dialysis against TBS with 2M urea, TBS with 1M urea and TBS with 0.5M urea at 4°C. Refolded proteins were immobilized with anti-FLAG beads by incubating on rotator overnight at 4°C.

CHAPTER III

IDENTIFICATION OF THE PROTEIN PARTNER OF SISA

The function of SisA in sex determination was first identified because of its sex-specific lethal effects. Genetic test showed that *sisA* mutants were female lethal and extra copies of *sisA*$^{+}$ reduced male viability [3, 26, 27, 32]. *SisA* is one of the earliest genes to be zygotically transcribed in Drosophila embryos. *In situ* hybridization showed that *sisA* transcripts can be detected as early as nuclear cycle 8, and that they are localized in or near the nuclei. The expression continued to increase and peaked around late cycle 12 or early cycle 13, displaying a uniform expression pattern at the embryo surface. Afterwards, *sisA* transcription decreased rapidly and from cycle 14 onwards, *sisA* mRNAs were restricted to the yolk [32] (Figure 4). This early expression pattern is consistent with the expected role of *sisA* in the initiation of *SxlPe* activity and its late expression in yolk is likely associated with midgut development. The regulation of *SxlPe* by *sisA* was further confirmed by experiments using embryos containing *SxlPe-lacZ* fusion constructs. Extra copies of *sisA* were able to induce ectopic expression of *SxlPe-lacZ* in males while *sisA* mutants showed decreased level of *SxlPe-lacZ* expression [26]. Apart from serving as a key XSE in somatic sex determination, recent work in Dr. Mark Van Doren's lab at Johns Hopkins University suggests that *sisA* is also expressed in the female germline later in development where it is suspected to be involved in regulating *SxlPe* activity in germline sex determination.

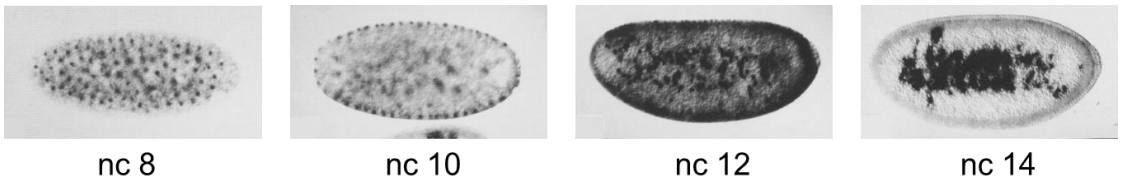

Figure 4. *SisA* expression in embryos.
Transcripts of *sisA* are first detected inclusively in nuclei at nuclear cycle 8. The expression continues to increase and reach peak around cycle 12 and become uniform in the embryo. By cycle 14, *sisA* expression greatly decrease and is restricted in the yolk. Reprinted from [32].

The attempt to decipher the structure of SisA protein is the first step to better understand how *sisA* works as a regulator of *SxlPe*. According to its sequence, SisA protein is predicted to be a non-canonical basic leucine zipper (bZip) transcription factor [26, 33, 57] (Figure 5). bZip proteins are known to have a basic DNA-binding domain followed by a leucine zipper dimerization domain, containing four to five heptad repeats of amino acids (designated as ‘a’, ‘b’, ‘c’, ‘d’, ‘e’, ‘f’ and ‘g’). Charged amino acids at ‘a’ position usually promotes heterodimerization and prevent homodimerization [33, 57]. As SisA contains charged amino acids, arginine and glutamic acid at the ‘a’ position of the second and third heptad, it is most likely that SisA needs dimerization partner to function as a bZip transcription factor. However, unlike the canonical bZip proteins, SisA does not have any attractive or repulsive g↔e’ interaction, suggesting it might have a novel partner [33].

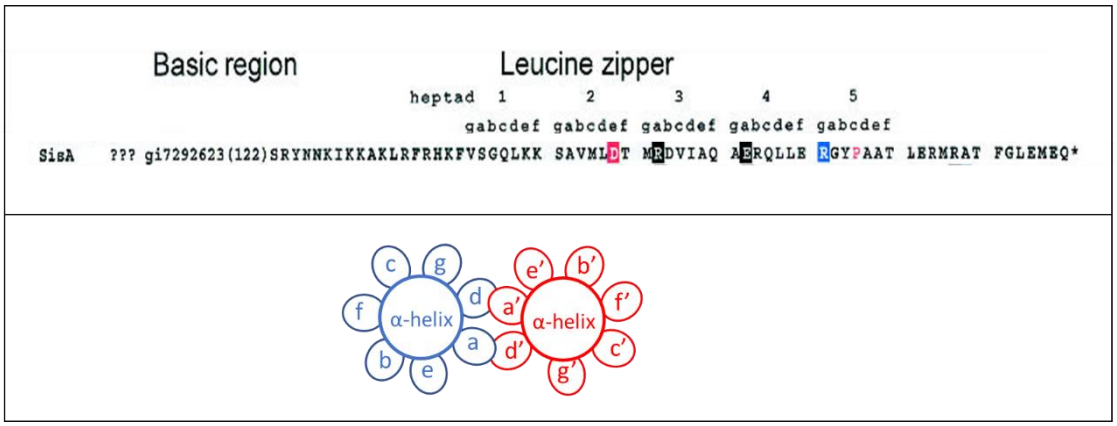

Figure 5. SisA heptad sequence and dimerization topology.
(Top Panel) SisA contains a basic DNA binding domain and a leucine zipper domain. (Lower Panel) Topology of bZip protein shows that g↔e' interaction and polarity or charge of amino acid at "a", "d" positions are crucial to predict dimerization properties of the protein.

Previous attempts to identify the dimerization partner of SisA

Sequence prediction of SisA indicates that it is highly likely to be insoluble according to C. Vinson group, whose research focuses on the gene regulation of bZip transcription factors. Experimental trails in our lab also showed SisA protein alone is highly insoluble when expressed in *E.coli*. Most important, neither refolded *E.coli*-expressed SisA and SisA protein translated *in vitro* from rabbit reticulocyte extracts were able to bind to *SxlPe* DNA in gel-shift assays (D. Yang and J. Erickson, unpublished). All these experiment results were consistent with the prediction that SisA protein cannot form homodimers and needs heterodimerization partner to function. The results do not, of course, rule out the possibility that SisA is not a transcription factor and that it regulates *SxlPe* activity with a unique mechanism that is independent of partner protein.

Yeast two hybrid screens(Y2H) are the most widely used method when it comes to the study of protein-protein interaction. Our lab has used Gal4 based Y2H assays to investigate the possible proteins that interact with SisA. Full length SisA and SisA leucine zipper domain were fused with the binding domain as baits and cDNA clones from 0-6 hours embryos were fused to the activation domain as preys. SisA appeared to be a sticky protein in Y2H assays and showed interactions with many proteins that were not recovered from controls. As bZip proteins presumably dimerize with bZip proteins, and none of the other proteins offered obvious promise, the candidates from Y2H was narrowed to three bZip proteins.

A previous Ph.D. student in our lab, Alejandra Gonzalez conducted detailed genetic experiments with the three bZip protein candidates, encoded by the *atf4*, *CG16813* and *CG16815* genes. Atf4, also known as CRC, is maternally deposited [58]. *CG16813* and *CG16815*, two genes located next to each other on the 2nd chromosome, were originally thought to be maternally deposited but proved to be early expressed zygotic genes with expression either before or during *SxlPe* activation [59]. All their temporal expression patterns meet the standard as being SisA partner in activating *SxlPe*. To eliminate maternally deposited ATF4, Alejandra Gonzalez generated germline clones with no atf4 and germline clones from hypomorphic *crc*^{1} mutant. Both showed no decrease in female viability, indicating that maternal Atf4 is not involved in sex determination. As *CG16813* and *CG16815* were closely linked, Alejandra Gonzalez generated double deletions of these two genes. Homozygous double deletions Δ (CG16815, CG16813)

were fully viable and fertile, a result inconsistent with the notion of either being the essential protein partner of SisA. When Δ (CG16815, CG16813) males were crossed with *sisA*^{1}/FM7 females (*sisA*^{1} is hypomorphic allele) to examine the zygotically effect in a sensitized genetic background there was no evidence for female-lethality. When *sisA*^{1}/+; Δ (CG16815, CG16813) females were crossed with *sisA*^{1}/Y; Δ (CG16815, CG16813) males to examine possible maternally effects, there was again no decrease in female viability. The above results imply that either none of those three genes are the real SisA partner *in vivo* or that redundancy exists and deleting these candidates alone (or in pairs for *CG16813* and *CG16815*) is not sufficient to sensitize female viability.

My Experimental approach--affinity purification of SisA from embryo extracts

Because the genetic tests of Y2H candidates failed to provide any evidence that they interact with SisA *in vivo*, a different approach was needed to identify possible SisA partners. The most straightforward approach is to biochemically purify SisA and its interacting protein, followed by identification using mass spectrometry.

