Engineered Underdominance as a Method of Insect Population Replacement and Reproductive Isolation

Thesis by
Anna Basalova Buchman

In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
2014
(Defended May 28, 2014)

To Galina “Babun” Dontsova

Acknowledgements

I must begin by expressing my deep gratitude to my mentor through this whole project, Bruce Hay. Thank you for always believing in our projects, for encouraging me to keep optimistic and move forward, and for never failing to amaze me with new, adventurous, and often brilliant ideas. Over the last five years, I have really come to appreciate the meaning of this quote by W.L. Bragg: "The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them." Thank you for teaching me how to think like a scientist.

I must also thank my committee members, Drs. Sternberg, Aravin, and Zinn, for their support and feedback, and for being the kinds of scientists a graduate student is truly inspired by.

The members of the Hay lab have been my source of information, guidance, support, and camaraderie over these years, and I could not have made it without them. I especially want to thank Geoff Pittman, for teaching me everything I needed to know to survive in the Hay lab; Kelly Matzen, for laying such a solid foundation for this project and having the patience to teach me the nuts and bolts of it; and Katie Kennedy, for having magical cloning powers and being a true friend.

I want to express my deep appreciation for the various members of our Caltech community that made every day functioning here possible, including Carole Worra, Divina Bautista, Alice Sogomonian, and Linda Schutz. I also want to thank the mentors I was fortunate enough to have before Caltech, including William Lutterschmidt, Chris Randle, Monte Thies, and Ted Brummel, for sparking and fostering my interest in biology.

Last but indubitably not least, I want to express my profound gratitude to my family. My parents have always told me that if I worked hard, I could do anything, and encouraged me to dream big. They have ceaselessly fostered my academic pursuits and offered unwavering support on this journey, and I could not have done this without them. This also would not have been possible without my husband, Rob, who has patiently been here with me through the ups and downs of this

process, and whose support and love are my greatest sources of comfort. (Nothing says true love like flipping your wife's fly stocks so she can focus on finishing her thesis.) And this road would not have been nearly as enjoyable without our Kronkite, who is forever willing to brighten a bad day with his continual outpour of doggy affection.

Finally, I want to dedicate this dissertation to my grandmother, Galina. She was always my most enthusiastic, unconditional, unwavering supporter, and I miss her greatly. I hope she would have been proud.

Abstract

Insect vector-borne diseases, such as malaria and dengue fever (both spread by mosquito vectors), continue to significantly impact health worldwide, despite the efforts put forth to eradicate them. Suppression strategies utilizing genetically modified disease-refractory insects have surfaced as an attractive means of disease control, and progress has been made on engineering disease-resistant insect vectors. However, laboratory-engineered disease refractory genes would probably not spread in the wild, and would most likely need to be linked to a gene drive system in order to proliferate in native insect populations. Underdominant systems like translocations and engineered underdominance have been proposed as potential mechanisms for spreading disease refractory genes. Not only do these threshold-dependent systems have certain advantages over other potential gene drive mechanisms, such as localization of gene drive and removability, extreme engineered underdominance can also be used to bring about reproductive isolation, which may be of interest in controlling the spread of GMO crops. Proof-of-principle establishment of such drive mechanisms in a well-understood and studied insect, such as Drosophila melanogaster, is essential before more applied systems can be developed for the less characterized vector species of interest, such as mosquitoes. This work details the development of several distinct types of engineered underdominance and of translocations in Drosophila, including ones capable of bringing about reproductive isolation and population replacement, as a proof of concept study that can inform efforts to construct such systems in insect disease vectors.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

Insect-Borne Diseases

Insects are vectors for a large number of diseases that significantly impact both human health and agriculture. Malaria, dengue, yellow fever, chikungunya, Chagas disease, West Nile, Japanese encephalitis, and Rift Valley fever are just a few of the insect-borne illnesses that afflict humans (Gubler 1998; Tolle 2009), while Huanglongbing, zebra chip disease, Pierce's disease and other vector-borne plant pathogens cause billions of dollars in agriculture-related loss annually (Lemon et al. 2009). Climate change and the accelerated global movement of plants, animals, people, and goods are causing the spread of vectors and pathogens to new hosts and environments, and threatening to increase the occurrence of many vector-borne diseases (Githeko et al. 2000; Gubler et al. 2001; Lemon et al. 2009; Carvalho et al. 2014). Resurgence of malaria, dengue, and other diseases in the 1970s after their effective control/elimination in the previous decades accentuates the threat that these vector-borne pathogens continually pose to global health and economics (Gubler 1998; Lemon et al. 2009).

This threat is underscored by specific examples of the burden inflicted by some of these diseases. Malaria, which is caused by the Plasmodium parasite and transmitted person to person by the Anopheles mosquito, causes approximately 225 million cases and almost 1 million deaths, mostly in African children and pregnant women, annually (Murray et al. 2012; Crompton et al. 2014). Approximately half of the world's population is at risk for malaria (Centers for Disease Control and Prevention 2014); it is the fifth leading cause of death from infectious diseases worldwide (World Health Organization 2000); and in Africa alone, malaria kills a child every 60 seconds (World Health Organization 2014a) and imposes at least US$12 billion in direct losses a year (Gallup and Sachs 2001).

Similarly, over 40% of the world's population is at risk for contracting dengue, a viral disease that is transmitted by the Aedes mosquito (World Health Organization

2014b). Dengue is currently the fastest spreading mosquito-borne disease in the world, infecting almost 100 million people yearly (World Health Organization 2014b; Centers for Disease Control and Prevention 2012), and there is no known cure or vaccine (World Health Organization 2014b). Dengue hemorrhagic fever (DHF), a severe form of the disease that affects approximately 500,000 people annually (World Health Organization 2014b), is a major cause of global hospitalization and death (Gubler 1998), and its recent reemergence has been marked by explosive epidemics (Gubler 1998; Lemon et al. 2009).

There are presently several major methods for combating these, and other, vector-borne diseases. Suppression of total insect populations can be achieved via removal of insect breeding grounds (such as water containers) and spraying of insecticides (World Health Organization 2014b), and through the release of sterile males (known as the Sterile Insect Technique, SIT; Morrison et al. 2010; Harris et al. 2011; Miller 2011). Disease-transmitting insect bites can be prevented through use of insect repellents, insecticide-treated bed nets, and application of insecticides to dwelling walls (Gould et al. 2006; Miller and Pierce 2009; World Health Organization 2014b). And drugs to treat some of these diseases, like malaria, are available (Enayati and Hemingway 2010). However, each of these control strategies comes with pitfalls: population suppression via insecticides or SIT leaves the insect's biological niche intact, so that it can rapidly re-colonize it after suppression efforts cease; avoidance efforts like bed nets can be costly, and are not as effective for insects (like the Aedes mosquito) that bite during the day (Gould et al. 2006); and evolution of insecticide and drug resistance by targeted insects is likely, if not inevitable (Miller and Pierce 2009; Enayati and Hemingway 2010; Ndiath et al. 2012; Wang and Jacobs-Lorena 2013). While recent efforts to reduce the impact of major insect-borne diseases like malaria have had some success (World Health Organization 2014a), available control methods have not been able to completely block disease transmission: the incidence of dengue has grown thirtyfold in the last 50 years (World Health Organization 2014a), and eradication of malaria with current techniques seems unlikely (Miller and Pierce 2009; Crompton et al. 2014).

Genetic Approaches to Controlling Insect Vectors

The inability of current strategies to successfully control insect vector-borne diseases has led to increased interest in disease control methods that involve release of genetically modified mosquitoes refractory to pathogen transmission (Gould et al. 2006; Sinkins and Gould 2006; Marshall and Taylor 2009; Hay et al. 2010). The idea of replacing wild insect populations with those engineered to be incapable of disease transmission was first discussed decades ago (Curtis 1968; Whitten 1971; Foster et al. 1972); however, at that time, such research was limited by lack of genetic manipulation techniques (Gould and Schliekelman 2006).

Since then, advances in molecular biology have generated a wealth of new tools for precise genetic manipulation (Groth et al., 2004; Gould and Schliekelman 2006; Ran et al. 2013), and a number of genes that confer refractoriness to malaria and dengue have been identified or engineered (de Lara Capurro et al. 2000; Ito et al. 2002; Moreira et al. 2002; Kim et al. 2004; Franz at al. 2006; Riehle et al. 2008; Corby-Harris et al. 2010). However, disease refractory genes are not expected to confer a fitness benefit to carrier mosquitoes (Schmid-Hempel 2005; Gould et al. 2006), and a large proportion of a mosquito population must be disease-refractive to achieve meaningful levels of disease control (Boete and Koella 2002). Thus, effective disease control via population replacement is generally thought to require the linkage of disease refractory genes to a mechanism capable of driving them into a wild population (Braig and Yan 2001; James 2005; Gould et al. 2006; Sinkins and Gould 2006).

Several naturally occurring selfish genetic elements, including transposons, meiotic drive, B-chromosomes, homing endonuclease genes (HEGs), Medea elements, and the intracellular bacterium Wolbachia, have been proposed as potential gene drive mechanisms (reviewed in Sinkins and Gould 2006), along with approaches relying on linking genes of interest to engineered chromosomes, such as translocations or compound chromosomes (Curtis 1968; Gould and Schliekelman 2004). Another approach, known as engineered underdominance, is discussed in much more detail below (Davis et al. 2001; Magori and Gould 2006). Some of these strategies, including Medea (Chen et al. 2007), Wolbachia (Walker et al. 2011), and

HEGs (Windbichler et al. 2011), have been shown to have some capacity to drive gene introgression in laboratory populations, and translocations have been used to effect insect population suppression (Asman et al. 1981; Baker 1984). However, a robust mechanism of gene drive capable of spreading chosen disease refractory genes into wild populations has not yet been developed in important vector species.

High-Threshold Drive Systems

Some drive systems, like Medea and HEGs, are predicted to have low release thresholds, and therefore be invasive, spreading to high frequency even if initially introduced at very low frequency (Marshall 2009; Marshall and Hay 2011; Deredec et al. 2008). On one hand, this is desirable, since disease-refractory genes must spread to high levels to achieve disease control (Boete and Koella 2002). It may also be disadvantageous, however—insects don't respect international borders, and forceful drive systems can spread genetically modified organisms (GMOs) into communities or countries before they've agreed to welcome their introduction (Knols et al. 2007). The Cartagena Protocol (the United Nations' set of regulations governing movement of GMOs) allows countries to decide for themselves whether to allow import of GMOs, and prohibits release of GMOs capable of invasively spreading across borders without prior international agreements (Marshall 2010). Furthermore, people have been shown to prefer that transgene spread be tested in isolated locations before releases occur in their own community (Marshall et al. 2010a; Marshall et al. 2010b), and public distrust of GMOs may be considerable (Alphey et al. 2002; Bohannon 2002; Gould et al. 2006). In light of these regulatory guidelines and societal views, it is generally thought that initial releases of transgenic insects must be confinable to the isolated areas where they are being tested (Marshall and Hay 2012).

Gene drive systems based on the phenomenon of underdominance may offer a way to spread transgenes to high frequencies locally without risk of spread into neighboring populations (Altrock et al. 2010; Marshall and Hay 2012). The simplest form of underdominance, or heterozygote disadvantage, occurs when the average fitness of a heterozygote is lower than that of either parental homozygote (Hartl and

Clark 1997). Although a single-allele underdominant scheme may be difficult to engineer (Marshall and Hay 2012), several genetic systems based on the underdominant principle—including chromosome translocations (Curtis 1968), inversions (Foster et al. 1972; Robinson 1975) compound chromosomes (Foster et al. 1972; Gould and Schliekelman 2004), and engineered underdominance (Davis et al. 2001; Magori and Gould 2006; Akbari et al. 2013) —have been proposed (and, in the case of Akbari et al., engineered). These systems all have high introduction thresholds (27%–67%), and act as a bi-stable switch: if the frequency of one allele or chromosome is above a crucial threshold, that allele will spread to fixation at the expense of the other allele, while if it is below the threshold, it will be lost from the population and the other allele type will be fixed (Davis et al. 2001; Sinkins and Gould 2006; Altrock et al. 2010; Altrock et al. 2011). (A more detailed theoretical framework of underdominance, including how it can result in population replacement and what approaches can be used to bring it about, will be presented in later chapters.)

This type of system is inherently removable, since the frequency of underdominant individuals can be diluted below the critical threshold by the addition of wild types, which in some cases may be preceded by a round of insecticide application to decrease total population numbers. A high threshold gene drive mechanism is also unlikely to spread to high levels in neighboring populations linked to the source population by low levels of migration, as the frequency of underdominant alleles is not likely to reach the needed threshold, and so confined releases are possible (Altrock et al. 2010; Altrock et al. 2011; Marshall 2009; Marshall and Hay 2012). Indeed, some analyses suggest that engineered underdominance is the safest gene drive mechanism in contexts in which transgenic containment in initial field cage experiments is likely to be critical (Marshall 2009). Finally, since underdominant schemes are based on two alleles, and since each allele can be engineered to carry a different disease refractory gene, such systems offer additional insurance against breakdown of the system due to mutation or loss of the disease resistance gene (Sinkins and Guld 2006).

The high threshold required for spread of underdominant mechanisms (as compared with other systems) will necessitate significant releases of transgenic individuals to achieve fixation, and considerable releases of wild types if transgene removal is required. However, necessary release ratios are still appreciably lower that those utilized with many SIT programs (Krafsur 1998; Gould and Schiekelman 2004; Alphey et al. 2010), which in the case of the Mediterranean fruit fly involves the release of billions of insects into the wild each week (Mumford 2012). And, unlike SIT, underdominant systems are self-perpetuating (Baker 1984; Robinson 1976a), while sterile males must be released each generation. Thus, creation of underdominant systems capable of gene drive would be a valuable addition to the emerging field of using genetically modified insects to control vector-borne diseases.

In this work, I describe the development of several high-threshold underdominant systems capable of gene drive in Drosophila melanogaster as a proof of principle study that can hopefully be applied to vector species of interest.

Chapter 2

Engineering UD: Protein Toxins

2.1 Introduction

Theoretical Framework

Underdominance, or heterozygote disadvantage, in its simplest form, occurs when the average fitness of a heterozygote is lower than that of either homozygote for a trait that is determined by two alleles at the same locus (Hartl and Clark 1997). Thus, for two alleles, *A* and *a*, the heterozygote *A/a* will be less fit than either the *A/A* or *a/a* parental homozygote (Figure 2.1A). In extreme underdominance, the heterozygous state is 100% lethal, so that *A/A* and *a/a* individuals are reproductively isolated from each other. This process contributes to speciation (e.g., Eppstein et al. 2009), and it can also be appropriated as a mechanism for gene drive. In a two-allele underdominant system, there exist two stable, and one unstable, equilibria (Magori and Gould 2006). One allele typically becomes fixed in the population while the other is lost, with the alleles' initial frequencies determining their fates (i.e., the more abundant allele tends to be fixed). This bi-stable, switch-like behavior is what makes underdominance appealing as a gene drive mechanism.