The main obstacle of this approach is that SisA is expressed only for a short period in early embryos and that the protein is probably in low abundance given the relatively low *sisA* mRNA level. To overcome the difficulty, high quality antibody with high affinity that can efficiently recognize and enrich SisA and its partner is crucial to the success of the biochemical approach. Unfortunately, no good anti-SisA antibody is available on the market and none have been created in our lab's multiple attempts to generate them.

Therefore, my approach was to construct transgenic lines containing tagged SisA and use high-affinity antibodies against the tag to detect and purify SisA and its partner from *Drosophila* embryos. As SisA is most likely to act as a transcription factor, I used protein nuclear extract from early embryos to eliminate possible contaminating proteins and concentrate my target proteins. The identification of the purified proteins was performed using mass spectrometry (Mass spec), the most sensitive and effective method to study protein-protein interaction over the past decade. Apart from SisA, I have also employed same approach to construct transgenic lines of tagged Sc. Although Sc is better studied and is known to heterodimerize with maternal Da, this approach will help us to search for other protein interactors that are involved in regulating of *SxlPe* activity as well as confirming the interaction of Sc and Da *in vivo*. Experiments concerning with Sc will be discussed in Chapter IV.

Construction of tagged SisA transgenic lines

As stated above, high-affinity antibody is crucial to the purification of SisA given its low mRNA levels and transient expression. A specific kind of antibody called a nanobody, which is engineered from Camelidae, is devoid of light chain and its single domain possesses high binding affinity to antigens [60]. A great number of researchers have employed nanobodies to detect proteins of interest and to perform immunoprecipitation and ChIP experiment. Therefore, we decided to tag SisA with eGFP, RFP and mCherry tags, which are well studied and have commercially made nanobody products for detection.

As we wanted to construct transgenic lines with fully functional tagged SisA, it was important that the integration site of the tags should not disrupt SisA structure. My initial plan was to insert the tags into disordered region of the protein following the advice of Dr. Hugo Bellen, who stated at a *Drosophila* meeting that tags in disordered region were unlikely to adversely affect the structure and functions of the protein. Therefore, I used online computational program DisEMBL to predict intrinsically disordered regions of SisA using SisA sequences from *D. melanogaster*, *D. pseudoobscura* and *D. virilis*. One of the predicted disordered regions in all three species was in the middle of the protein but in front of leucine zipper domain. The amino acid sequences of this region are not conserved between different species and are 6 repeats of glycine-serine in *D. melanogaster* SisA, which had previously been considered to be a linker sequence to connect the N-terminal and C-terminal halves of the protein. [53]. As a result, we chose this region to insert the tags.

To construct tagged-SisA, I used *galK* recombineering technique to introduce the tags into the specific region of the SisA sequence in a P[acman] vector. The plasmid with P[acman] vector carrying tagged SisA and flanking sequence were then integrated into Drosophila genome via transgenic integration. The three tags I fused with SisA were eGFP-FlAsH-StrepII-TEV-3xFlag, RFP-3xHA and mCherry, which will be refered as eGFP, RFP, mCherry in the rest of the thesis.

Once the transgenic lines were obtained, they were tested for the ability to complement the null *sisA*^{5} mutation. *sisA*^{5} is not only defective in *SxlPe* activation, but also disrupt fly development, leaving both homozygous females and hemizygous males dead [32]. My genetic tests (Table 1) show that the tagged SisA transgene fully rescued hemizygous *sisA*^{5} males, showing those constructs provided the vital SisA function(s) needed in later in fly development. However, single copies of the tagged transgenes only partially complemented the sex determination function of sisA^5. The eGFP-tagged SisA barely rescued any homozygous *sisA*^{5} females (0.8%) while the RFP-tagged and mCherry-tagged versions were able to rescue around 30% of the females.

$disordered\ region\ TG$	eGFP tagged		RFP tagged		mCherry tagged		
Genotype	%	#	%	#	%	#	
$\frac{y\ w\ v\ sisA^5}{FM7}; \frac{+}{+}$	100	123	100	75	100	66	Reference
$\frac{y\ w\ v\ sisA^5}{y\ w\ v\ sisA^5}; \frac{+}{+}$	0	0	0	0	0	0	♀ Control
$\frac{y\ w\ v\ sisA^5}{y\ w\ v\ sisA^5}; \frac{sisA-TG}{+}$	0.8	1	29.3	22	30.3	20	♀ Experiment
$\frac{y\ w\ v\ sisA^5}{Y}; \frac{+}{+}$	0	0	0	0	0	0	♂ Control
$\frac{y\ w\ v\ sisA^5}{Y}; \frac{sisA-TG}{+}$	111.4	137	137.3	103	110.6	73	♂ Experiment

Table 1. Complementation of *sisA*^{5} mutant by single copy of tagged-*sisA* transgenes
Crosses set up at 25 ◦C. females $\frac{y\ w\ v\ sisA5}{FM7}$ × males $\frac{y\ w\ v\ sisA5}{Y}; \frac{sisA-TG}{+}$.

Since those transgenic lines with tags within the predicted disordered region failed to effectively complement wild type SisA function, it suggests that the presence of a disordered region at this location, rather than the specific sequence, serves an important purpose in the case of SisA function. Insertion of a folded peptide like eGFP, RFP and mCherry into the disordered region likely confers order or disrupts the region so that either sisA folding or perhaps its interactions with other proteins are disrupted. As disordered region-tagged *sisA* constructs did not meet our needs, I constructed new transgenic lines with the tags inserted at the N terminal of SisA. Our lab had previously generated N terminally tagged GST-*sisA* transgenic lines which were able to complement *sisA* null mutations. Although GST itself does not act as a good tag for my purification because its homozygous lethal probably due to its P-element insertion site, it provides a good insight of potential location to introduce the tag. Hence, I generated new constructs with eGFP, RFP and 3xFLAG at N-terminal of SisA via *galK* recombineering and the constructs were injected into flies. I chose to use 3xFLAG in addition to the fluorescent proteins, because this FLAG tag is relatively small, which might cause less interference with SisA structure and because there are a number of anti-FLAG antibodies on the market that have proven to be efficient for IP and ChIP experiments. Having a second tagged version also provides a means of screening potential false-positives interactors as the antibodies used to purify tagged sisA should interact with different sets of contaminating proteins.

N-tag-*sisA*	N-eGFP		N-RFP		N – 3xFLAG		
Genotype	%	#	%	#	%	#	
$\frac{y\ w\ v\ sisA^5}{FM7}; \frac{+}{+}$	100	97	100	87	100	65	Reference
$\frac{y\ w\ v\ sisA^5}{y\ w\ v\ sisA^5}; \frac{+}{+}$	0	0	0	0	0	0	♀ Control
$\frac{y\ w\ v\ sisA^5}{y\ w\ v\ sisA^5}; \frac{\mathrm{N-tag-}sisA}{+}$	10.3	10	114.9	101	120.0	78	♀ Experiment
$\frac{y\ w\ v\ sisA^5}{Y}; \frac{+}{+}$	0	0	0	0	0	0	♂ Control
$\frac{y\ w\ v\ sisA^5}{Y}; \frac{\mathrm{N-tag-}sisA}{+}$	114.4	111	139.1	121	152.3	99	♂ Experiment

Table 2. Complementation of *sisA*^{5} mutant by single copy of N-tag-*sisA* transgenes
Crosses set up at 25 ◦C. female $\frac{y\ w\ v\ sisA5}{FM7}$ × male $\frac{y\ w\ v\ sisA5}{Y}; \frac{N-tag-sisA}{+}$.

In comparison with the previous constructs, transgenic lines with N-tag-*sisA* display much higher female viability rates (Table 2). One copy of N-RFP-*sisA* or N-3xFLAG-*sisA* transgene is sufficient to fully rescue homozygous *sisA*^{5} females. These two transgenes were also able to fully rescue hypomorphic *sisA*^{1} mutant (not shown). *sisA*^{1} is probably a point mutation in the probable DNA-binding domain, which affect its function in sex determination and is homozygous lethal for females [26]. One odd finding is that eGFP tag has a lower ability to provide sisA function compared to the RFP and 3xFLAG tags. This was true for both the disordered region and N-terminally tagged constructs. I do not know the reason. One possibility is that the extra tandem tag, FlAsH-StrepII-TEV-3xFlag following eGFP negatively affects the proper folding of

SisA or otherwise interferes with sisA function. A second possibility is the eGFP, itself, perhaps due to slower folding, is more disruptive to SisA function than the RFP or 3X FLAG tags.

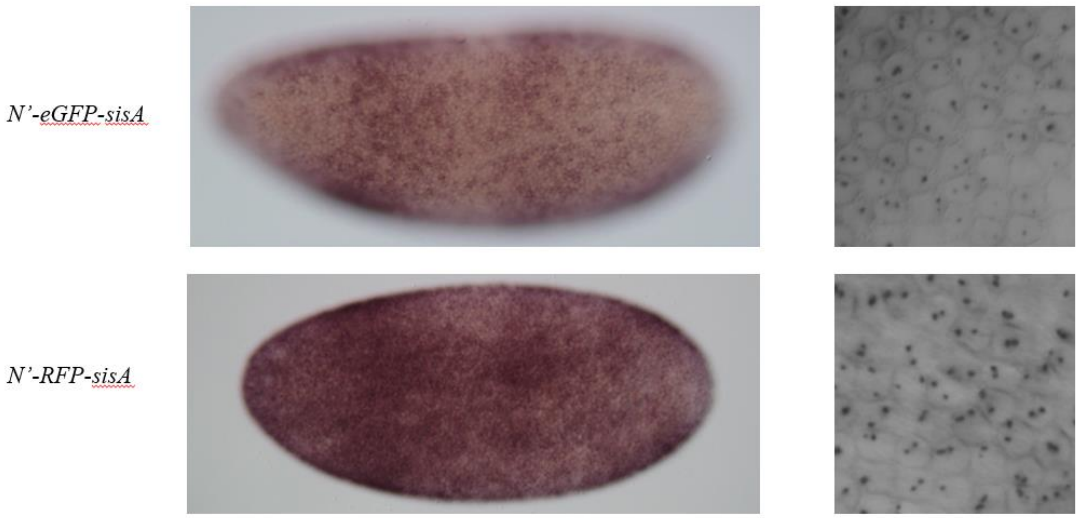

Figure 6. *In situ* hybridization of *SxlPe* in *sisA*^{5}/ *sisA*^{5}; N-tag-*sisA*/ N-tag-*sisA* embryos in early cycle 14.
(Top Panel) N-eGFP-*sisA* embryos showed mutant like, weak *SxlPe* expression. Many of the nuclei failed to express *SxlPe* as illustrated by the presence of only one or the absence of any nuclear dots which represent nascent *SxlPe* transcripts. (Lower Panel) N-RFP-*sisA* embryos showed wild type like, strong uniform *SxlPe* expression. All nuclei expressed *SxlPe* as illustrated by the presence of two nuclear dots.

To confirm that those N-tag-*sisA* constructs were able to activate *SxlPe in vivo*, I performed *in situ* hybridization on homozygous *sisA*^{5} mutant embryos carrying two copies of N-tag-*sisA* transgenes using probes that can specifically detect *SxlPe*. Because these embryos lack any endogenous *sisA*, the N-tag-*sisA* transgenes are the only source

of SisA and are thus responsible for regulating *SxlPe* activity. The results of *in situ* were consistent with the genetic tests (Figure 6). N-RFP-*sisA* embryos (and 3X FLAG -- not shown) displayed a strong, uniform expression pattern similar to wild type while N-eGFP-*sisA* embryos had a mutant like expression pattern with much weaker expression level (Figure 6). This suggests that low rescue of N-eGFP-*sisA* transgene in females is likely due to its failure to properly activate *SxlPe*.

Purification of SisA — protein nuclear extract and immunoprecipitation (IP)

The first requirement in SisA purification is to acquire sufficient amount of SisA and its partner protein(s). According to temporal expression pattern of SisA and its function in activating *SxlPe*, both sisA and its partner(s) must be present during early development stages. Thus, I used 1 to 4 hour old embryos with two copies of N-tag-*sisA* transgenes as my starting material. Protein nuclear extraction was performed on those embryos to enrich the amount of SisA and its partner and avoid contaminants since SisA is predicted to be a transcription factor and suspected to be more abundant in the nuclei. To ensure sufficient yield of protein for Co-immunoprecipitation (Co-IP) after nuclear extract, around 10 grams of embryos were used. I modified the nuclear extract protocol from the methods described in papers of Strubbe [54] and Lewis [56]. Embryos were first manually homogenized using Wheaton Dounce homogenizer and filtered using cell strainers, followed by centrifugation to pellet the nuclei. The nuclei were resuspended and a high salt solution (~ 0.4M ammonium sulfate) was used to extract proteins from the nuclei and separated from insoluble materials by centrifugation. Salt level was then

further increased (~ 2.4M ammonium sulfate) to precipitate the soluble proteins. The pelleted proteins were dissolved by overnight dialysis against TBS with 10% glycerol (Figure 7). All the above procedures were performed in 4◦C cold room and/or on ice.

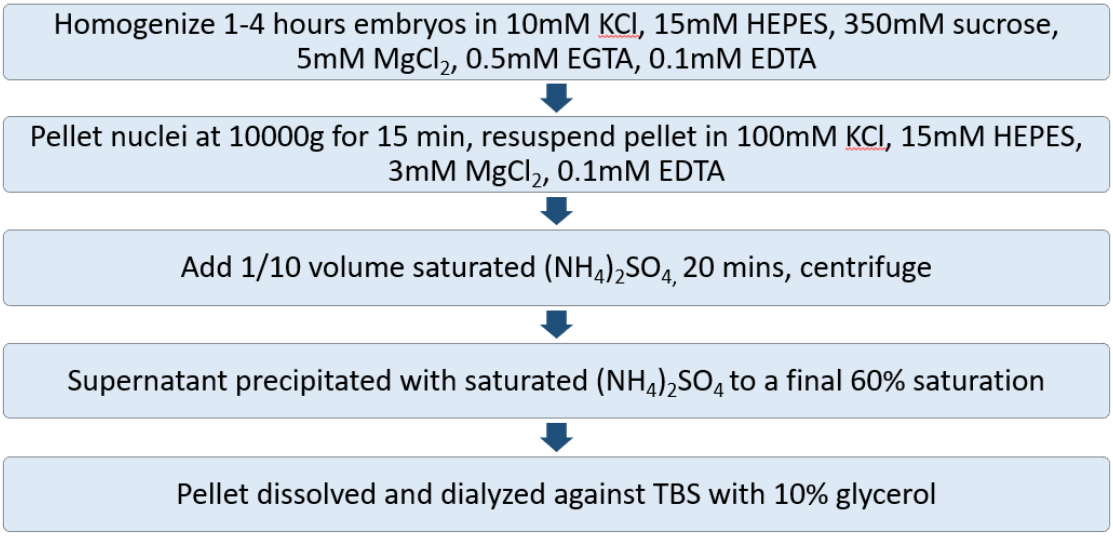

Figure 7. Flow chart of protein nuclei extract protocol.

The dialyzed proteins were incubated with RFP-Trap® magnetic agarose beads (Chromotek) or Anti-FLAG M2 magnetic beads (Sigma-Aldrich), respectively. Co-IP protocols provided by the companies were followed for purification of N-RFP-SisA and N-3xFLAG-SisA. Proteins extracted from N-RFP-*sisA* embryos were used as control to incubate with Anti-FLAG M2 magnetic beads and proteins extracted from N-3xFLAG-*sisA* embryos were used as control to incubate with RFP-Trap® magnetic agarose beads. Beads were washed with washing buffer supplemented with 50ug/ml ethidium bromide

to eliminate chromatin contamination by disrupting DNA-protein interactions. Proteins were finally eluted using 0.2 glycine pH 2.5 or by boiling the beads with SDS sample buffer.

I attempted to determine how many proteins might be associated with the beads by analyzing the proteins, eluted from the RFP-Trap beads by checking the eluates on SDS-PAGE gel, either by Coomassie blue or silver staining. I was unable, however, to detect any proteins in the final samples. In contrast, I did observe proteins in the eluates from ANTI-FLAG beads by silver staining. Importantly, I could not observe FLAG-SisA on the gel (~25 KDa). Since SisA is a small peptide and probably of low abundance, along with its partner, I thought it was possible that sisA could be present in the samples but at too low a concentration to be visible in the gel stained samples. Accordingly, I sent my samples to the mass spectrometry facility at Rutgers University. Unfortunately, the mass spec failed to detect any SisA in my sample. The list below shows some of the proteins that were unique in my sample or were of high abundance in the sample comparing to the control (Table 3). But, given the fact that SisA was not purified and pulled down, those proteins are unlikely to be present due to specific binding to the beads and should not be taken into consideration as protein partner of SisA.