In engineered underdominance, the fit homozygotes are generally envisioned as possessing either two wild type or two transgenic alleles, with their heterozygous offspring having reduced (or zero, in the case of extreme underdominance) fitness. This requires that we be able to bring about a situation in which a transgenic allele in the presence of a wild type genome, in a heterozygote, is unfit, while transgenic homozygotes (which lack any copies of the corresponding wild type sequence at this locus) are fit (Matzen 2012). One way to engineer such behavior is to create a system of pairs of unlinked lethal genes (toxins), each of which is connected with some repressor of lethality (antidote) associated with the expression of the other lethal gene (Figure 2.1B; Davis et al. 2001; reviewed in Sinkins and Gould 2006). In such a system, only "homozygous" engineered individuals, which are really transheterozygotes carrying both engineered lethal gene/repressor elements, are

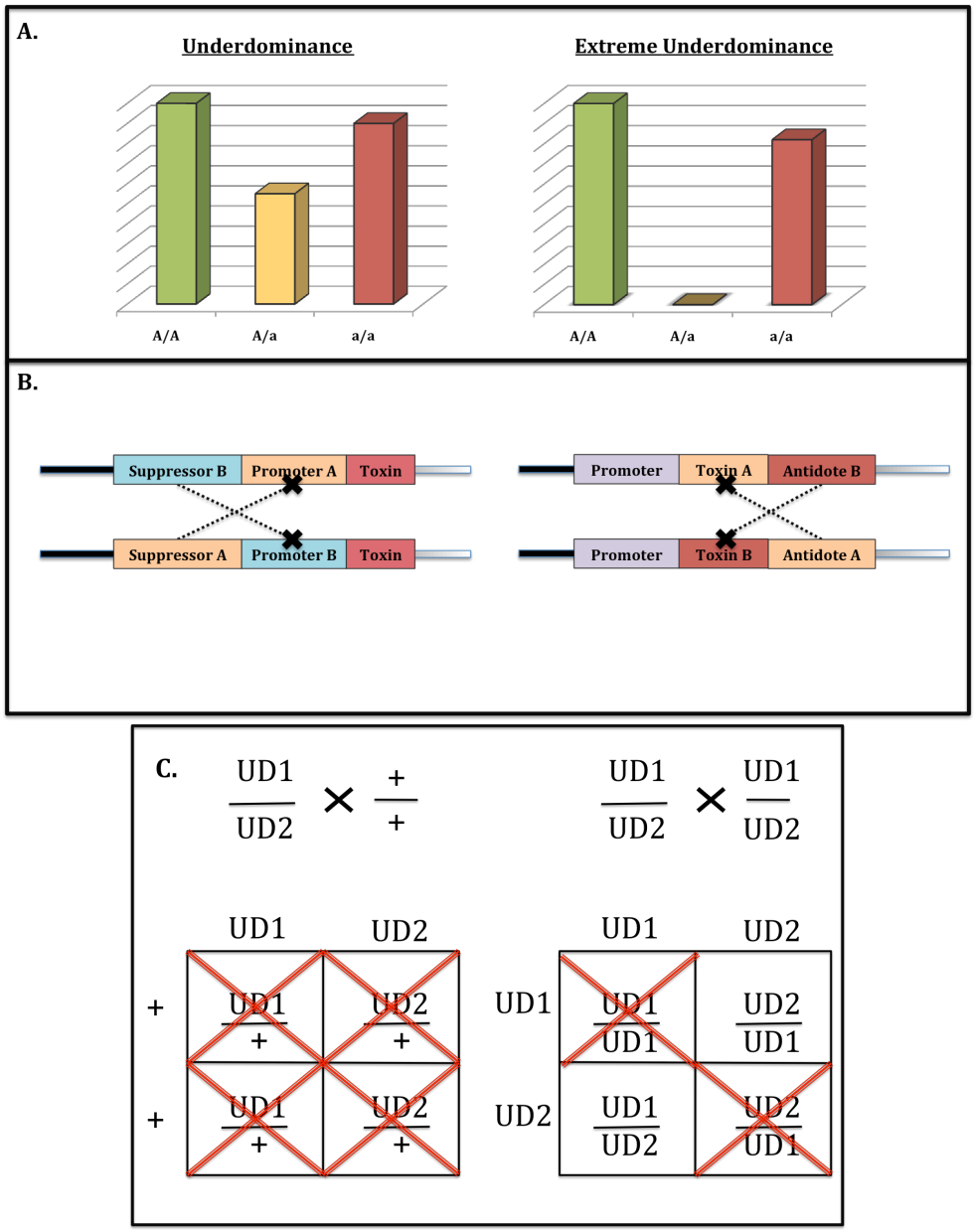

Figure 2.1: A. Underdominance is a case of heterozygous disadvantage, where the fitness of the heterozygote is lower than that of either homozygous parent. In extreme underdominance, the heterozygote fitness is zero, i.e., it is not viable. B. Two ways of engineering underdominance. On the left, two different promoters expressing the same toxin and the promoters' respective suppressors make up the underdominant system (Davis et al. 2001). On the right, the underdominant system is comprised of the same promoter driving two toxins and their antidotes. C. A single-locus extreme underdominant system. Crosses to wild type (+) produce no viable progeny, as only UD1/UD2 transheterozygotes survive. Self-crosses result in 50% viable offspring.

viable (Figure 2.1C). The original system proposed by Davis et al. 2001 relied on each allele of the pair containing a different repressible promoter expressing a toxin, and the repressor for the promoter of the opposite pair. Therefore, only an individual that had both alleles would have the toxins fully suppressed and would be viable. However, identifying promoter/repressor pairs that are strongly expressed and yet not at all leaky can be difficult (Matzen 2012). Because of this, the work detailed here involves a streamlined system that does not involve repressible promoters, but rather relies on expressed toxins and antidotes.

Conceptually, such a system can be designed in one of two ways. There are a number of proteins that, when expressed in sensitive tissues, can cause whole-animal death, and thus can be used as the toxin portion of the underdominant allele. Small RNAs targeting the toxic protein can be used as an antidote. If two slightly different versions of the toxic protein can be recoded so as to be sensitive to one set of small RNAs but not to another, then linking protein version A with RNAs targeting version B in one transgenic allele, and vice versa in the other allele, should produce a two-allele system that depends on the presence of both alleles for animal survival. Another way of engineering toxin/antidote systems based on expressed transcripts relies on using some means, such as small RNAs, to repress the expression of an essential gene as a toxin. Expressing a rescue version of the protein that is insensitive to the RNAs can serve as an antidote, as long as that antidote can be split between two alleles. With either approach, it should be possible to engineer transgenic underdominant individuals that cannot successfully reproduce with wild types.

Drive characteristics

In the underdominant systems described above, two different transgene-bearing alleles are utilized. These can be implemented in two different configurations: single-locus and two-locus underdominance (Figure 2.2A). For single-locus underdominance, both engineered alleles are placed at the same location in the genome, and mating to wild type produces either unfit heterozygotes (underdominance) or no viable progeny (extreme underdominance). In two-locus

underdominance, the two underdominant alleles are located on different chromosomes. In this system, individuals bearing four transgenes, two copies of each, can be generated. When these are out crossed to wild type all progeny are viable, but further outcrosses result in lethality for some offspring.

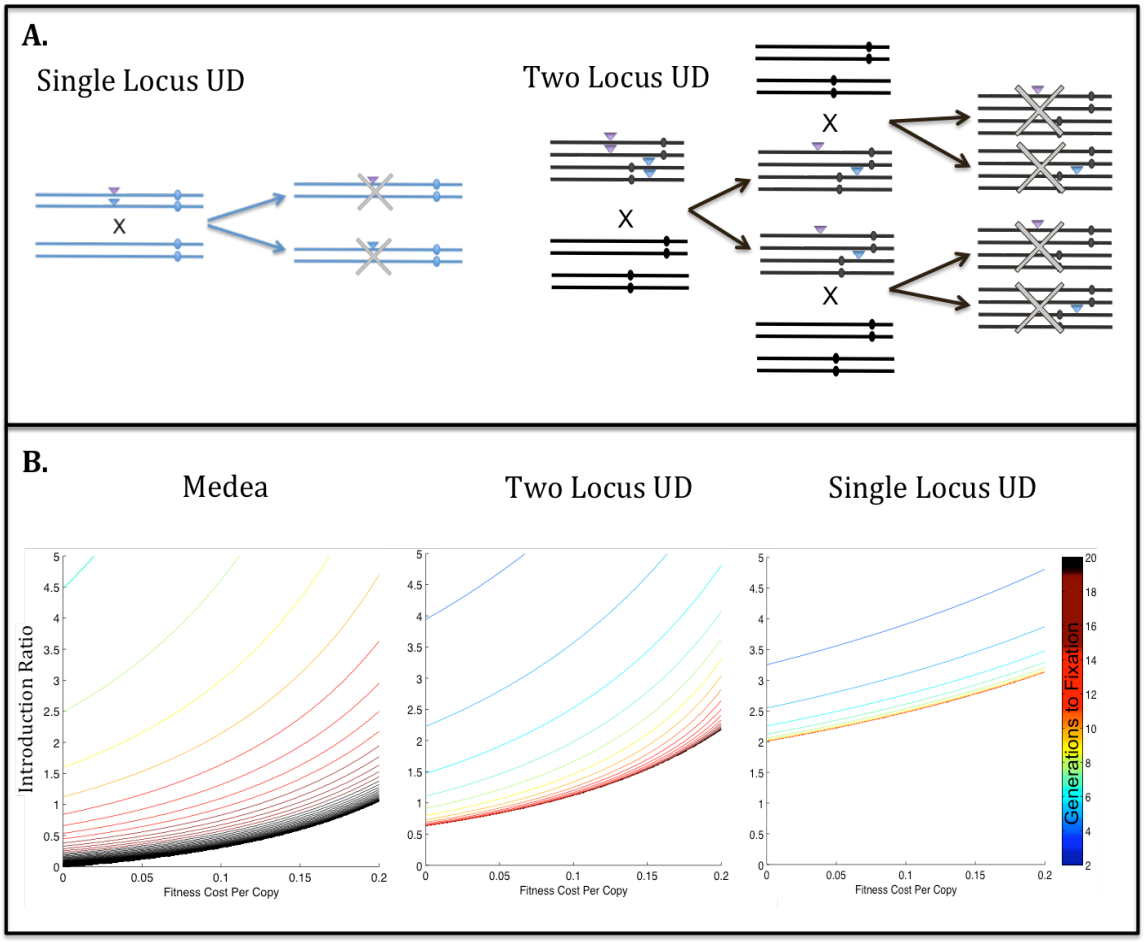

Figure 2.2: A. Outcomes of single-locus and two-locus extreme underdominant outcrosses to wild type. For the single-locus case, outcross to wild type produces no viable progeny. For the double-locus system, every offspring in the first generation survives, but further out crossing to wild type leads to some lethality in the second generation. B. These graphs show drive dynamics of Medea, single locus underdominance, and two-locus underdominance given initial release ratios and fitness costs. The area under the lowest line for each graph represents conditions where drive will not occur. C. Ward generated the graphs.

Single locus and two-locus underdominance have distinctive drive characteristics. In single locus underdominance many progeny die, even when the underdominant chromosomes are present at high frequency, because mating to wild type always results in loss of underdominant chromosomes. In contrast, as two-locus underdominance chromosomes spread, the fraction of progeny that die approaches zero. This occurs because as the alleles spread, the probability that progeny inheriting one underdominant allele also inherit the other increases. Generally, single-locus underdominance is not considered practical as a drive mechanism (Curtis 1968), largely because for fixation of a transgenic allele to occur, it must be introduced at very high frequency—greater than 67% (Davis et al. 2001). However, extreme single-locus underdominance (where the heterozygous state is 100% lethal) can be used to bring about reproductive isolation, which may be useful in development of GMO safety mechanisms designed to preserve the integrity of wild populations and thus alleviate public fears regarding transgenic organisms (Marris 2010; Pardo et al. 2009; e.g., containment of genetically modified crops, Kwit et al. 2011), as well as for studies of speciation (Moreno 2012). Two-locus underdominance requires a population frequency of 27% to spread (Davis et al. 2001), thus necessitating a smaller initial release and making it a viable candidate for gene drive. Both types of underdominance still possess lower thresholds than some of the release ratios utilized in sterile insect techniques, which can be as high as 90% (Krafsur 1998). In addition, as noted in Chapter 1, underdominant drive systems are self-perpetuating, while sterile males must be released on an ongoing basis.

Underdominant systems have a considerable introduction threshold. While this might be considered a disadvantage if the goal is to spread transgenes far and wide, in fact it has many benefits, especially when population replacement is in testing phases, or when public acceptance of transgenic release is not universal (Marshall et al. 2010a). Firstly, accidental releases of insects at sub-threshold levels are unlikely to lead to transgene spread, making confined release possible (Marshall and Hay 2012). Furthermore, while releases above threshold would be expected to lead to local spread, the introduced transgenes are unlikely to spread to high frequency in

nearby populations, since the threshold needed for fixation is not likely to be crossed via low-level migration (Altrock et al. 2010; Altrock et al. 2011). Also, introduced transgenic alleles can be removed from the release site through the sustained release of wild type individuals, since diluting transgenic allele frequency below the threshold will cause the alleles to fall out of the population (Altrock et al. 2010; Marshall and Hay 2012). Finally, the ability to reproductively isolate transgenic individuals from wild types can be crucial to securing greater public acceptance for transgenic releases and protecting the integrity of native populations. Thus, development of engineered underdominance would serve as an important tool for furthering disease control strategies based on refractory insect release, and potentially for assuring reproductive isolation of genetically modified organisms.

Engineering UD: An Essential Problem

To generate transgenic underdominant individuals of the kind discussed above, one has to bring two underdominant alleles into the genome of the same individual. The essential problem with this approach is that each allele is supposed to impart a significant fitness cost and/or death (for extreme underdominance) when present without its partner allele, yet the alleles can only be introduced individually.