Protein	Description	Sample	Control
FBpp0081261	Prat:p,	9	0
FBpp0089396	Rnp4F:p,	8	0
FBpp0086061	Dcr-2:p,	8	0
FBpp0076375	CG6638:p,	7	0
FBpp0073572	Bap60:p,	7	0
FBpp0084350	exo84:p,	6	0
FBpp0086441	CG30085:p	6	0
FBpp0070596	lva:p, lava lamp	6	0
FBpp0303335	no protein information	11	1
FBpp0079869	dgt2:p,	11	1
FBpp0304986	Dynein heavy chain 64C	30	3
FBpp0087646	CG8777:p,	10	1
FBpp0271890	CG14299:p,	9	1
FBpp0082353	RpII140:p,	9	1
FBpp0085157	CG1635:p,	9	1

Table 3. Several proteins that were uniquely presented in the sample or had a higher ratio in the sample comparing to the control.
Mass spec was performed on proteins eluted using ANTI-FLAG antibody. Proteins extracted from N-3xFLAG-*sisA* embryos were used as sample and proteins extracted from N-RFP-*sisA* embryos were used as control. The numbers in the table represented the count of reads in mass spec.

The mass spec results indicated that SisA was either lost during nuclear extract and CoIP or was not able to efficiently bind to the beads. To identify where SisA was throughout the sample preparation, I analyzed saved materials from different steps of the experiment and tried to locate SisA using western blots. I first tried GFP polyclonal antibody from BioVision to detect N-RFP-SisA in my samples. This antibody was claimed to be able to detect RFP. However, no signal was detected in any of my samples either by ECL western blot or fluorescence western blot. As stated earlier, a 3xHA tag also present right after the RFP tag, I then tried anti-HA monoclonal antibody from SIGMA and

failed to get any signal. The result was unanticipated since I expected at least to detect RFP- or HA-tagged SisA in sample from the first step, the homogenized embryos. It is possible that GFP antibody may have difficulty reacting with RFP tag and the fusion of RFP, 3xHA with SisA might also negatively affect the antibodies ability to react with the tags.

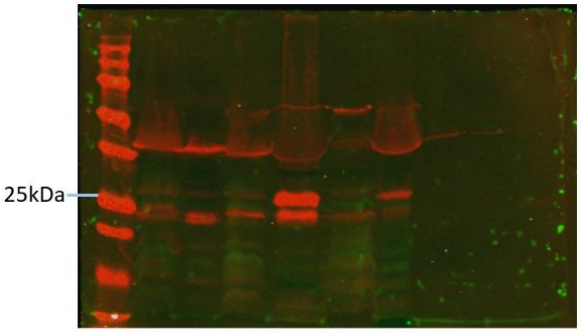

Figure 8. Western blot using anti-FLAG antibody.
From Left to right: Precision Plus Protein™ All Blue Prestained Protein Marker, homogenized N-3xFLAG-*sisA* embryos, supernatant after 1st spin, resuspended pellet after 1st spin, after adding 1/10 vol of saturated $(NH_4)_2SO_4$, insoluble after incubation with 1/10 vol of saturated $(NH_4)_2SO_4$, unbound proteins after incubating with Anti-FLAG M2 magnetic beads, IP wash (TBS with 50ug/ml EtBr), IP elution (0.1M glycine HCl, pH 3.0).

Since N-RFP-SisA proved hard to detect, I used anti-FLAG antibody for Western blots of samples obtained from embryos carrying N-3xFLAG-*sisA* genes. Numerous proteins were detecting using this antibody (Figure 8). Nevertheless, these signals did not help me to confirm the presence of SisA due to two complications. First, there was high

staining with this antibody so a great number of nonspecific bands were detected. Second, the specific MW of N-3xFLAG-SisA protein as expressed in *Drosophila* embryos is not known. Since *E.coli* expressed N-3xFLAG-SisA protein was known to migrate around 25kDa, I expected to find the protein to be similar MW in *Drosophila*. Unfortunately, there happened to be two bands around 25kDa, making it hard to distinguish SisA from the nonspecific signal. Hence, it was hard to be certain about the presence of SisA in the samples. While proteins around 25kDa could be detected in various steps including the final protein nuclear extracts, no such signal was obtained from wash and elution samples from IP experiments suggesting that the sisA protein was lost or failed to bind to the beads. In addition to the antibodies used to detect the tags, I used an Anti-SisA antiserum produced by Cocallico Biologicals, Inc for our laboratory. This antibody showed more background staining than the anti-FLAG antibody even after purification using CNBR activated sepharose 4B (details of this method are discussed in Chapter II and IV). In spite of that, the results of Western blots using anti-SisA antibody resembled the results using anti-FLAG antibody. Signals around 25kDa were detected in various steps of nuclear protein extracts and proteins unbound to the beads during IP, but not in IP wash and elution. Although we could not confirm the presence of SisA in the samples that do show signals around 25kDa, one thing is certain is that no or very low amount of N-3xFLAG-SisA actually bound to the beads even if SisA existed in nuclear extract.

IP by immobilizing *E.coli* overexpressed SisA with antibody beads

Direct use of nuclear extracts for IP was proved unsuccessful because we could not recover SisA itself from the samples. As SisA is known to be highly insoluble when expressed in *E.coli* and refolded, it is possible that even if we obtained a decent amount of SisA in the nuclear extracts, much of it might not be in the native conformation and fail to bind with the beads conjugated with antibodies. Hence, I decided to try another method, which was to overexpress N-3xFLAG-SisA in *E.coli*, and refold it. I then immobilized refolded SisA on Anti-FLAG magnetic beads and incubated immobilized SisA with nuclear extract. The purpose was to ensure that a relatively high level of SisA was bound with the beads first, so that the immobilized protein could interact with its partner in the nuclear extract, so that both can be eluted from the beads (Figure 9).

To do this, N-3xFLAG-SisA constructs were cloned into PET28a vector. Protein was expressed in *E.coli* cells by induction with 0.5mM IPTG. Doing this also allowed us to confirm that N-3xFLAG-SisA in *E.coli* runs at ~ 25kDa, which served as reference when trying to detect the protein from embryos in Western blot experiment. After induction, *E.coli* cells were lysed with 4M urea to extract N-3xFLAG-SisA proteins as SisA is insoluble by itself. Refolding of the protein were performed by lowering the urea to a final concentration of 0.5M via a series of dialyses. Refolded SisA, in 0.5 M urea was then incubated with Anti-FLAG magnetic beads. After 4 hour of incubation, bound SisA could be eluted by boiling the beads with SDS sample buffer and easily visualized on SDS-PAGE gel.

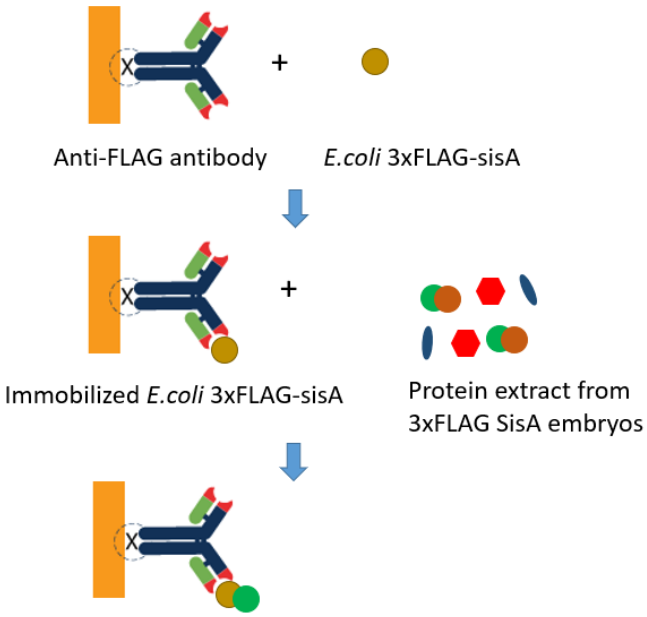

Figure 9. IP using immobilized *E.coli* overexpressed N-3xFLAG-SisA.
E.coli expressed N-3xFLAG-SisA was first incubated with Anti-FLAG antibody beads for 4 hours. Beads with immobilized *E.coli* N-3xFLAG-SisA were incubated overnight with protein nuclear extracts from embryos carrying N-3xFLAG-*sisA* transgenes. Possible protein partner of SisA interacted with *E.coli* N-3xFLAG-SisA.

With this knowledge, I performed IP by incubating nuclear extracts from embryos carrying N-3xFLAG-*sisA* transgenes with Anti-FLAG magnetic beads immobilized with *E.coli* N-3xFLAG-SisA. My control was nuclear extract from embryos carrying N-RFP-*sisA* transgenes with just Anti-FLAG magnetic beads. Both were then sent to Taplin Mass Spec Facility at Harvard Medical School. This time, SisA was successfully immunoprecipitated as expected and was the most abundant unique protein in the experimental sample and was absent from the control. However, every other protein that was uniquely detected in the sample was of low abundance (less than 10 reads). Careful

scrutinizing of them fail to discover any bZip proteins or other transcription factors, making it hard to judge if any of them is a plausible SisA partner. Table 4 shows some of the most abundant proteins that were identified in the SisA-containing samples.

Gene Symbol	Sample	Control
sisA	86	0
sti	12	0
thoc5	8	0
CysRS	6	0
SH3PX1	8	0
Aar2	5	0
Dmel\CG4452	5	0
mRF1	5	0
CG8080	5	0
DhpD	4	0
Cg17746	4	0
mRpS18B	4	0
Dmel\CG5126	4	0
mRpS2	4	0
alph	4	0
Hexo2	4	0
Tina-1	4	0
Zir	4	0
Tm1-RA	4	0
CG8230	4	0
Pdk	4	0

Table 4. Some of the most abundant proteins that were present uniquely in the sample.
The table showed several of the most abundant proteins that were detected only in the sample. The numbers in the table represented the count of reads in mass spec.

Although *E.coli* expressed N-3xFLAG-SisA can interact with Anti-FLAG beads, it seems that the immobilized SisA does not have the ability to interact with its protein partner in cell extracts. One possibility is that the SisA insolubilized on the beads because it aggregated and precipitated as the urea was diluted with the addition of nuclear extract. Alternatively the interaction between *E.coli* SisA and the partner protein was not strong enough to displace the partner from its interaction with SisA produced in the fly. It is also possible that the amount of the partner was simply too low in nuclear extract to efficiently bind to the immobilized *E.coli* SisA.

Alternative approaches

Overall, I was unable to purify SisA and its partner using antibody affinity purification and unable to identify its partner partner(s) using mass spec. Nevertheless, several alternative approaches might be useful to proceed with this project. One approach could be to use *Drosophila* S2 cells to purify SisA and its partner. We did not attempt to use S2 cells as they are derived from fly embryos at a much later stage of development and do not express *SxlPe*, making it unlikely to contain sisA and or its partner. However, preliminary results in Dr. Mark Van Doren's lab at John's Hopkins University suggests that overexpressing SisA in S2 cells can induce *Sxl* expression, suggesting the existence of a SisA partner in S2 cells. We will keep communication with them to get updates on this.

A second approach would be to carry out a maternal-effect genetic screen using a set of deletion lines covering the genome to identify SisA's partner. Although no other candidates were identified from genetic tests in previous research of *Sxl* or *sisA*, it may be worthwhile to perform genome-wide deficiency screen using X, 2nd and 3rd chromosome deletion lines available from flybase. Meanwhile, there are several deletion lines are good to start with as they contain genes that we are interested in testing if they are involved in the regulation of *SxlPe* by SisA. Alejandra Gonzalez followed candidates from Y2H screen and constructed double mutant of *CG16813* and *CG16815.* She did not see genetic interaction of this double mutant with *sisA* or *Sxl* and she also suspected redundancy of SisA partner. A recent screen for early activated zygotic genes found several other bZip proteins or transcription factors. Hence, we want to test whether these genes possibly encode SisA partner and have redundancy with previously tested *CG16813* and *CG16815*. We ordered 3 deletion lines. Df(2L)Exel6012 deleted early activated bZip protein encoded by *CG14014*. Df(2L)ED250 deleted another early activated bZip protein gene *Bsg25A*. Df(2L)BSC344 contain not only the cluster of bZip protein genes *CG16813*, *CG16815* and *CG15479* but also another early activated transcription factor gene next to them, *CG15480*. These deletion lines will be crossed with $sisA^{1}$, sxl^{-} , $sisA^{1}$ sxl^{-}, sc^{-} $sisA^{1}$, etc. to test for genetic interaction and double deletion or triple deletion lines can also be obtained via recombineering if redundancy do exist. Hopefully, the genetic tests will provide insights in finding SisA partner protein.

CHAPTER IV

OTHER WORK RELATED TO XSE AND SXL

This chapter will cover some of my other work involved in the study of how XSEs regulate *SxlPe* activity.

Identification of protein interactors involved in the regulation of *SxlPe* by *sc*

Similar to SisA, Sc is a strong XSE that is indispensable for *SxlPe* activation. Previous research found that Sc, as a bHLH transcription factor, heterodimerizes with maternal Da to bind *SxlPe* to stimulate the transcription [34-36].Sc/Da binding sites on *SxlPe* have also be identified and work in our lab showed that mutations affecting a single binding site has profound effect on *Sxl* expression (J. Rajendren and J. Erickson, unpublished). I was interested to know what allowed Sc to function so effectively and whether there were any other interactors, apart from Da involved in the regulation of *SxlPe* expression. Moreover, with the SisA project going on, I thought it efficient to use the same method to purify Sc by IP and identify potential protein interactors via mass spec. In the case of Sc, Da can served as a positive control to confirm the effectiveness of this approach in studying endogenous protein-protein interaction. In addition, with tagged Sc constructs, we can also perform ChIP to study the previously found Sc/Da binding sites and their affinities *in vivo*.

Genotype	eGFP-*sc*		Genotype	eGFP-*sc*		
	%	#		%	#	
$\frac{y\ sc^{72}\ w}{FM7}; \frac{+}{+}$	100	54	$\frac{sc^{M6}\ w}{FM7}; \frac{+}{+}$	100	112	Reference
$\frac{y\ sc^{72}\ w}{y\ sc^{72}\ w}; \frac{+}{+}$	0	0	$\frac{sc^{M6}\ w}{sc^{M6}\ w}; \frac{+}{+}$	0	0	♀ Control
$\frac{y\ sc^{72}\ w}{y\ sc^{72}\ w}; \frac{TG}{+}$	24.1	13	$\frac{sc^{M6}\ w}{FM7}; \frac{TG}{+}$	26.8	30	♀ Experiment

Table 5. Complementation of *sc* mutants by single copy of eGFP-*sc* transgenes.
Crosses set up at 25 ◦C. female $\frac{y\ sc^{72}\ w}{FM7}$ × male $\frac{y\ sc^{72}\ w}{Y}; \frac{TG}{+}$ or female $\frac{sc^{M6}\ w}{FM7}$ × male $\frac{sc^{M6}\ w}{Y}; \frac{TG}{+}$

It is known anecdotally that both N-terminal and C-terminal insertion of tags will negatively affect Sc function. Hence, I searched regions are not conserved and also predicted to be disordered among different *Drosophila* species. Various prediction methods agreed that the region located around 84th-101st amino acid is very likely to be disordered, a region upstream of the bHLH domain. Therefore, I inserted eGFP, RFP, mCherry tags (constructs were same as the ones used for tags fused with SisA) after the 95th amino acid of Sc, respectively, via *galK* recombineering techniques. P[ac]man vector carrying those tagged-Sc constructs were sent to Bestgene, Inc. to integrate into fly genome through injection. Unluckily, due to the big size of P[ac]man and also maybe the less effectiveness of the integration site I chose, only eGFP tagged Sc transgenic line were obtained. Complementation test were performed to see if this transgene can rescue sc^{72} or sc^{M6} mutants, both alleles are homozygous lethal for females. Table 5 showed

that one copy of transgene only partially rescued homozygous sc^{72} or sc^{M6} females, with around 25% viability in both cases.

Genotype	eGFP-*sc*		RFP-*sc*		mCherry-*sc*		
	%	#	%	#	%	#	
$\frac{y\ sc^{72}\ w}{FM7}; \frac{+}{+}$	100	68	100	54	100	64	Reference
$\frac{y\ sc^{72}\ w}{y\ sc^{72}\ w}; \frac{+}{+}$	0	0	0	0	0	0	♀ Control
$\frac{y\ sc^{72}\ w}{y\ sc^{72}\ w}; \frac{TG}{+}$	1.5	1	92.6	50	34.4	22	♀ Experiment

Table 6. Complementation of sc^{72} mutants by single copy of tagged *sc* transgenes. Crosses set up at 25 ◦C. female $\frac{y\ sc^{72}\ w}{FM7}$ × male $\frac{y\ sc^{72}\ w}{Y}; \frac{TG}{+}$

The possible reason that *sc* function was greatly impaired by the insertion of eGFP tag was that even thought the insertion site was not in the bHLH domain, it was very close to the bHLH domain and might negative affect its function. Alternatively, as seems to be the case with sisA, the insertion may have disrupted a key unstructured region. Thus, I generated new constructs that had the tags inserted after the 85th amino acid, which was 10 amino acid more upstream comparing to the previous constructs. All three transgenic lines, eGFP-*sc*, RFP-*sc*, mCherry-*sc*, were obtained and tested to see their ability to rescue sc^{72} mutants. Interestingly, although the new eGFP-*sc* constructs seemed to be almost completely devoid of the Sc function, RFP-*sc* constructs fully rescued sc^{72} mutants while mCherry-*sc* constructs partially rescued sc^{72} mutants with 34.4% viability

(Table 6). This might imply that not only the insertion location of the tag, but the tag itself may interfere with the protein function. The observation of eGFP tag with lowest viability in both *sisA* and *sc* cases also suggested that this eGFP tag may not be ideal to use to tag proteins for use in the narrow time window of the precellular embryo.