To get around this issue, one has to create underdominant alleles in which the toxin function can be temporarily and reversibly suppressed until both alleles are present in the same animal. This can be achieved in several different ways. Promoters that are off in the presence of some substance (for example, tetracycline) can serve as this required conditional rescue: animals can be reared on tetracycline to suppress toxin expression until both underdominant alleles are crossed into the same individual. Another way to provide conditional rescue is to separate the chosen promoter from the toxin in such a way that no toxin is expressed until the separating factor is purposefully removed. Inserting a blocker sequence between toxin and promoter can serve this purpose; and if specific recombinase sites are designed to surround the blocker, expressing the recombinase in animals that have both underdominant alleles can catalyze the blocker's permanent removal.

Whatever the approach used, designing a conditional or reversible rescue is essential to any attempt to engineer underdominance.

Engineering UD: Additional Considerations

Another prerequisite for the creation of single-locus underdominant individuals is the capability to insert two different alleles at the same genomic position. In Drosophila, this can be achieved by using the PhiC31 system (Groth et al., 2004). Briefly, this system relies on the PhiC31 integrase, derived from bacteriophage, to carry out recombination between two different sites, attP and attB (Thorpe et al., 2000). Since there is a multitude of existing Drosophila lines with attP sites in different parts of the genome, one can use a plasmid containing the complementary attB site to insert different alleles into the same genomic locus.

The PhiC31 system is utilized extensively in the Hay lab, and for good reason. It is very efficient, making transgenic fly generation straightforward and reliable. It also allows for controllability—different transgenic alleles can be compared in the same genetic background, eliminating the need to calibrate for potential position effects. However, as we found out recently, when my thesis research was nearing completion, there is a caveat to using PhiC31 lines to engineer underdominant systems. A number of the lines (including most of the ones we've used for this project) have a fitness cost in that the females have significantly reduced fecundity, sometimes laying half as many eggs as the white minus stock typically used as wild type in the laboratory. While this does not affect our ability to generate underdominant systems and test their functionality, it does prevent us from doing drive experiments against wild type flies (since our transgenic individuals are affected by a significant fitness cost unrelated to their underdominant alleles). Thus, all drive experiments described in this work involve testing our underdominant populations against the empty attP insertion lines used to generate them. This is by no means invalidating—development of any drive mechanism in Drosophila melanogaster serves simply as proof of principle, since it is not a pest species and there is no interest in transforming wild populations via gene drive. However, my

work does highlight that fact that all future attempts to transfer this system to vector species of interest should evaluate potential fitness costs of planned genomic insertion sites.

In the studies that follow, I describe the development and optimization of a conditional rescue system meant to facilitate the creation of underdominant Drosophila individuals. I also discuss the generation of an underdominant Drosophila population based on the protein toxin approach.

2.2 Results and Discussion

2.2.1 Conditional Rescue

Blocker Sequence

In order to attempt creation of any underdominant (UD) systems, I had to have a method of conditional rescue that blocked toxin expression well enough to allow generation of viable heterozygotes, and could be easily and efficiently removed. Previous work on underdominance in the lab carried out by Kelly Matzen utilized both tetracycline-induced expression and presence of a removable blocker sequence to prevent toxin expression until desired. However, Kelly's work showed that tetracycline-inducible systems were difficult to clone and tricky to calibrate (Matzen 2012), so all of my work relied on using a blocking sequence originally tested by Kelly (Figure 2.3A). The sequence—usually GFP and a 3' untranslated region (UTR)—is flanked by FRT recombinase sites, and can be excised by the addition of FLP, a site-directed recombinase (Zhu and Sadowski 1995). Not only does such a sequence separate the promoter used from the toxin open reading frame (ORF), it serves as a visual marker for the unflipped versus flipped state.

FLP-Expressing Fly Lines

In order to remove the blocking sequence, a source of FLP must be crossed into UD allele-bearing flies. If a germline promoter-driven FLP was introduced into flies that had unflipped UD constructs, FLP would catalyze the removal of the blocker in those

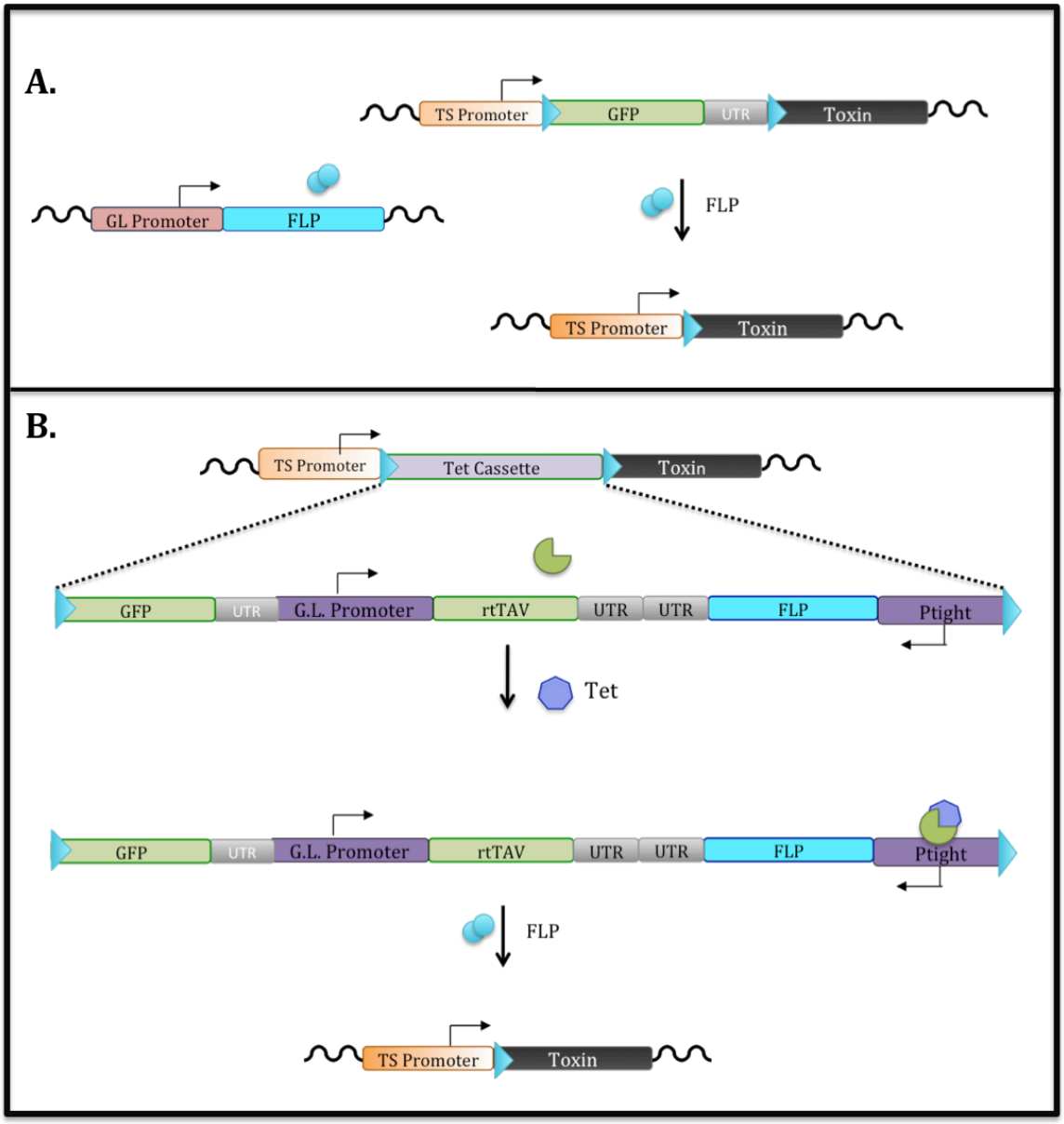

Figure 2.3: A. Schematic of FLP-catalyzed excision of the removable rescue blocker. When the blocker is present, GFP is expressed instead of the toxin. After addition of FLP, the blocker is excised and the toxin is expressed. FRT sites are represented by blue triangles. B. Schematic of the tetracycline (Tet) blocking cassette. A germline-specific promoter is driving the expression rtTAV. Upon addition of Tet, Tet and rtTAV induce expression of FLP from P_{tight} , which then catalyzes the blocker's excision.

flies' progeny. Two FLP lines, both obtained from the Bloomington Stock Center (Bloomington, IL), were used in the lab for this purpose—male-germline-specific ß-tubulin-FLP and female-germline-specific ovo-FLP. They were, however, suboptimal: neither resulted in high frequency germline excision of our blocking sequences; the ovo-FLP line produced very inconsistent flipping results; and the ß-tubulin-FLP line had a very dark eye color, which made it difficult to disambiguate from combinations of UD alleles that were also marked with a red eye color. This made the process of blocker removal cumbersome and unnecessarily complex.

To address this issue, in collaboration with Kelly, I created and tested novel promoter-FLP lines in hopes of developing a highly efficient blocker removal system. Several newly characterized germline promoters were obtained from O. Akbari and used to drive FLP. The first iteration used a vasa promoter fragment, which drives expression in both male and female germlines; this gave no observable flipping at two attP insertion sites. At this point, we also discovered that there were two versions of FLP, and that we had been using the one with a low flipping efficiency (Nern et al. 2012). The next set of constructs utilized the more efficient form of FLP, and tested four different promoters: a long fragment of the Exu promoter meant to express in males (ExuL), a medium Exu fragment meant to express in females (ExuM), and two very strong female-specific promoters, Deadhead (Dhd) and Bicaudal C (BicC).

The constructs were evaluated at one or two attP sites (9753 and 9724, or 9724 only), and the flipping efficiencies observed for the more efficient of the sites (9724), as well as the efficiency for the previously described FLP lines, are listed in Table 2.1. The variable flipping rates observed with the ExuL promoter seemed correlated with the physical distance between the FRT sites: larger blocker sequences (~ 5 kb+) were removed with lower efficiency, while minimal blockers (~ 1.5 kb) were excised with high efficiency. This is not surprising, as efficiency of FLP recombination is known to decrease with increasing distance between FRT sites (Ringrose et al. 1999). However, since most of the blockers used in subsequent underdominance work were small, this line was efficient enough to be used for all future male-specific blocker excision.

Promoter	Flipping Rate
β-tubulin (male-specific)	60-90%
Ovo (female-specific)	0-60%
Vasa (male and female)	0%
Exu long (male-specific)	50-100%
Exu medium (female-specific)	42%
BicC (female-specific)	100%
Deadhead (female-specific)	98%
Deadhead male	36%

Table 2.1. Efficiencies of various germline promoter-FLP constructs. Flipping efficiencies were calculated by counting flipped versus unflipped progeny, as evidenced by a color change from GFP to dsRed. All generated FLP constructs were tested at the same genomic insertion site.

The BicC and Dhd promoters not only resulted in nearly 100% efficient flipping when present in the germline of parents with unflipped alleles, but could flip constructs in trans (i.e., if a male with an unflipped allele is crossed to a BicC-FLP or Dhd-FLP female, the resulting progeny are all flipped). This simplifies testing for killing by various toxins a great deal, since the toxin-bearing unflipped parents don't need to have a germline source of FLP crossed in before the blocker can be removed. Furthermore, all of the generated efficient lines are at an attP site marked with light orange eye color, and thus can be easily told apart from UD individuals. Taken together, the above fly lines provided everything I needed to attempt engineering underdominance, and were essential to this project.

B3 Recombinase

During my work on various underdominant systems, I tested an approach that utilized FLP as part of the rescue (described in detail in Chapter 3). Because FLP had to be present in the underdominant construct, it could not be used to remove the conditional rescue blocker. To be able to employ the removable blocker system, I needed to use another recombinase that did not cross-react with FRT sites. To this end, I tested the B3 recombinase from Zygosaccharomyces bisporus, which was shown to work well in Drosophila (Nern et al. 2012), with the ExuM, ExuL, BicC, and Dhd promoter fragments. The construct designs were identical to the FLP ones described above, except that FLP was replaced by B3, and removable cassettes were flanked with B3 recombinase sites. The efficiency of cassette removal using this system was almost identical to that seen with FLP; the only exception was ExuM, which gave lower levels of blocker excision (< 20%) in females. The generated promoter-B3 lines allowed me to work on the FLP-containing underdominant system, and can be utilized in the future if FLP/FRT recombination cannot be used.

Tetracycline Cassette

While the tools for conditional rescue removal described above worked well, ideally one can effect blocker excision without needing any further fly crosses, e.g., via addition of a drug. This would be especially useful if underdominant systems were transferred to a vector species like the mosquito, where generation of germline FLP lines may pose a considerable challenge.

To this end, Kelly and I designed an FRT-flanked blocker cassette that would self-excise upon the addition of tetracycline (Tet; Figure 2.3B). Besides the usual GFP blocker, this cassette contained rtTA2^{S}-M2 (rtTAV) driven by a germline promoter, and a tetracycline-response P_{tight} promoter driving FLP. In the presence of Tet, rtTAV—a prokaryotic Tet operon-binding domain fused to minimal repeats of the VP16 activation domain (Matzen 2012) —binds and induces expression from the P_{tight} promoter. Thus, rtTAV is always present in the germline, but only in the presence of Tet should it cause FLP to be expressed from P_{tight}, and to catalyze the removal of the cassette.

The initial version of this cassette, which used a vasa germline promoter, the original P_{tight} promoter, and the inefficient allele of FLP, did not give any flipping. I first replaced the FLP with the efficient version and tried using the BicC promoter instead of vasa, since we knew it was strongly expressed in the germline. However, females bearing this cassette were infertile, presumably because very high levels of rtTAV in the germline were toxic. It also came to our attention that the P_{tight} promoter, while adequate for somatic expression, may lack the elements to be sufficiently expressed in the Drosophila germline.

To find both an appropriate germline promoter and a version of P_{tight} that would function well in the germline, I decided to test the components with a GFP induction system. A set of constructs was generated that had different germline promoters driving rtTAV, as well as the original P_{tight} promoter or one of two modified versions driving GFP. When inserted into the Drosophila genome, these constructs should not express GFP without the presence of Tet, but should show germline GFP expression when Tet is added. Using this system, I tested two germline promoters, Dhd and zpg (both provided by O. Akbari), and the original P_{tight} plus two novel versions, for a total of six constructs. The original P_{tight} promoter used contained seven TetO operator sequences followed by a core CMV promoter; the versions I designed had nine TetO operator sequences followed by either the basal hsp70 promoter (TetO-hsp70) or the P transposase minimal promoter (TetO-P_{el}), which is known to work in the Drosophila germline (Rorth 1998). The original P_{tight} and TetO-hsp70 did not appear to function at all in the female germline, while TetO-P_{el} was readily inducible in the presence of Tet and rtTAV (but gave no GFP expression in their absence). Both the Dhd and zpg promoters induced robust expression of GFP from TetO-P_{el} in fly ovaries in the presence of Tet, but Dhd was also leaky, with some GFP observable in ovaries without Tet. Thus, zpg and TetO-P_{el} were chosen for testing in the full Tet cassette. No GFP expression was ever detected in male testes.