Thus, I obtained RFP-s*c* transgenic lines that with wild type *sc* function that can be used to for IP experiments. However, due to the difficulties with purification of the tagged sisA from embryos, I have not yet tried IP with this line.

Effects of TAGteam binding sites mutations on *SxlPe*

Zelda is a zinc finger transcription factor that is also considered as an activator of *SxlPe* in addition to the four XSEs. Maternally deposited Zelda acts as a pioneering transcription factor that is essential to the transcription of many early genes before and during maternal-to-zygotic transition (MTZ) [40-42]. Many early developmental gene, including *Sxl*, *sisA* and *sc*, were found to contain TAGteam sites, a *cis*-regulatory heptamer motif of CAGGTAG or related sequences that serves as Zelda specific binding sites [40-42, 61]. In the 400bp promoter region of *SxlPe*, there are a cluster of four TAGteam sites, referred as TAGteam 1, TAGteam 2 and TAGteam doublet as the last two sites overlapped with each other [42]. Previous research had shown that mutating these sites in 1.4kb *SxlPe-lacZ* transgene could lead to reduced *lacZ* expression [42]. The exact mechanism of how Zelda is involved in *SxlPe* activity is not known, but research suggests that Zelda might work as transcription activator by increasing chromatin

accessibility via its binding to TAGteam sites [43, 62, 63]. *SxlPe* might be an interesting subject for this theory due to the their being binding sites for both activators and repressors near the Zelda binding sites. TAGteam 1 is close to Sc/Da binding site while TAGteam 2 is next to two Dpn repressor binding sites as well as a Sc/Da site (Figure 6). If Zelda binding does affect chromatin accessibility, it might encourage the binding of repressors as well as activators depending on the nearby binding sites offering an explanation for how zelda can function as both a positive and negative regulator of transcription.

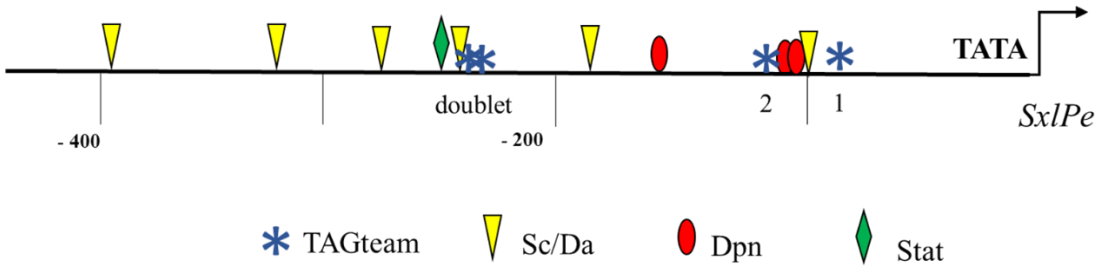

Figure 10. TAGteam sites and other activator and repressor binding sites at *SxlPe*. TAGteam 1 site is close to activator Sc/Da binding site 1. TAGteam 2 site is next to two repressor Dpn sites. TAGteam doublet contain two overlapping site and they overlap with activator Sc/Da binding site 3.

Jayashre Rajendren, a previous PhD student in the lab, has characterized various binding sites on *SxlPe*. She had designed and made specific mutations for all four TAGteam sites (Table 7), obtained three transgenic fly lines, TAGteam 1, TAGteam 2 and TAGteam

doublet and performed genetic tests and *in situ* on TAGteam doublet line. Her results showed that mutating TAGteam doublet site does negatively affect *SxlPe* acitivity.

Binding Site	Sequences	Mutations
TAGteam 1	CTACCTG	GTAGTTC
TAGteam 2	TAGGTAG	GAACTAC
TAGteam doublet	catCTGCCTGCCTG	catCTGCGCCCCTG

Table 7. Sequences of TAGteam sites mutation.
Table showed the original sequences and the mutated sequence of each site. For TAGteam doublet mutation, both sites were mutated but the overlapping Sc/Da 3 site was not affected. The Sc/Da binding site 3 is underlined with non-TAGteam sequences shown in lower case.

I followed up by performs complementation tests with all three lines with *Sxl* mutants, Sxl^{f1}, Sxl^{f9} and Sxl^{7BO}. Sxl^{f9} allele affects early *Sxl* activity and is usually the weakest of all three mutants [8, 64]. All three transgenic lines completely or near completely rescued Sxl^{f9} mutants (Table 8). Sxl^{f1} is a null allele that affects later maintenance of *Sxl* [65] . Inexplicably I was not able to recover any progeny from Sxl^{f1} females when crossing with Sxl^{f1}/Y; TAGteam 1^{-}/+ males. Nevertheless, transgenes carrying TAGteam 2 and TAGteam doublet mutation rescued Sxl^{f1} mutants very poorly, with just 8.6% and 18.9% viability, respectively (Table 9). Sxl^{7BO} is another null allele that has a complete deletion of *Sxl* and several adjacent loci [5]. The complementation tests showed that mutations of these sites all had moderate effects on *SxlPe* activity (Table 10).

Genotype	TAGteam 1		TAGteam 2		TAGteam doublet		
	%	#	%	#	%	#	
$\frac{w\ sxl^{f9}\ ct}{w\ sxl^{f9}\ ct}\ ;\ \frac{+}{+}$	0	0	0	0	0	0	♀ Control
$\frac{w\ sxl^{f9}\ ct}{w\ sxl^{f9}\ ct}\ ;\ \frac{TG}{+}$	91.6	55	106.4	50	120.5	47	♀ Experiment
$\frac{w\ sxl^{f9}\ ct}{Bincy}\ ;\ \frac{+}{+}$	100	60	100	47	100	39	Reference

Table 8. Complementation of *Sxl^{f9}* mutants by single copy of transgene carrying TAGteam site mutation.

Crosses set up at 25 ◦C. female $\frac{w\ sxl^{f9}\ ct}{Bincy}$ × male $\frac{w\ sxl^{f9}\ ct}{Y}\ ;\ \frac{TG}{+}$

Genotype	TAGteam 2		TAGteam doublet		
	%	#	%	#	
$\frac{y\ w\ cm\ Sxl^{f1}\ ct}{y\ w\ cm\ Sxl^{f1}\ ct}\ ;\ \frac{+}{+}$	0	0	0	0	♀ Control
$\frac{y\ w\ cm\ Sxl^{f1}\ ct}{y\ w\ cm\ Sxl^{f1}\ ct}\ ;\ \frac{TG}{+}$	8.6	5	18.9	10	♀ Experiment
$\frac{y\ w\ cm\ Sxl^{f1}\ ct}{Bincy}\ ;\ \frac{+}{+}$	100	58	100	53	Reference

Table 9. Complementation of *Sxl^{f1}* mutants by single copy of transgene carrying TAGteam site mutation.

Crosses set up at 25 ◦C. female $\frac{y\ w\ cm\ Sxl^{f1}\ ct}{Bincy}$ × male $\frac{y\ w\ cm\ Sxl^{f1}\ ct}{Y}\ ;\ \frac{TG}{+}$

Genotype	TAGteam 1		TAGteam 2		TAGteam doublet		
	%	#	%	#	%	#	
$\frac{y\ pn\ w\ Sxl^{7BO}}{y\ pn\ w\ Sxl^{7BO}}\ ;\ \frac{+}{+}$	0	0	0	0	0	0	♀ Control
$\frac{y\ pn\ w\ Sxl^{7BO}}{y\ pn\ w\ Sxl^{7BO}}\ ;\ \frac{TG}{+}$	41.0	25	74.0	37	66.2	43	♀ Experiment
$\frac{y\ pn\ w\ Sxl^{7BO}}{Bincy}\ ;\ \frac{+}{+}$	100	61	100	50	100	65	Reference

Table 10. Complementation of Sxl^{7BO} mutants by single copy of transgene carrying TAGteam site mutation.

Crosses set up at 25 ◦C. female $\frac{y\ pn\ w\ Sxl^{7BO}}{Bincy}$ × male $\frac{y\ pn\ w\ Sxl^{7BO}}{Y}\ ;\ \frac{TG}{+}$

The genetic tests above only showed evidence that TAGteam 2 site functions as an activator site just as the other two mutations, TAGteam 1 and TAGteam doublet. However, genetic tests may not be ideal as Zelda binding to TAGteam 2 sites might not only increase the chromatin accessibility of Dpn to the two adjacent repressor binding sites, but also encourage Sc/Da binding to the nearby activator binding sites. Therefore, *in situ* hybridization of *SxlPe* might be a better method to check if TAGteam 2 site has dual role in regulation of *SxlPe* activity as it may reveal very subtle ectopic *SxlPe* expression that is not strong enough to affect viability. Also we could bring transgene carrying TAGteam 2 site mutation into *sc* mutant background to solely focus on its effect on Dpn repressor sites via genetic tests and *in situ*.