To avoid dealing with any toxins or other complex components, I cloned the modified Tet cassette (with zpg and TetO-P_{el}) between a Ubq promoter fragment (provided by O. Akbari) and dsRed. With this architecture, a color switch from GFP (unflipped) to dsRED (flipped) is expected if flipping occurs. This construct was

injected into two attP sites (86Fa and 51D), and the blocker was efficiently flipped out if a source of germline FLP was crossed at both insertion sites. When dosed with Tet, the construct did catalyze its own removal, but at a fairly low efficiency (12-15%, depending on the site). Although further optimization of this system has not been carried out, and the final version of this cassette was not used in underdominant constructs, this work shows that such drug-inducible conditional rescue removal is possible.

2.2.2 Underdominance with a Protein Toxin

The underdominant system described here uses Hid, a Drosophila RHG cell death protein, as a toxin, and was originally designed and tested by Kelly (Matzen 2012). The Hid^{Ala3} allele (Bergmann et al. 1998) is recoded at the DNA level to produce two slightly different versions, each of which can be targeted by a specific set of synthetic small hairpin RNAs (shRNAs). Each shRNA set can target only its assigned allele, and not the other allele or endogenous Hid. The Hid alleles, arbitrarily labeled 1 and 2, are paired with the shRNA set that targets the opposite allele (i.e., Hid1 with shRNA set 2 on one construct, Hid2 with shRNA set 1 on another) and expressed in an essential tissue. In this way, both constructs need to be present in order for Hid to be fully suppressed, and individuals carrying only one of the constructs should not be viable (Figure 2.4A).

This system was initially tested in fly eyes using the GMR promoter, and resulted in both death (small eye phenotype in flies with only one construct) and rescue (restoration of normal eyes in flies with both constructs). It was then moved to several essential tissues, and expression in the heart, under the hand enhancer (Sellin et al. 2006), produced strong (90-95%) killing of individuals with only one underdominant allele at several attP sites. However, rescue in the heart was never obtained, and Kelly completed her thesis studies before she could figure out why.

Achieving Rescue in the Heart

The architecture of the Hid constructs I began working with is shown in Figure 2.4B. In these, the hand enhancer drives the expression of Hid and the rescue shRNA set,

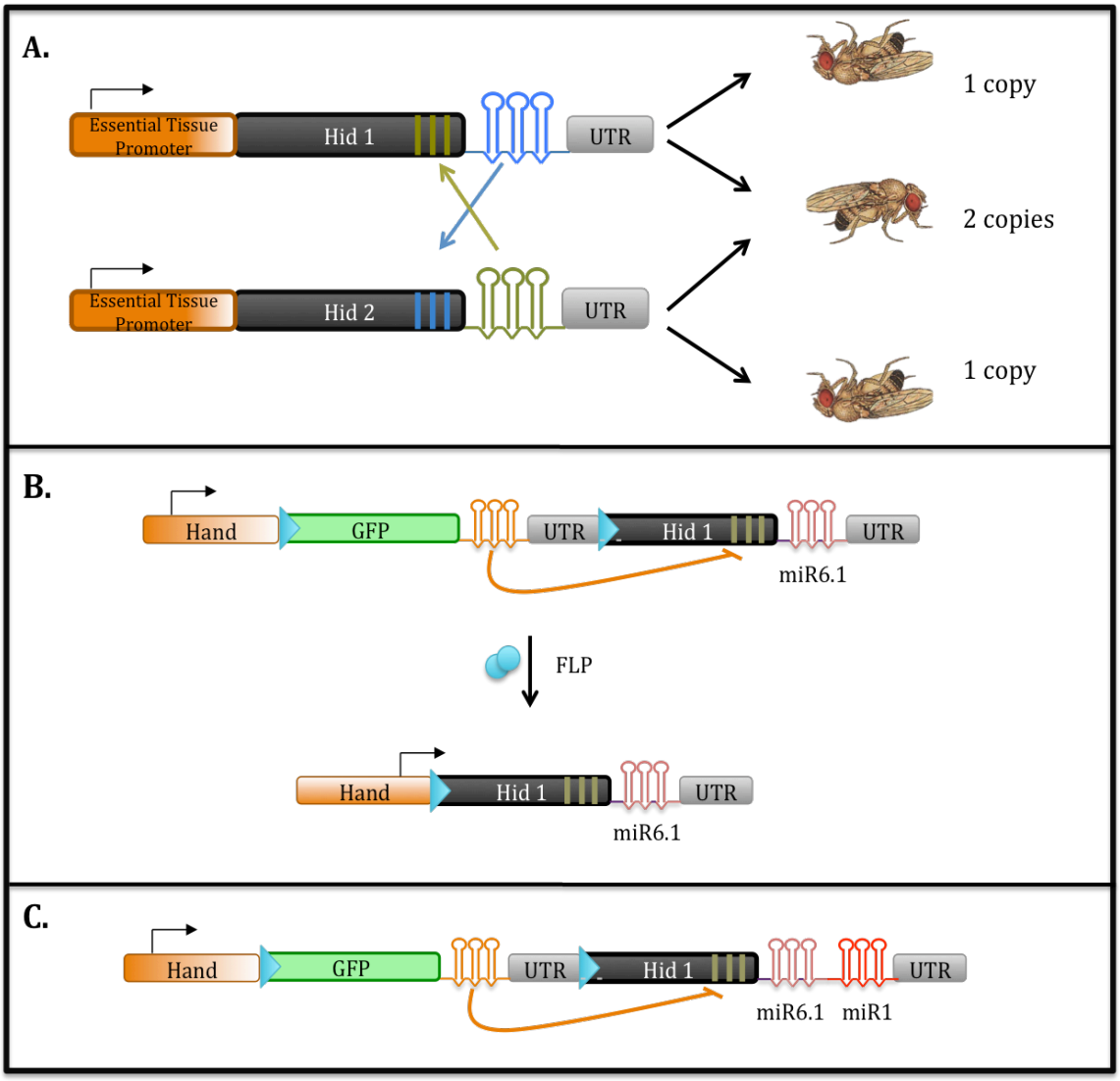

Figure 2.4: A. Schematic of the protein toxin underdominance system using Hid expression in an essential tissue. Each construct has one version of Hid (the toxin) and the version of shRNAs targeting the opposite Hid (the antidote). When only one construct is present in an animal, Hid is not suppressed and the animal dies; when both constructs are present, the animal is viable. B. Architecture of initial Hid constructs that killed but didn't rescue. The GFP blocker prevents expression of Hid; upon the addition of FLP, the blocker is excised and Hid is expressed in the heart. C. The final construct (with additional miR-1-based shRNAs) that gave both killing and rescue.

which is placed in an intron in the Hid 3'UTR. The blocker sequence, flanked by FRT sites, consists of GFP and the opposite set of shRNAs in the GFP 3'UTR (this second set targets the Hid version in the construct, and is included as part of the conditional rescue to prevent Hid leakage during transformant generation). Following flipping, this conditional rescue is removed, and Hid is expressed in the heart.

As Kelly's results suggested, I found that these constructs gave good killing at one attP insertion site (51D), although the precise level of killing was difficult to gauge because hand-driven GFP was not visible at this site, and it was thus impossible to tell whether the surviving heterozygotes had the conditional rescue blocker flipped out. Transheterozygous animals could not be obtained, however, even though the rescue appeared to function adequately in the eye.

The triplets of shRNAs used for the rescue were based on the Drosophila miR-6.1 backbone, which was known to work well in the lab (e.g., Chen et al. 2007). However, it wasn't clear if this backbone was ideal for driving down expression in muscle tissue. I decided to try making rescue shRNAs based on miR-1, engineered versions of which have been shown to effectively knock down a variety of Drosophila genes (Ni et al. 2011), in the hopes of improving rescue efficacy. These new shRNAs targeted the same sites in Hid, and each set consisted of an shRNA triplet in an intron of the Hid 3'UTR, just like the original version, but their backbone was based on the miR-1 design described in Ni et al. 2011. The new rescue sets were added to the Hid 3'UTR in addition to the original shRNAs (Figure 2.4C).

When tested at attP site 51D, these new constructs gave the same strong kill observed with the previous version (as would be expected, since the toxin wasn't altered), but also effected rescue. Transheterozygous underdominant individuals were obtained at expected frequencies and appeared healthy and robust, and an underdominant fly stock was created for use in future gene drive experiments.

Improving Killing

Although an underdominant population was generated in the above experiments, the ultimate goal of this project was to engineer extreme underdominance (with no heterozygote survival) so that reproductive isolation could also be achieved. Since

the Hid system was working quite well already, I set about trying to improve the toxin's function to achieve 100% kill. Pfeiffer et al. 2012 showed that addition of certain translational enhancer (TE) sequences can significantly boost transcript expression levels, so I added a 5' MHC IVS16 intron and the p10 3'UTR (from Pfeiffer et al. 2012) to Hid, but saw no improvement in killing. (If anything, I saw a slight decrease of killing; this was thought to potentially stem from splicing machinery being overwhelmed by having too many different introns to process).

Another approach to potentially improve killing involved trying new promoters to drive Hid expression. I tested two other heart-specific enhancers: the minimal hand enhancer (handm; Han and Olson 2005), which was supposed to give stronger expression in the heart than the full-length version of hand, and the tinman enhancer (tinC; Lo and Frasch 2001). I also tried two enhancers that express in oenocytes, which regulate water tightness of the trachea and are required for larval viability (Parvy et al. 2012): desaturase 1 (dst1; Billeter et al. 2009) and a seven-up enhancer (svp; Gutierrez et al. 2007). However, none of these enhancers improved killing (Table 2.2). The handm enhancer produced comparable killing to the full-length version, and was used for all subsequent work in the heart; tinC resulted in a weak kill; and neither of the oenocyte enhancers generated any observable killing.

Promoter	Killing	Rescue
hand original	Strong	Yes
hand minimal	Strong	Yes
tinC	Weak	Yes
dst1	None	—
svp	None	—

Table 2.2: Fly results from Hid constructs driven by various promoter fragments. All flies were scored at the same genomic insertion site (51D).

Drive Experiments

Gene drive experiments with the one underdominant Hid line generated in this work were initially set against white minus (w-; the laboratory equivalent of wild type). Since this was a functional single-locus underdominant system, drive could theoretically be expected if the frequency of the initial underdominant release was greater than 67% of the population (i.e., 67% underdominant to 33% wild type; Davis et al. 2001; Sinkins and Gould 2006). However, even at significantly higher release ratios (e.g., 80%), most of the experimental replicates showed rapid loss of transgenic individuals from the population (Figure 2.5). (There were two replicates that showed spread of the transgenic alleles when set at an 80% release threshold; the reasons for this are unclear.) We assumed this stemmed from either incomplete killing of heterozygotes, incomplete rescue resulting in fitness costs to the transheterozygotes, or both. At this point, I set aside this project to focus on underdominant systems with a small RNA toxin.

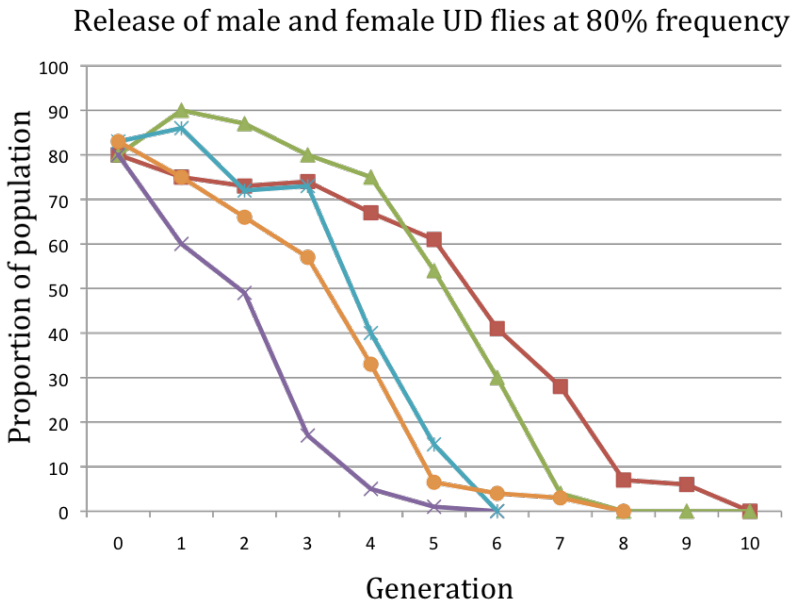

Figure 2.5: Drive experiments were set against white minus stock by releasing male and female Hid underdominant individuals at 80% frequency. Each line represents a biological replicate. Most replicates showed rapid loss of Hid underdominant alleles out of the population.

However, as mentioned in the introduction, our lab has recently discovered that many of our regularly used attP lines have intrinsic fitness costs that disadvantage them in competition with w-. Thus, in order to obtain a fair comparison, any drive experiments involving transgenics created at these attP lines have to be done against the empty attP lines themselves. Hid drive experiments were therefore set up again using the empty 51D attP line, and are currently ongoing. Initial results show that, when released at high threshold, the transgenes quickly spread to fixation (Figure 2.6). The drive actually appears too forceful: even when released below the frequency predicted to lead to fixation, the underdominant alleles rapidly increase instead of falling out of the population.