Purification of SisA antibody

In addition to using RFP, FLAG and HA antibody tagged SisA protein detection, I have also tested anti-SisA antibody available in our lab to see how effectively this antibody can detect SisA. Our lab had obtained anti-SisA antisera raised in guinea pigs from Cocalico Biologicals, Inc. Dot blot experiment showed the antibody can detect *E.coli* expressed SisA, but produced no signal with my negative control, purified Sxl peptide. The signal for SisA was very strong, as staining is still visible when the antibody is diluted $1:10^7$. Antibody staining of 0-7 hour wild type embryos revealed that this antibody had very strong background issues. Early embryos even before the initiation of *sisA* expression were stained and later embryos were all darkly stained, with no differentiation between different sexes. A large series of dilutions did nothing to bring up any specific sisA signal.

Since the specificity of this antibody was in question, I decided to try to purify this antibody as it would be helpful to have anti-SisA specific antibody to help track down SisA in my IP experiment. The strategy I employed was to couple purified SisA protein with cyanogen bromide (CNBR) activated sepharose 4B beads and then use it to affinity purify anti-SisA antiserum. First, I constructed His-tagged SisA and overexpressed the protein in *E.coli*. As SisA s insoluble, I lyzed the *E.coli* cells and purified His tagged SisA protein using Ni-NTA agarose beads under denaturing condition. The purified SisA was coupled with CNBR activated sepharose 4B beads and then anti-SisA antisera was incubated with the beads at 4 °C overnight before eluted with 0.2M glycine pH 2.8.

However, the background issue of the antibody was not resolved by this purification method. 0-7 hour embryos were still universally darkly stained regardless of their developmental stages. When using this purified anti-SisA antibody to perform western blot with materials from different steps of nuclear extraction and IP, it could detect a broad band around 25kDa, the predicted size of SisA, but it also had more unspecific bands and background signals comparing to using anti-FLAG antibody. Therefore, even though there is anti-SisA antisera in the lab, I could not successfully purify it to the extent that it can be effectively and specifically used to detect SisA proteins in embryos or in IP experiments.

CHAPTER V

CONCLUSIONS

Even though my efforts in trying to identify SisA's protein partner were not successful, some valuable experience gained through this process may be helpful for people who would like to carry out similar projects. First, disordered regions proved not to be an ideal region to insert epitope tags in sisA and perhaps sisB. This is true even the sequence of the region is not conserved among different species. Therefore, the conventional choice of inserting tags at N- or C-termini should still be the first options tested. Second, the eGFP tag I fused with SisA in my experiment seemed to have a much more negative influence on SisA function comparing with other tags used. This also applied to my tagged Sc constructs, with eGFP-tagged Sc also exhibiting much lower activity than RFP-tagged and mCherry-tagged Sc when inserted at the same location. It is possible that the structure of this eGFP negatively interact structure of other proteins and I would recommend using other tags for similar experiment. The poor performance of eGFP tag could also be due to the fact that the tandem tags FlAsH-StrepII-TEV-3xFlag following eGFP were affecting the protein function. This could be tested by using just eGFP alone. An alternative possibility is that folding of eGFP is too slow to allow SisA or Sc function in the brief time period available for sex determination. The third and most important lesson for future experiments, is the importance of obtaining effective antibodies that can detect the tag or SisA in cell extracts. This would help track the amount of SisA in every step from preparation of the nuclear extract until IP. Such

antibodies might also be used to directly purify SisA and its partner. My experiences strongly suggest it would not be worthwhile to repeat similar experiments until we find any antibody that has higher specificity and affinity against SisA, RFP or 3xFLAG to effectively detect tagged SisA. Alternatively, we can switch to other epitope tags that have better antibodies. A small tag is highly recommended as the previous experiments suggest that big tags with slow folding rate might affect SisA's conformation and function. Fourth, I suggest to use RFP tagged *sc* transgenic lines to set up positive control experiment before proceeding with purification of SisA and its partner. As Sc is known to heterodimerize with Da, the mass spec is expected to identify Da specifically in the sample. This is a useful standard to optimize the protocol of nuclear extract, protein refolding and IP by altering slat concentration, pH, urea concentration, etc. In previous experiment, I was not able to detect RFP tagged *sisA* in western blot, suggesting the specific version RFP in my construct might not interact well with RFP nanobody and RFP antibody. Therefore, using RFP tagged *sc* can also help to verify whether the RFP tag I used can be detected by RFP nanobody and RFP antibody. Importantly, we can use the relative abundancy of Sc and Da obtained in the mass spec as a reference to have a better idea how much how much relative abundancy of protein partner/interactor is expected to be detected in mass spec. SisA itself is only around 25kDa and its possible partner is probably small as well, not to say it could even be multiple redundant partners, it is very possible that the reads of the partner can be very low and one major obstacle in previous screen is that we do not know what threshold should be ideal for selecting partner candidates. If Da is identified but of much lower

reads comparing to Sc, it might be worthwhile to re-examine the previous mass spec list and take into consideration of those proteins that are not obvious candidates as transcriptional regulators. Last, trying to overexpress tagged SisA in flies may be another way to increase the yield and efficiency of SisA purification. However, there is not an easy to achieve that. The most convenient method is to increase the copies of tagged SisA transgene via P element insertion.

In addition to the above suggestions, it is necessary to remember that we cannot exclude the possibility that SisA actually acts in a unique way in the regulation of *SxlPe* independent of any partner protein. If that is the case other strategies must be devised to determine how sisA works.

REFERENCES

1. Bridges, C.B., *Non-Disjunction as Proof of the Chromosome Theory of Heredity.* Genetics, 1916. **1**(1): p. 1-52.

2. Erickson, J.W. and J.J. Quintero, *Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila.* PLoS Biol, 2007. **5**(12): p. e332.

3. Cline, T.W., *The Drosophila sex determination signal: how do flies count to two?* Trends Genet, 1993. **9**(11): p. 385-90.

4. Cline, T.W. and B.J. Meyer, *Vive la difference: males vs females in flies vs worms.* Annu Rev Genet, 1996. **30**: p. 637-702.

5. Salz, H.K. and J.W. Erickson, *Sex determination in Drosophila: The view from the top.* Fly (Austin), 2010. **4**(1): p. 60-70.

6. Bell, L.R., et al., *Sex-lethal, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins.* Cell, 1988. **55**(6): p. 1037-46.

7. Penalva, L.O. and L. Sanchez, *RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation.* Microbiol Mol Biol Rev, 2003. **67**(3): p. 343-59, table of contents.

8. Siera, S.G. and T.W. Cline, *Sexual back talk with evolutionary implications: stimulation of the Drosophila sex-determination gene sex-lethal by its target transformer.* Genetics, 2008. **180**(4): p. 1963-81.

9. Valcarcel, J., et al., *The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA.* Nature, 1993. **362**(6416): p. 171-5.

10. Burtis, K.C. and B.S. Baker, *Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides.* Cell, 1989. **56**(6): p. 997-1010.

11. Heinrichs, V., L.C. Ryner, and B.S. Baker, *Regulation of sex-specific selection of fruitless 5' splice sites by transformer and transformer-2.* Mol Cell Biol, 1998. **18**(1): p. 450-8.

12. Horabin, J.I. and P. Schedl, *Regulated splicing of the Drosophila sex-lethal male exon involves a blockage mechanism.* Mol Cell Biol, 1993. **13**(3): p. 1408-14.

13. Demir, E. and B.J. Dickson, *fruitless splicing specifies male courtship behavior in Drosophila.* Cell, 2005. **121**(5): p. 785-94.

14. Gergen, J.P., *Dosage Compensation in Drosophila: Evidence That daughterless and Sex-lethal Control X Chromosome Activity at the Blastoderm Stage of Embryogenesis.* Genetics, 1987. **117**(3): p. 477-85.

15. Bashaw, G.J. and B.S. Baker, *The regulation of the Drosophila msl-2 gene reveals a function for Sex-lethal in translational control.* Cell, 1997. **89**(5): p. 789-98.

16. Gelbart, M.E. and M.I. Kuroda, *Drosophila dosage compensation: a complex voyage to the X chromosome.* Development, 2009. **136**(9): p. 1399-410.

17. Kuroda, M.I., et al., *The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila.* Cell, 1991. **66**(5): p. 935-47.

18. Duncan, K., et al., *Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3' UTR: translational repression for dosage compensation.* Genes Dev, 2006. **20**(3): p. 368-79.