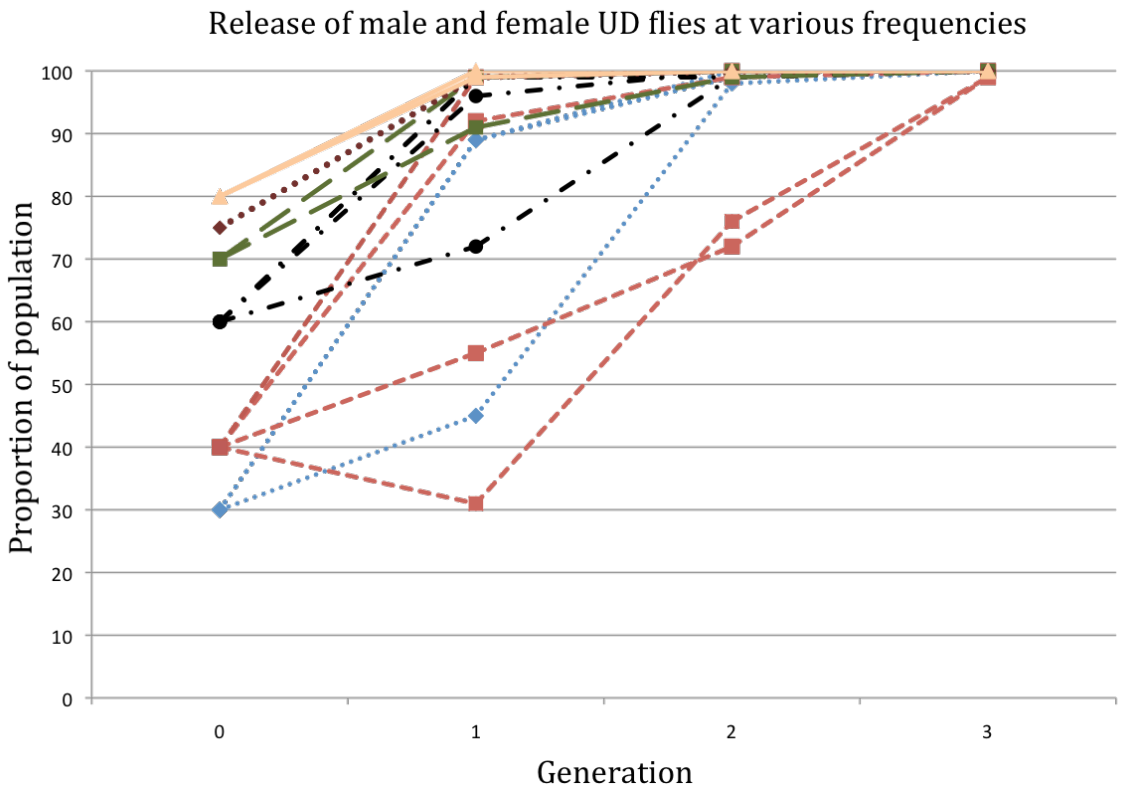

Figure 2.6: Drive experiments were set against the empty attP insertion line used to generate Hid transgenics by releasing male and female Hid underdominant individuals at various frequencies (from 30%-80%). Each line represents a biological replicate. Most replicates showed rapid spread of Hid underdominant alleles to population fixation, even when released at frequencies below the predicted drive threshold of 67%.

The exact causes of this deviation from expected drive dynamics are unclear. Differences in visual capabilities between the Hid and attP individuals may play a considerable role: the attP flies lack a functional *white* gene and thus have white eyes, while the Hid underdominant flies are red-eyed, as *white* is resupplied as part of the underdominant allele (to facilitate screening for transgenics). White-eyed flies have significantly impaired vision compared to red-eyed ones, especially in high light intensities (Borycz et al. 2008; Krstic et al. 2013), and since the above drive experiments were carried out in brightly lit incubators, the inferior visual ability of the attP individuals may have imposed a substantial fitness cost on them. This would facilitate accelerated spread of the underdominant alleles, since their carriers would be at an advantage. To address this possibility, drive experiments have now been re-set up in the dark (where the white-eyed flies are not at a great disadvantage) to see whether that changes drive dynamics.

Another possible explanation for the unexpected drive characteristics of the Hid system has to do with unforeseen fitness costs of the attP line as compared to the Hid population. We do not know what imposes the fitness costs observed in the attP lines themselves: they may be associated with the insertion site itself (and any associated marker genes), with deleterious recessive alleles present in the stocks, or with both. If there are significant fitness effects stemming from recessive mutations, the transgenic Hid flies might have had some of them recombined and/or selected out, as they had been integrogressed with w- and other lines during the process of transformant generation and blocking sequence removal. Should this be the case, attP insertion individuals can be outcrossed to w- for several generations and then homozygosed again to create a stock that can be used for future drive experiments. Together, the above strategies should make it feasible to show drive in and drive out at expected frequencies with the Hid system, and experiments to implement them are ongoing.

2.3 Future Directions

Conditional Rescue

The minimal conditional rescue blockers and the recombinase lines used to remove them described here function well enough to allow efficient testing of underdominant systems. Nevertheless, development of a rescue cassette that self-excises upon drug addition would further simplify underdominance work in Drosophila, and would provide a blueprint for creating such a system in vector species. The fact that the Tet cassette detailed in this work functions as expected is a promising starting point. However, its efficiency of drug-induced self-excision is low, and would have to be improved for it to be useful.

One way to make this cassette more efficient is to increase expression levels of FLP and rtTAV, potentially by adding the translational enhancers described earlier. Although they did not improve Hid-induced killing, they do seem to increase expression in tissue culture constructs (Hay lab, unpublished data), and may function better when present without other intronic elements. Since this work was carried out, optimized versions of both rtTAV (rtTA-V10) and P_{tight} (P_{TRE3G}) have also become available (Zhou et al. 2006; Clontech, Mountain View, CA), and may improve Tet-induced FLP expression. Furthermore, other recombinases, such as B3, can be tested to see if greater recombination efficiency can be achieved. All together, some combination of the above improvements may well result in a Tet cassette that is effectual enough for practical use.

Underdominance with a Protein Toxin

The Hid underdominant system already functions well—it has a robust rescue and a strong kill. An ideal system, however, would bring about reproductive isolation, and it might be worthwhile to attempt further improvement of the Hid kill to achieve this goal. One obvious target for modification is the position of the shRNA rescue. Recent work in our lab has shown that placing shRNAs in the UTR of a gene significantly reduces its expression level. Moving the shRNAs to a separate part of the construct and expressing them under the same promoter used to drive Hid will

complicate construct design, but it may lead to significantly increased Hid levels, and thus improved killing. If the intronic shRNAs are moved away, it may also be reasonable to re-test Hid with translational enhancers, in case their previous lack of function was really caused by too many introns on one transcript. Additional enhancers can also be tested to see whether the system will operate in other essential tissues, and the plethora of transcriptional profiling data currently available can guide selection of enhancers likely to work well with this system.

Finally, since insertion at different genomic positions has been shown to greatly affect killing in earlier Hid work, further attP sites can be tested to see if a more optimal site than the one currently used can be found. Preferably, this would be a site with no associated fitness costs, so that drive experiments against wild type can be performed. This would eliminate the issue of the stronger-than-predicted drive observed when performing releases against the current attP line. However, if drive against wild type cannot be achieved, it will be important to carry out drive experiments against such populations that allele fixation and loss can be achieved at theorized frequencies, since the switch-like, threshold-dependent property is a crucial component of underdominant drive. The approaches described earlier will hopefully elucidate a way to conduct such experiments.

2.4 Materials and Methods

Fly Culture and Strains

Fly husbandry and crosses were performed under standard conditions at 25°C. BestGene Inc. (Chino Hills, CA) and Rainbow Transgenics (Camarillo, CA) carried out all of the fly injections, with Rainbow Transgenics performing the vast majority of them with consistent reliability.

Additional fly strains utilized in this study were ovo-FLP (Bloomington Stock Center #8727), ß-tubulin-FLP (#7196), and attP lines 86Fa (#24486), 51D(#24483), #9724, and #9753.

Fly crossing schemes used to introduce FLP into the germline and set up kill and rescue crosses are detailed in Appendix C.

To test tetracycline-induced GFP expression and Tet cassette excision, Tet food was made. Standard fly food was microwaved until liquefaction and allowed to cool to ~ 45-50°C. Then a saturated Tet solution in 70-100% ethanol was added to the food at a 1:100 ratio, and the food was poured back into standard vials to solidify. Since Tet is degraded by light, food vials were stored in the dark, usually covered by tin foil, at 4°C. To induce cassette excision, day-old flies were placed on Tet food, and 10-30 μL more of the Tet solution was added directly to the food every two or three days. The progeny of these flies were then scored for flipping, as evidenced by a switch from GFP to dsRed expression. To check for induction of GFP expression in the germline, flies were kept on Tet food as described above, or on regular food; ovaries and tested were dissected after ~ five days to check for GFP expression.

The process of setting up drive experiments is described in Appendix C.

Conditional Rescue Constructs

The germline-expressing FLP constructs were cloned into a multiple cloning site (MCS)-containing Drosophila attB backbone using the enzymatic assembly (EA) cloning method described by Gibson et al. 2009 (this procedure is detailed in Appendix A). FLP was PCRed from lab sources, and originally contained a glycine residue at position 5 of the protein. For subsequent constructs, the glycine was mutated to an aspartic acid via PCR to produce a more active form of FLP (Nern et al. 2011). The SV40 3'UTR was PCRed from lab sources, and cloned 3' of FLP. Germline enhancer fragments vasa, ExuM, ExuL, BicC, and Dhd were PCRed from constructs provided by O. Akbari or from Drosophila genomic DNA, and cloned 5' of FLP. Flipping efficiencies of the germline FLP lines were tested via crosses to the GFP/dsRed Tet cassette line described in this chapter (overall FLP cross schematics are detailed in Appendix C). B3 recombinase-FLP constructs were cloned the same way. B3 recombinase and B3 sites were PCRed from Addgene plasmids 32136 and 32139, respectively (Addgene, Cambridge, MA). Efficiencies of B3 recombinase lines were tested using underdominance constructs described in Chapter 3.

The GFP-induction constructs were cloned into the MCS of the attB vector using the EA method. To prevent any read through, they were designed so that rtTAV would be transcribed from the bottom strand relative to the Tet-responsive GFP, with the two transcriptional units separated by a Gypsy insulator (Chun et al. 1993; Gdula et al. 1996). The rtTAV coding sequence and original P_{tight} fragment were PCRed from the original Tet rescue cassette provided by K. Matzen. Each germline promoter was cloned (in the reverse orientation) onto the bottom strand of the attB plasmid, and was followed by rtTAV and then the p10 3'UTR (PCRed from UASt; Rorth 1998). The Gypsy insulator was PCRed from lab sources and cloned 3' of the germline promoters. TetO repeats were PCRed from the P_{tight} fragment in groups of three, and the nine-repeat multimer was originally assembled via Golden Gate cloning methods (described in Appendix A) into backbone OA669 (from O. Akbari). These were then moved into the above attB backbone 3' of Gypsy, followed by either the hsp70 basal promoter (PCRed from lab sources) or the P transposase minimal promoter (PCRed from UASp), then GFP and the SV40 3'UTR (both PCRed from lab sources).

K. Matzen assembled the original Tet-inducible cassette (Matzen 2012). It was flanked by FRT sites, and contained ttAV driven by the vasa promoter and FLP driven by the P_{tight} promoter. It was moved to the attB backbone using EA cloning, and assembled together with the paramyosin (prm) enhancer (Hess et al. 2007; Marco-Ferreres et al. 2005) 5' of the cassette, and dsRed (provided by O. Akbari) and tubulin 3'UTR (provided by K. Matzen) 3' of the cassette. In the final version, the prm enhancer was replaced by Ubq (provided by O. Akbari), vasa was replaced by zpg, FLP was replaced by a more efficient version, and the P_{tight} promoter was replaced by TetO-P_{el}. All of the above manipulations were done with EA cloning.

The FRT site, B3 site and TetO sequences, as well as the first and last 30 base pairs (bps) of all other DNA fragment used in the described constructs, are listed in Table 2.3. All transcribed proteins utilized had a Kozak sequence (CAACAAA) 5' of the first ATG codon added via PCR.

Name	Sequence
FRT Site	GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGA AAGTATAGGAACTTC
TetO	TCCCTATCAGTGATAGAGA
B3 Site	GGTTGCTTAAGAATAAGTAATTCTTAAGCAACC
Name	30 bps of the 5’ and 3’ ends of each fragment, 5’ to 3’
P transposase minimal	CCGTAGCTTACCGAAGTATACACTTAAATT CATTGGTACCCGCCCGGGGATCAGAATTGG
P_{tight} original	CGAGTTTACTCCCTATCAGTGATAGAGAAC CTCGTTTAGTGAACCGTCAGATCGCATCGA
Hsp70 basal	AGCGCCGGAGTATAAATAGAGGCGCTTCGT AAGCGCAGCTGAACAAGCTAAACAATCTGC
Gypsy	AATTGATCGGCTAAATGGTATGGCAAGAAA GAATAAATTCAACGCACACTTATTACGTGA
Zpg	ATGACCACCACACTGTGCGACAGACTGAGC CCGAAATTAAATAAAATAAATACGCCATCA
rtTAV	ATGTCTAGACTGGACAAGAGCAAAGTCATA GATTTTGACCTTGACATGCTCCCCGGGTAA
p10 3’UTR	CTAGAATGAATCGTTTTTAAAATAACAAAT TGGCTTGGATAGCGATTCGAGTTAACGGCC
Dhd	TTGCGGCCTTCTTGGATTGAAAAAAAAGTG AAACACTTTGCTAATCCTCGCGACATAAAA
SV40 3’UTR	ACATTGATGAGTTTGGACAAACCACAACTA TGTGGTATGGCTGATTATGATCAGTCGACC
B3 recombinase	ATGAGCTCGTATATGGATCTTGTTGATGAT CGGCGCGTCCTGGCACCAATTAACCGGTAA
ExuL	CTGTCATCTCTGGCGCAACCATTTGAAATT TAGATTGCTTTTCGCAGATCATATAAGAAA
ExuM	TTCATTACGTTTTAGTTAGTCGTACGAATT TAGATTGCTTTTCGCAGATCATATAAGAAA
BicC	ATAATTATATAATAATAAACTGCATGCCGC TGTGGAATTCGGATGATGATGATGATCACT
FLP	ATGCCACAATTTGATATATTATGTAAAACA TTATATGCGTCTATTTATGTAGGATGAAAG
GFP	ATGGTGAGCAAGGGCGAGGAGCTGTTCACC ACTCTCGGCATGGACGAGCTGTACAAGTAA
dsRed	ATGGTGCGCTCCTCCAAGAACGTCATCAAG CTACAGGAACAGGTGGTGGCGGCCCTCGGT

Table 2.3: Full sequences of short fragments, and first and last 30 bps of longer fragments, used in the conditional rescue constructs.

Protein Toxin Constructs

K. Matzen assembled the original Hid constructs, attB-hand-rH1m2 and attB-hand-rH2m1 (Matzen 2012). Two sets of new shRNAs, targeting Hid versions 1 and 2, were designed based on the miR-1 backbone described in Ni et al. 2011, and targeted the same sites as the original shRNAs (Table 2.4). ShRNA triplets were assembled using the Golden Gate method into OA669, and were flanked by the same 5' and 3' intronic consensus sequences utilized in the original set (5' site cagGTaagt, 3' site tactaattcttctttttccctttttttttAGg, with splice donor and acceptor capitalized). They were then subcloned into the corresponding attB Hid plasmids (set 1 with Hid version 2 and vice versa) 3' of the existing shRNAs using traditional cloning methods (AscI).

In later versions, the MHC16 IVS intron (Pfeiffer et al. 2012) was cloned directly 5' of Hid, and the tubulin 3'UTR was replaced by the p10 3'UTR (Pfeiffer et al. 2012). Both fragments were PCRed from templates provided by O. Akbari, and cloned in via EA. The minimal hand enhancer (handm; Han and Olson 2005), tinman enhancer (tinC; Lo and Frasch 2001), desaturase 1 enhancer (dst1; Billeter et al. 2009), and seven-up enhancer (svp; Gutierrez et al. 2007) were PCRed from genomic DNA and inserted into the Hid constructs instead of the original hand, with a minimal hsp70 promoter downstream of each fragment, using EA methods. The first and last 30 base pairs (bps) of all DNA fragments used in the described constructs are listed in Table 2.5.

Name	Sequence (5' to 3')
Hid 1.1	GCATAGCTCCTCCTCCAACAAC
Hid 1.2	AGGCTGAATCCGGAGCTATTCA
Hid 1.3	CTACAGTAGTTACGGTATTTGT
Hid 2.1	CCACTCCTCGAGTTCGAACAAT
Hid 2.2	AAGCAGAGAGTGGCGCAATACA
Hid 2.3	CGACCGTCGTAACAGTCTTTGT

Table 2.4: Sites in Hid1 and Hid2 targeted by rescue shRNAs.

Name	30 bps of the 5’ and 3’ ends of each fragment, 5’ to 3
Hand full	TGATCGCTTTGAATTTTGGATTTAAGTTTA TCCAATTAATAATGACAATTTATATTTCAG
Hand minimal	CTTCAAAACTTCAATATTTCCAAAAAGGCA CGCCCCAGTGCAGAAAAAGGGGGAGAAAAT
tinC	CATGAACAGCTTTCGATCGGCCACCGATCC CAATGCAACTGTAAAATGTTTCCCTGCCTC
Dst1	CACAGGTTCGTTTTCCACCTGGCCACTTGT TAAGCATGCCACCAATTCTGGCAATTATTT
Svp	AAGCTTGGAAAATGCGTTGGAGTAATAGCC GGAGCACAAACACATGGGCCCTGCCACTTT
Hid	ATGGCCGTGCCCTTTTATTTGCCCGAGGGC ATGGGCTGTGGCTTCTTTGCGGCGCGATGA
MHC16 IVS	AAAAGGTAGGTTCAACCACTGATGCCTAGG GAGGGTACCAACTTAAAAAAAAAAATCAAA
Tubulin 3’UTR	GCGTCACGCCACTTCAACGCTCGATGGGAG AGAAATAAGAAAACCCCACTGTTTTTCTTT

Table 2.5: First and last 30 bps of fragments used in the protein toxin constructs that were not listed in Table 2.3.

Chapter 3
Engineering UD: shRNA toxins

3.1 Introduction

It is important that any underdominant module to be released into the wild is robust on an evolutionary time scale. Potential frailty is a major disadvantage of protein toxin-based underdominance: a single point mutation can compromise toxin function and disable the drive system. Thus, we sought to develop an underdominant scheme that employs a more mutation-resistant toxin. Our progress in implementing such a system is detailed in the work below.

Theoretical Framework

Instead of introducing something toxic and neutralizing it with a two-component rescue, another way of engineering underdominance can involve the removal of something essential as a "toxin" and re-introduction of it from an exogenous source as a "rescue". For example, knocking down an essential gene in a sensitive tissue can produce animal death, and re-supplying that gene in a form that is insensitive to the knockdown can rescue that death. If the rescue can be re-supplied in two pieces, one located on each of the underdominant chromosomes, then this "toxin"/"rescue" scheme can be used to generate an underdominant system. As with the protein-based toxin approach described in Chapter 2, such a method would utilize two underdominant alleles. Each would contain the same tissue-specific promoter, as well as a transcript (containing one or more shRNAs) that would decrease levels of a gene product whose presence in the targeted tissue is required for organismal viability. Each underdominant chromosome would also express, as a part of the same transcript, a fragment of the essential targeted gene, recoded so as to be insensitive to the death-inducing shRNAs. In this system, rescue of essential gene function only occurs when both transgenes are present and expressed, together reconstituting a functional version of the essential transcript.

Small Hairpin RNAs as Toxins

RNA silencing is a powerful method widely used to artificially silence genes in all phyla (McManus et al. 2002; Perrimon et al. 2010). Based on endogenous small RNAs and their processing machinery, RNA silencing techniques introduce artificial small RNAs into an organism that bind to mRNAs in a sequence-specific manner, promoting their degradation and/or inhibition of their translation (Filipowicz et al. 2005). Several methods are commonly used to induce synthetic RNA silencing of specific genes (Perrimon et al. 2010), including long double-stranded RNAs (sdRNAs), synthetic small interfering RNAs (siRNAs), and small hairpin RNAs (shRNAs). In Drosophila, most initial RNA silencing techniques utilized long dsRNAs produced from intron-spliced hairpin RNA molecules (Perrimon et al. 2010). However, these do not effectively work in the female germline (Ni et al. 2011), and can be difficult to clone and work with. To overcome these limitations, several research groups have designed shRNAs based on endogenous micro-RNA (miRNA) backbones (Chen et al. 2007; Ni et al. 2011). These shRNAs, which utilize silencing pathways employed by endogenous miRNAs and siRNAs, are quite straightforward to clone, and have been shown to systematically silence numerous Drosophila genes with high efficiency (Ni et al. 2011). Thus, they are excellent candidates for the role of silencing "toxins" in Drosophila underdominance constructs.

Two-Component Rescue

Designing an antidote that is distributed between two alleles can be challenging, since neither part can provide any rescue function when on its own, yet rescue must be robust when both parts are present in the same animal. There are several distinctly different ways of doing this, but all involve a version of the gene being targeted that is recoded at the DNA level to be impervious to the silencing shRNAs.

One way of achieving a two-component rescue is through the use of a split transcription factor. In this mode of recue, each underdominant allele has half of a transcription factor, with the full-length recoded rescue gene under the control of a transcription factor-activated enhancer. When both alleles are present, the two parts of the transcription factor will associate non-covalently and activate

expression from their target enhancer, which will result in production of the rescue. There are several split transcription factor/promoter systems available, and one (LexA::VP16; Arndt et al., 2001; Luan et al., 2006) had been previously tested in the lab as part of an underdominant system (Matzen 2012). However, such a system has several potential drawbacks, including leakiness (i.e., expression from the promoter in the absence of both transcription factor parts), ineffectual activation of the promoter even when the full transcription factor is present, and toxicity of transcription factor halves (Matzen 2012). Additionally, when previously tested in the lab, this mode of rescue ultimately did not work (Matzen 2012).

Another way to build a two-part rescue involves expressing the target gene itself in two pieces and reconstituting the pieces via intein-based protein splicing. Inteins (short for internal proteins) are enzymes that catalyze their own excision from a protein-intein fusion and splice the host protein back together, generating the free intein and a mature protein (Lockless and Muir 2009; Schwartz et al. 2007; Cheriyan et al. 2013). They work by a number of different mechanisms, and are ubiquitous in single-celled organisms from all three domains of life (Carvajal-Vallejos et al. 2012). Since they can tolerate heterologous sequences at their extein ends, and since split inteins can splice in *trans* (i.e., when fused with different proteins), they have many potential biotechnological applications, and optimized intein versions are becoming increasingly more utilized (Carvajal-Vallejos et al. 2012; Cheriyan et al. 2013). Split inteins have been shown to work efficiently in mammalian cells (Dhar and Mootz 2011), and some data shows they function in Drosophila *in vivo*, as well (Schwartz et al. 2007). Splitting the recoded target gene in two and attaching each part to an efficient split intein piece can thus allow for the distribution of the rescue between two alleles.

Finally, instead of splitting the transcription factors required for rescue expression or splitting the rescue itself, another approach to building a two-part antidote entails expressing the recoded gene from one or both alleles in the essential tissue in such a way that expression from one allele alone is not enough to rescue. For example, one allele can encode a version of the rescue transgene in which a portion of the gene is inverted, thereby generating a transcript that makes a

non-functional protein. The second allele can encode a recombinase that restores the inverted rescue transgene to its wild type configuration. In such a system, in which both chromosomes express gene-silencing shRNAs, the animal can only survive when both chromosomes are present. This kind of irreversible flipping can be carried out using a FLP-catalyzed FLEX switch (Schnutgen and Ghyselinck 2007) that consists of FRT and F3 recombinase sites. When arranged in a particular order around a sequence, these sites plus the FLP recombinase will cause the portion of the recoded target gene between the sites to be permanently re-oriented in the direction that leads to the expression of the full length recoded rescue transcript.

A different way to implement the rescue scheme proposed above is to target a haplolethal gene. Haplolethal genes require two functional copies of a gene for animal survival. If a haplolethal gene is knocked down by shRNAs, one recoded copy expressed under its own promoter will not be sufficient to restore viability. However, two identical copies should restore transcript levels to those needed for animal survival. If each allele contains an shRNA toxin and one copy of the recoded rescue driven by its endogenous promoter, animals heterozygous for the construct should not be viable, but homozygotes will have sufficient amounts of target protein to survive. The appeal of such an approach is its simplicity, since only one underdominant construct will need to be designed. And, as antidote function does not require protein-protein interaction or re-assembly of multiple parts, rescue may be easier to achieve. Indeed, it is with this system in Drosophila that we had the greatest success, engineering several underdominant populations, some with reproductive isolation. Since all animals have haplolethal genes, this system is also potentially adaptable to vector species of interest.

Advantages of shRNA Toxin-Based Underdominance

A major benefit of using shRNAs as toxins is their evolutionary stability: since numerous shRNAs targeting different mRNA sites can be included as part of the toxin, mutations in any one shRNA or its target site should not de-activate the toxin's lethal properties. If deactivating mutations in the rescue gene occur, the animals bearing them will simply die and the creation of broken kill-rescue alleles

will be unlikely. This can be an important point of consideration in designing underdominant modules that will be released into wild populations, as there will be considerable selection against any toxin component of such a system.

Another advantage of this type of underdominance scheme is the range of options available for all of its key components. Suppression of a number of genes can be lethal in sensitive essential tissues, and there are numerous options for providing a split rescue, so there are a lot of variables that can be manipulated to create an underdominant system with the desired properties. This is likely to be true for other species as well as for Drosophila, and makes this approach appealing, since our ultimate goal is to create underdominant systems in various vector organisms.

This chapter describes various shRNA toxin-based approaches to creating a single-locus underdominant system in Drosophila.

3.2 Results and Discussion

3.2.1 Underdominance with CG5266

Kelly Matzen originally selected the genes targeted by the various shRNA-toxin underdominance systems described here. In her work with Hid-based underdominance, she had had some success in driving toxin/antidote expression with the heart-specific hand enhancer, and decided to try using it in an shRNA-based underdominance scheme (Matzen 2012). Research by Neely et al. 2010 showed that depletion of numerous transcripts in the heart, driven by tinman-expressed RNAi, was lethal during development. Kelly chose two of these sensitive transcripts, CG5266 and Rpl35a, as initial targets for shRNA-based approaches. After she graduated, I continued to work on both targets, largely in parallel. Rpl35a-centered approaches are detailed in a later section of this chapter. Here, I describe research using CG5266, a proteasome subunit that has a homologue in Aedes, a vector species of interest.

Intein-Based Rescue

Initially, Kelly tested an shRNA triplet that was based on the endogenous Drosophila miR-6.1 backbone; the three shRNAs targeted the 5'UTR, 3'UTR, and start codon of CG5266, respectively. This shRNA set did not produce any killing in flies. She then designed another triplet, also based on miR-6.1, targeting the coding regions (CDs), and had Geoff Pittman design a different miR-6.1 triplet targeting the CDs based on his ideas what an ideal target site should look like. Kelly never got the chance to test these, and I began working with them after her departure from the lab. The sequences of all these target sites are listed in Appendix B.

Kelly focused on bringing about rescue using a split transcription factor, but was unsuccessful in getting significant levels of regulated expression. I decided to try a protein slicing-based rescue using the Nostoc DnaE split intein (Zettler et al. 2009), which has been shown to work efficiently in mammalian cells (Dhar and Mootz 2011). I split the recoded CG5266 gene, with its introns, at a point that I surmised would tolerate exposure until the mature protein was spliced together, and attached each piece to the appropriate half of the intein (i.e., 5' protein end to the N intein portion, 3' protein end to the C intein portion). In this way, when both halves were present, the two pieces should be post-transcriptionally spliced together to form a mature rescue protein (Figure 3.1A).

Cell Culture Tests

Before any fly constructs were built, I tested the various components of this system in Drosophila S2 tissue culture. To determine if shRNAs targeting CG5266 suppressed transcript levels I used a proxy for proteasome function, two short-lived GFP proteins: Ub-R-GFP (Dantuma et al. 2000), an N-end rule substrate, and pd2EGFP (Clontech, Mountain View, CA), which has a PEST sequence on the C terminus of the protein. Each of these degrons results in the encoded protein having a very short half-life in otherwise wild type cells, and therefore cells expressing them have low levels of fluorescence. I hypothesized that fluorescence intensity would increase if proteasome function was disrupted.