19. Duncan, K.E., C. Strein, and M.W. Hentze, *The SXL-UNR corepressor complex uses a PABP-mediated mechanism to inhibit ribosome recruitment to msl-2 mRNA.* Mol Cell, 2009. **36**(4): p. 571-82.

20. Gebauer, F., M. Grskovic, and M.W. Hentze, *Drosophila sex-lethal inhibits the stable association of the 40S ribosomal subunit with msl-2 mRNA.* Mol Cell, 2003. **11**(5): p. 1397-404.

21. Gonzalez, A.N., H. Lu, and J.W. Erickson, *A shared enhancer controls a temporal switch between promoters during Drosophila primary sex determination.* Proc Natl Acad Sci U S A, 2008. **105**(47): p. 18436-41.

22. Bell, L.R., et al., *Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in Drosophila.* Cell, 1991. **65**(2): p. 229-39.

23. Cline, T.W., *Autoregulatory functioning of a Drosophila gene product that establish es and maintains the sexually determined state.* Genetics, 1984. **107**(2): p. 231-77.

24. Horabin, J.I. and P. Schedl, *Sex-lethal autoregulation requires multiple cis-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5' splice site.* Mol Cell Biol, 1993. **13**(12): p. 7734-46.

25. Nagengast, A.A., et al., *Sex-lethal splicing autoregulation in vivo: interactions between SEX-LETHAL, the U1 snRNP and U2AF underlie male exon skipping.* Development, 2003. **130**(3): p. 463-71.

26. Erickson, J.W. and T.W. Cline, *A bZIP protein, sisterless-a, collaborates with bHLH transcription factors early in Drosophila development to determine sex.* Genes Dev, 1993. **7**(9): p. 1688-702.

27. Cline, T.W., *Evidence That Sisterless-a and Sisterless-B Are 2 of Several Discrete Numerator Elements of the X/a Sex Determination Signal in Drosophila That Switch Sxl between 2 Alternative Stable Expression States.* Genetics, 1988. **119**(4): p. 829-862.

28. Duffy, J.B. and J.P. Gergen, *The Drosophila segmentation gene runt acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene Sex-lethal.* Genes Dev, 1991. **5**(12A): p. 2176-87.

29. Jinks, T.M., et al., *The JAK/STAT signaling pathway is required for the initial choice of sexual identity in Drosophila melanogaster.* Molecular Cell, 2000. **5**(3): p. 581-587.

30. Sanchez, L., B. Granadino, and M. Torres, *Sex determination in Drosophila melanogaster: X-linked genes involved in the initial step of sex-lethal activation.* Dev Genet, 1994. **15**(3): p. 251-64.

31. Sefton, L., et al., *An extracellular activator of the Drosophila JAK/STAT pathway is a sex-determination signal element.* Nature, 2000. **405**(6789): p. 970-3.

32. Walker, J.J., et al., *The Drosophila melanogaster sex determination gene sisA is required in yolk nuclei for midgut formation.* Genetics, 2000. **155**(1): p. 191-202.

33. Fassler, J., et al., *B-ZIP proteins encoded by the Drosophila genome: evaluation of potential dimerization partners.* Genome Res, 2002. **12**(8): p. 1190-200.

34. Torres, M. and L. Sanchez, *The scute (T4) gene acts as a numerator element of the X:a signal that determines the state of activity of sex-lethal in Drosophila.* EMBO J, 1989. **8**(10): p. 3079-86.

35. Torres, M. and L. Sanchez, *The sisterless-b function of the Drosophila gene scute is restricted to the stage when the X:A ratio determines the activity of Sex-lethal.* Development, 1991. **113**(2): p. 715-22.

36. Yang, D., et al., *Interpretation of X chromosome dose at Sex-lethal requires non-E-box sites for the basic helix-loop-helix proteins SISB and daughterless.* Molecular and Cellular Biology, 2001. **21**(5): p. 1581-1592.

37. Kramer, S.G., et al., *Direct activation of Sex-lethal transcription by the Drosophila Runt protein.* Development, 1999. **126**(1): p. 191-200.

38. Avila, F.W. and J.W. Erickson, *Drosophila JAK/STAT pathway reveals distinct initiation and reinforcement steps in early transcription of Sxl.* Curr Biol, 2007. **17**(7): p. 643-8.

39. Harrison, D.A., et al., *Drosophila unpaired encodes a secreted protein that activates the JAK signaling pathway.* Genes Dev, 1998. **12**(20): p. 3252-63.

40. Liang, H.L., et al., *The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila.* Nature, 2008. **456**(7220): p. 400-3.

41. Nien, C.Y., et al., *Temporal coordination of gene networks by Zelda in the early Drosophila embryo.* PLoS Genet, 2011. **7**(10): p. e1002339.

42. ten Bosch, J.R., J.A. Benavides, and T.W. Cline, *The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription.* Development, 2006. **133**(10): p. 1967-77.

43. Foo, S.M., et al., *Zelda potentiates morphogen activity by increasing chromatin accessibility.* Curr Biol, 2014. **24**(12): p. 1341-1346.

44. Schulz, K.N., et al., *Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo.* Genome Res, 2015. **25**(11): p. 1715-26.

45. Sun, Y., et al., *Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.* Genome Res, 2015. **25**(11): p. 1703-14.

46. Barbash, D.A. and T.W. Cline, *Genetic and molecular analysis of the autosomal component of the primary sex determination signal of Drosophila melanogaster.* Genetics, 1995. **141**(4): p. 1451-71.

47. Younger-Shepherd, S., et al., *deadpan, an essential pan-neural gene encoding an HLH protein, acts as a denominator in Drosophila sex determination.* Cell, 1992. **70**(6): p. 911-22.

48. Lu, H., et al., *Maternal Groucho and bHLH repressors amplify the dose-sensitive X chromosome signal in Drosophila sex determination.* Dev Biol, 2008. **323**(2): p. 248-60.

49. McLarren, K.W., et al., *The mammalian basic helix loop helix protein HES-1 binds to and modulates the transactivating function of the runt-related factor Cbfa1.* J Biol Chem, 2000. **275**(1): p. 530-8.

50. McLarren, K.W., F.M. Theriault, and S. Stifani, *Association with the nuclear matrix and interaction with Groucho and RUNX proteins regulate the transcription repression activity of the basic helix loop helix factor Hes1.* J Biol Chem, 2001. **276**(2): p. 1578-84.

51. Venken, K.J., et al., *MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes.* Nat Methods, 2011. **8**(9): p. 737-43.

52. Warming, S., et al., *Simple and highly efficient BAC recombineering using galK selection.* Nucleic Acids Res, 2005. **33**(4): p. e36.

53. Erickson, J.W. and T.W. Cline, *Key aspects of the primary sex determination mechanism are conserved across the genus Drosophila.* Development, 1998. **125**(16): p. 3259-68.

54. Strubbe, G., et al., *Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners.* Proc Natl Acad Sci U S A, 2011. **108**(14): p. 5572-7.

55. Pazin, M.J., *Preparation of nuclear extracts from Drosophila embryos.* CSH Protoc, 2007. **2007**: p. pdb prot4709.

56. Lewis, P.W., et al., *Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex.* Genes Dev, 2004. **18**(23): p. 2929-40.

57. Vinson, C.R., P.B. Sigler, and S.L. McKnight, *Scissors-grip model for DNA recognition by a family of leucine zipper proteins.* Science, 1989. **246**(4932): p. 911-6.

58. Hewes, R.S., A.M. Schaefer, and P.H. Taghert, *The cryptocephal gene (ATF4) encodes multiple basic-leucine zipper proteins controlling molting and metamorphosis in Drosophila.* Genetics, 2000. **155**(4): p. 1711-23.

59. Kwasnieski, J.C., T.L. Orr-Weaver, and D.P. Bartel, *Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns.* Genome Res, 2019. **29**(7): p. 1188-1197.

60. Ries, J., et al., *A simple, versatile method for GFP-based super-resolution microscopy via nanobodies.* Nat Methods, 2012. **9**(6): p. 582-4.

61. Blythe, S.A. and E.F. Wieschaus, *Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis.* Elife, 2016. **5**.

62. Harrison, M.M., M.R. Botchan, and T.W. Cline, *Grainyhead and Zelda compete for binding to the promoters of the earliest-expressed Drosophila genes.* Dev Biol, 2010. **345**(2): p. 248-55.

63. Harrison, M.M., et al., *Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition.* PLoS Genet, 2011. **7**(10): p. e1002266.

64. Maine, E.M., et al., *Sex-lethal, a link between sex determination and sexual differentiation in Drosophila melanogaster.* Cold Spring Harb Symp Quant Biol, 1985. **50**: p. 595-604.

65. Cline, T.W., *A female-specific lethal lesion in an X-linked positive regulator of the Drosophila sex determination gene, Sex-lethal.* Genetics, 1986. **113**(3): p. 641-63.