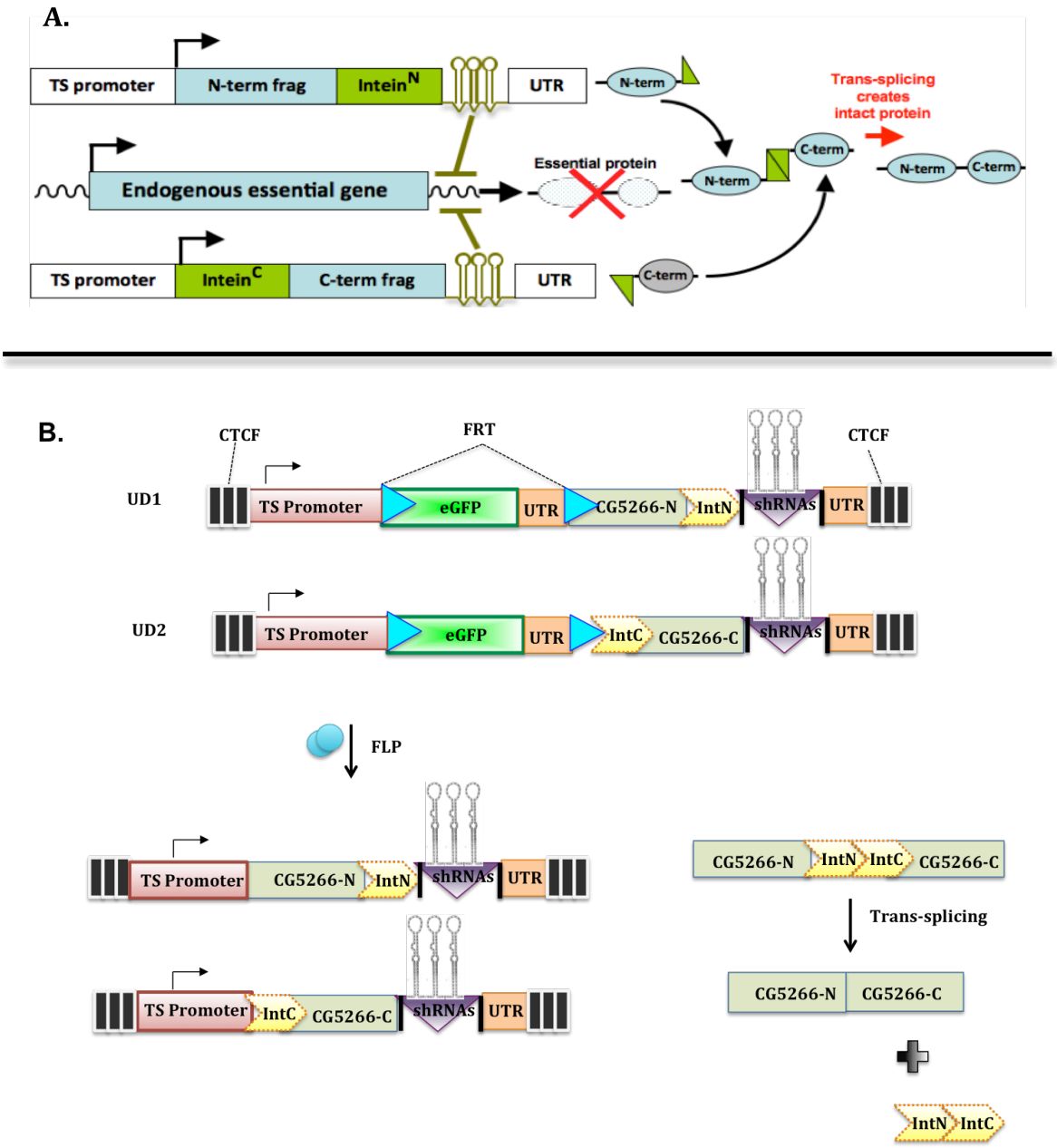

Figure 3.1: A. Schematic of underdominant constructs with shRNA toxins and a self-splicing protein rescue. Each underdominant allele contains a tissue specific promoter driving the expression of shRNAs that silence some essential gene in the 3'UTR of a recoded gene-intein fragment. When the two inteins are present in the same individual, they catalyze the splicing of their extein protein fragments, creating a mature recoded rescue protein. B. Architecture of the actual CG5266-targeting intein constructs tested in various tissues in Drosophila. In addition to parts described above, each construct has a blocking sequence that is excised upon the addition of FLP, and is flanked by CTCFs.

The two GFP-degron coding regions, as well as all three shRNA sets targeting CG5266 built by Kelly (one targeting the UTRs and two targeting the CDs, with each set in an intron), were cloned separately into the pAc5.1-HisB vector downstream of the Actin5 promoter. The entire CG5266 recoded rescue protein and the N- and C-portion protein-intein pieces were also individually cloned into pAc5.1-HisB. In this way, I could test whether the shRNAs suppressed CG5266 levels, whether providing a full-length recoded rescue could reverse this suppression, and whether intein splicing of the split CG5266 protein parts would produce a comparable rescue effect.

As predicted, expression of either destabilized GFP protein (in otherwise wild type cells) resulted in very faint, short-lived GFP signal in S2 culture. The addition of shRNAs targeting CG5266 resulted in an increase in GFP levels, with the degree of stabilization determined by the specific shRNA set used. The set targeting the CG5266 UTRs resulted in a very weak effect (as was expected, since Kelly saw no phenotype when she expressed this set in the fly heart), while the sets targeting the CDs produced a very significant increase in observable GFP levels, with the triplet designed by Geoff Pittman (which clustered shRNA target sites in one portion of CG5266) generating a slightly stronger effect. Addition of the full-length recoded CG5266 gene to cells expressing the above shRNAs resulted in a drop in GFP expression levels, suggesting rescue of the shRNA-induced phenotype. While neither protein-intein piece on its own brought about rescue, addition of both fragments at the same time produced a rescue effect comparable to, though slightly less than, that achieved by expression of the full-length protein. Given the above data, I concluded that, at least in tissue culture, the components of this underdominance scheme functioned as designed.

GMR Constructs and Results

I initially tested the above system in the fly eye, using the eye-specific GMR promoter, as proof of principle for later work in essential tissues. The general architecture of constructs generated for this experiment, and of all essential tissue constructs deriving from this work, are shown in Figure 3.1B. Each underdominant allele contained an identical set of intronic miR-6.1-based shRNAs in the 3'UTR of

one of the two split rescue/intein pieces, expression of which was driven by the tissue-specific promoter (the first round of GMR-driven constructs contained either one or the other set of CD-targeting shRNAs, not both). The promoter was separated from the intein-protein open reading frame (ORF) by either the Tet cassette or the minimal GFP blocker described in Chapter 2. (Which blocker was used depended on what phase of testing we were in, and changed from construct to construct; however, this did not in any way affect the actual underdominance-related results.) Additionally, the entire allele was flanked by artificial CTCF insulators (Kyrchanova et al. 2008) to prevent any possible genomic insertion effects from influencing expression.

The assembled GMR constructs were inserted into two attP insertion sites, and the transgenic flies generated were then crossed to germline FLP lines to excise the blocking cassette (detailed in Appendix C). When expressed in the fly eye under the GMR promoter, both shRNA toxins resulted in a glassy, decaying eye phenotype; additionally, there was some observed pupal lethality, likely due to leaky expression of GMR in non-target tissues. Transheterozygous flies were obtained and appeared to regain the wild type eye phenotype, suggesting that the rescue was functional (Figure 3.2). It was evident, however, that rescue may not have been complete, as some flies began to show an intermediate eye phenotype (i.e., not as healthy as wild type, not as defective as kill-only flies) with age. The same kill and rescue phenotypes were observed at both of the attP insertion sites tested (Table 3.1).

I also tried using both sets of CD-targeting shRNAs in one construct in order to achieve the small-eye phenotype observed with Hid-based underdominance alleles driven by GMR (Matzen 2012). This did not result in small eyes, but did produce a more pronounced glassy and decaying eye appearance. Transheterozygous flies with the same characteristics as the ones above were obtained for these double-kill alleles as well, and were used to set up underdominant fly stocks.

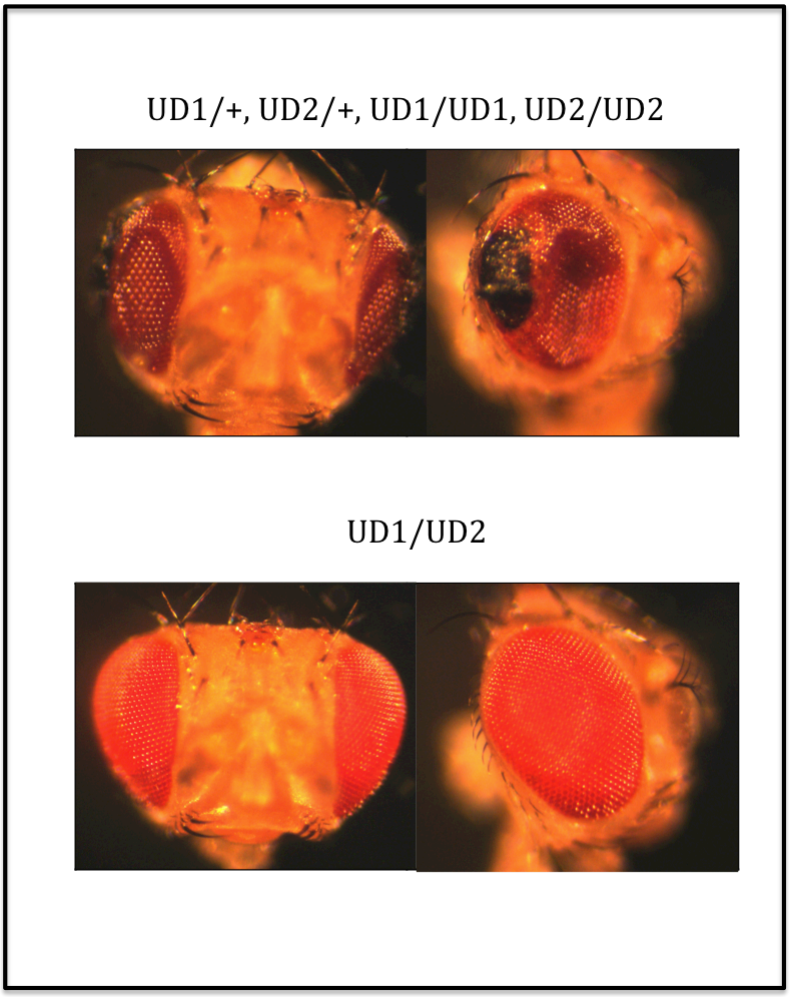

Figure 3.2: Eye phenotype of flies bearing only one type of GMR-driven CG5266 UD allele (above) versus both types of alleles (below). The heterozygous individuals have abnormal colored, deformed eyes while the transheterozygotes eyes are rescued back to a wild type phenotype.

Essential Tissue Constructs and Results

Since the CG5266 system appeared to function in tissue culture and in *vivo* in the fly eye, it was moved to essential tissues. As Hid-based underdominance had shown some success with the heart-specific hand enhancer, I first decided to test this system with the minimal hand enhancer (handm; Han and Olson 2005), since it was shorter (~ 500 bps long) and seemed to work at least as well as the full-length version (~ 1500 bps long). Constructs identical to those diagrammed in Figure 3.1B, with the handm enhancer followed by a basal hsp70 promoter, were injected into two attP sites. Following FLP-mediated blocker removal, the shRNA toxin resulted

Promoter	Insertion Site	Killing	Rescue
GMR	86Fa 9724	Strong[1] Strong[1]	Yes[2] Yes[2]
handm	86FA 51D	Strong Strong	No Yes[3]
tinC	86FA 51D	Very weak Very weak	— —
dst1	86FA 51D	100% 100%	No No
svp	86FA 51D	100% 100%	No No

Table 3.1: Fly results with CG5266 shRNAs and intein-based rescues.
[1.] Killing in the eye refers to a severe eye phenotype.
[2.] Rescue refers to restoration of wild type eyes.
[3.] Although transheterozygous rescue individuals were obtained, there were few of them and they were sickly.

in significant (~ 85-90%) killing. The flies that did survive were generally weak and short-lived, with a delicate, fluid-filled wing phenotype; however, they were fertile, and thus reproductive isolation was not achieved with this system. Transheterozygous animals were recovered at one attP site (51D), but the majority of them were not 100% healthy, indicating that the rescue was only partially functional.

As neither killing nor rescue was optimal, I attempted to improve both. Per Pfeiffer et al. 2012, addition of certain translational enhancer (TE) sequences can significantly boost expression levels. Thus, to enhance rescue expression, I added a 5' MHC IVS16 intron and the p10 3'UTR (from Pfeiffer et al. 2012) to the CG5266-intein fragment. This, unfortunately, did not produce any improvement in rescue.

To increase killing, I tried using both sets of CD-targeting shRNAs instead of just one, but this gave only a very slight increase in killing. Additionally, I built up six new shRNAs that targeted the same sites as the existing shRNAs, but were based on the miR-1 backbone (Ni et al. 2011), since work on Hid underdominance suggested they may work better in the heart (see Chapter 2). These were placed in an intron in

the 3'UTR of the CG5266-intein ORF, both separately and behind the existing shRNAs. By itself, this set produced killing that was no stronger (and potentially slightly weaker) than the killing obtained with the two original miR-6.1 triplets. When combined with the existing shRNA sets, killing improved only marginally.

I also tested the original constructs (with only the stronger set of original miR-6.1 shRNAs) with different enhancer fragments to see whether stronger toxin expression could be obtained. I utilized the enhancers previously tested with Hid-based underdominance (Chapter 2), all with a basal hsp70 promoter: heart-specific tinman (tinC; Lo and Frasch 2001) and oenocyte-specific desaturase 1 (dst1; Billeter et al. 2009) and a seven-up fragment (svp; Gutierrez et al. 2007). TinC-driven killing was very weak at both insertion sites I tried. Both dst1 and svp gave 100% killing at the two sites; however, I was not able to generate transheterozygous flies with either construct set, suggesting that the recoded splicing protein was not sufficient to rescue with these promoters. The results obtained for all essential tissue work described in this section are listed in Table 3.1.

CG5266 Underdominance with a Flipping Rescue

As discussed in the introduction, another way to build a two-component expressed transcript rescue is to invert a portion of the recoded target gene in such a way that, as is, the expressed transcript is nonsense. If this nonsense transcript is provided on one underdominant allele while the other allele supplies some means of permanently flipping around the inverted portion back to its correct orientation, a functional rescue will only be expressed if both alleles are present.

Typically, FLP recombinase catalyzes the excision of fragments located between two FRT sites in the same orientation, and the reversible inversion of fragments located between two oppositely oriented FRT sites (Zhu and Sadowski 1995). However, irreversible flipping can be brought about using a FLP-catalyzed FLEX switch (Schnutgen and Ghyselinck 2007). This switch consists of classic FRT sites and the F3 variant of FRT sites (each site type is unable to react with heterospecific sites). When arranged in a particular order around a sequence, these sites plus the

FLP recombinase will cause permanent inversion of the fragment located between them.

We decided to integrate this switch into a different rescue modality for the CG5266-based underdominant system. In this scheme (Figure 3.3), each construct has the same promoter placement, intronic shRNA toxin, and flanking CTCFs as the original intein alleles. However, in place of the CG5266-intein fragments, one construct has FLP, and the other has the full-length recoded CG5266 gene with the first exon inverted and flanked by the FLEX switch-dictated arrangement of FRT and F3 sites. On its own, this inverted version should not produce an operative protein; however, in the presence of FLP, the first exon should be inverted back to its correct orientation, and functional rescue transcript should be made. Additionally, B3 recombinase sites instead of FRT sites flank the removable blocker sequence.

This system was first tested in the heart with the handm enhancer. After blocker removal via crosses to germline B3 recombinase lines, the handm-FLP construct gave killing identical to the handm-intein alleles (~ 85-90%), as would be expected since the shRNA toxin remained constant. The construct with the partially inverted CG5266 gave a similar percentage of killing, but the surviving heterozygotes, in addition to being sickly and having fluid-filled wings, were also smaller than wild type flies. This suggested that expression of the nonsense version of CG5266 had some effect independent of the toxin, but we never determined why exactly that effect arose. I was not able to obtain transheterozygotes for this set of constructs.

The flipping rescue scheme was also tested with the dst1 and svp oenocyte enhancers. For both enhancers, I observed the same complete killing as with the oenocyte enhancer-intein constructs, but no rescue was brought about. To see whether the FLEX switch was re-orienting the CG5266 exon as designed, I crossed flies bearing the partially inverted CG5266 rescue allele, with the blocking sequence still present, to the BicC FLP line, and then PCR-amplified the fragment containing the inverted region from progeny genomic DNA. Sequencing of this region showed that the FLEX switch was working as expected, and that the first exon of CG5266 was now in the correct orientation. I also further out crossed these flipped flies to

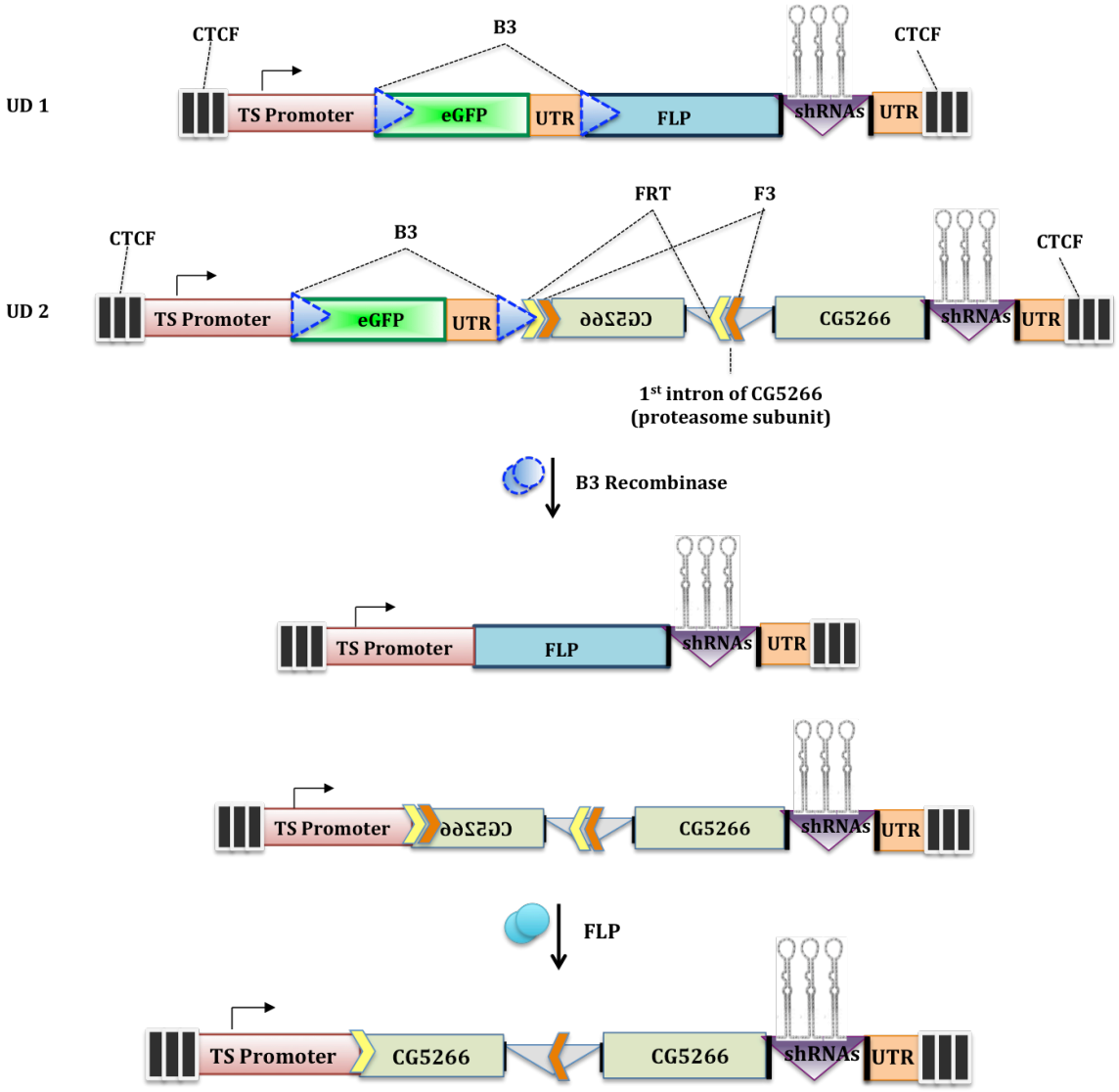

Figure 3.3: Alleles constructed for the CG5266 flipping rescue underdominance system. UD1 has a tissue-specific promoter driving FLP with CG5266-targeting shRNAs in its 3'UTR; the promoter is separated from the FLP ORF by a blocker sequence (flanked by B3 recombinase sites), and is surrounded by CTCFs. UD2 is identical, except that FLP is replaced by full-length recoded CG5266, with its first exon inverted and surrounded by the FLEX switch arrangement of FRT and F3 sites. Upon addition of B3 recombinase, the blocker sequence will be excised from both constructs. FLP will then cause re-orientation of the inverted CG5266 exon to its proper direction, and the full-length recoded rescue protein will be expressed.

the germline B3 recombinase lines to remove the blocker, and saw that, for the dst1 promoter line, some unblocked flies survived. This suggests that, in the proper orientation, the recoded CG5266 gene produces enough transcript to rescue when one set of shRNAs is present. However, whether the overall flipping rescue mode did not work because FLP activity in the oenocytes was not adequate to catalyze the FLEX switch, or because the reoriented CG5266 did not restore sufficient transcript levels in the presence of two shRNA sets, could not be determined.

Killer-Rescue with CG5266

At this point, with 100% killing using the oenocyte enhancers but no rescue, we decided to take a more empirical approach to troubleshooting the antidote. Before we could really test whether the intein splicing or FLEX switch rescue modalities could work, we needed to know whether the recoded CG5266 gene in one piece was sufficient to rescue.

To this end, I built separate kill and rescue constructs (Figure 3.4). Not only would this clarify whether enhancer-driven CG5266 expression was sufficient for rescue, but it would also potentially create a novel drive mechanism, as killer-rescue has been theoretically shown to have self-limiting gene drive capability (Gould et al. 2008). For the kill constructs, a single intronic shRNA set was placed into the 3'UTR of dsRed, and the GFP blocker separated the oenocyte enhancer from the dsRed ORF. For the rescue constructs, the full-length recoded CG5266 gene was directly driven by an oenocyte enhancer (no blocker was necessary, since no toxin was present). Both alleles were flanked by CTCFs.

These constructs were tested at three different attP insertion sites (86Fa, 51D, and 51C). At all three sites, killing after blocker removal was complete with both dst1 and svp. Rescue alleles inserted at the same sites as the kills gave robust rescue, with killer/rescue individuals appearing healthy. Rescue could also be achieved with both promoters when the rescue constructs were inserted at a different genomic site than the kill. However, very few double homozygotes (K/K; R/R) could be generated, and no two kill/one rescue flies (K/K; R/+) were found. This was problematic for two reasons.

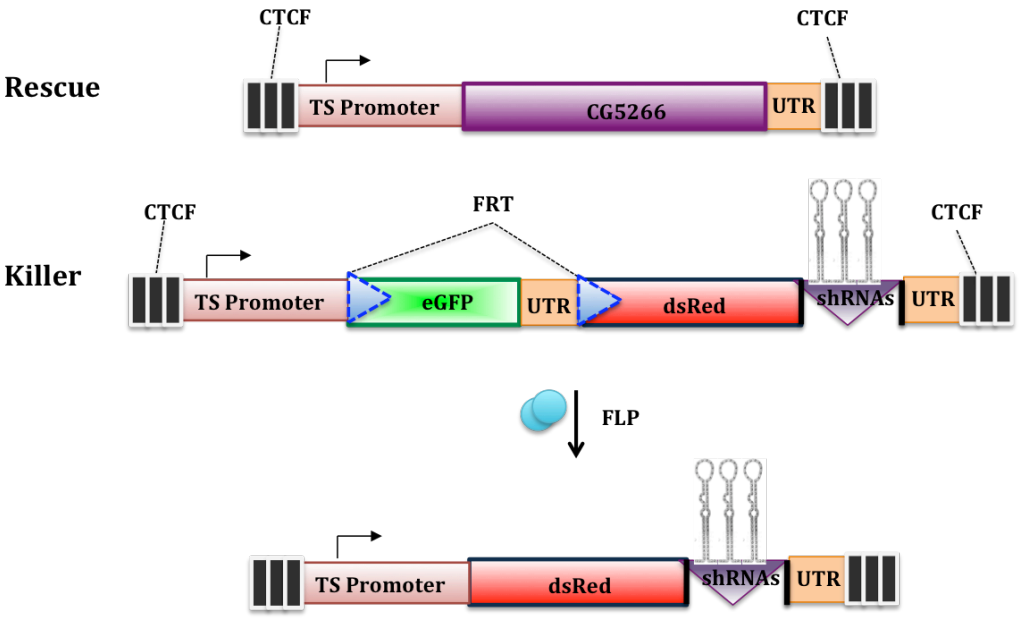

Figure 3.4: Architecture of CG5266 killer and rescue constructs. The rescue construct consists of a tissue-specific promoter driving the full-length recoded CG5266 gene. The killer has the same promoter, which separated from dsRed and CG5266-targeting shRNAs in its UTR by a GFP blocking sequence. FLP catalyzes the excision of the blocker, leading to expression of dsRed and the shRNA toxin. Both killer and rescue cassettes are flanked by CTCFs.

Firstly, healthy double homozygotes would be required for the killer-rescue system to work as a drive mechanism. Additionally, as designed, both split-rescue underdominance modalities with CG5266 only have one copy of the recoded antidote gene for two sets of shRNAs, and therefore cannot be expected to give rise to healthy flies unless that single copy of the antidote transgene is sufficient to rescue the lethality induced when both shRNA sets are expressed. In short, the above results suggest that one copy of the rescue transgene, expressed under the control of the oenocyte promoters, was insufficient to provide rescue, calling into question our ability to engineer underdominance using the above approaches.

There were several potential reasons that could explain poor rescue by one copy of CG5266. Position effects could be responsible for differential expression of the toxin and antidote, tipping the scale against robust rescue. Expression from oenocyte enhancers may not have restored CG5266 transcript levels to those naturally found in this tissue. Also, recoding of CG5266 may not have been sufficient

to prevent some shRNA targeting of the rescue allele, thus reducing the amount of transcript available. The last hypothesis was considered a likely source of problems. This is because the shRNAs targeted the CD portion of CG5266, limiting my options for recoding while still preserving the identical protein coding sequence (CG5266 is a highly conserved protein).

To address this possibility, I built a final set of killer rescue constructs. I assembled two new sets of shRNAs, each consisting of four stem-loops targeting either the 5'UTR (in one set) or the 3'UTR (in the second set) of CG5266, and cloned them into the kill backbones in place of the original toxins. Since the rescue does not utilize the endogenous UTRs of CG5266, insufficient recoding could not be an issue. I also put the previously described translational enhancers (Pfeiffer et al. 2012) around the full-length recoded CG5266 gene, to hopefully enhance rescue expression.

The above constructs were injected into two attP lines (51D and 86Fa), and experiments with them are currently in progress. Initial results show that all but one construct at one site (svp-driven 5'UTR shRNAs at 86Fa) produce 100% killing. Future work will determine whether rescue can be attained, and whether robust double homozygotes and two kill/one rescue flies can be generated.

3.2.2 Underdominance with Rpl35a

Although it was not originally chosen as a target for this reason, Rpl35a is a haplolethal gene (Cook et al. 2012; Marygold et al. 2007). This is advantageous for two reasons: since levels of this transcript above 50% are required for proper function, it was likely that sufficient suppression to bring about lethality could be achieved using artificial shRNAs; this also presented an opportunity to try a different approach to two-component rescue. Additionally, the mode of rescue is such that only one construct needs to be built, simplifying cloning. And since haplolethal genes are found in all organisms, it is likely that, if this approach to building underdominance is successful in Drosophila, it can be transferred to vector species of interest. Below, I describe successful construction of an Rpl35a-based underdominant system in Drosophila.

Construct Architecture

The original shRNA set utilized in this work, cloned by Kelly Matzen, had three shRNAs – one targeting the 5'UTR, one targeting the 3'UTR, and the last one targeting the start codon of Rpl35a. Kelly had tested this set as a part of a split transcription factor-based underdominance approach, and showed that it produced 100% lethality (Matzen 2012), but no rescue, when driven by the hand enhancer. Since this shRNA set seemed to suppress Rpl35a well in the heart, I decided to use it as a part of a haplolethal rescue scheme.

The basic structural design of all Rpl35a constructs described in this work is shown in Figure 3.5A. The toxin portion is comprised of a tissue-specific enhancer (handm plus hsp70 basal promoter for the initial construct), basic GFP blocker, and intronic set of shRNAs in the 3'UTR of VP16, all flanked by CTCFs to prevent inappropriate rescue activation by the enhancer. (VP16 has no actual function in the construct, and was transferred to my system from Kelly's arrangement to simplify cloning and provide a fragment to place in front of my shRNAs.) The rescue portion consists of full-length Rpl35a (with its introns) recoded to be insensitive to the shRNAs, along with its 5' and 3'UTRs and enough surrounding genomic sequence to (hopefully) capture elements needed for its apposite transcription. If shRNA knockdown of endogenous Rpl35a is adequate, and if expression of the recoded gene is temporally and quantitatively sufficient, then individuals with one copy of the Rpl35a underdominant allele should not have enough transcript to survive (since all of their endogenous Rpl35a is suppressed and they only have one rescue copy), but those with two copies should be viable. Outcrosses of such underdominant homozygotes to wild type should produce 100% offspring lethality (Figure 3.5B).

Initial Fly Results

Transformants for this initial construct were generated at two attP sites (51D and 86Fa) and crossed to germline FLP lines to remove the blocker sequence. After blocker removal, most heterozygotes survived, indicating that killing was very weak. Since we knew that this particular set of shRNAs expressed under handm previously

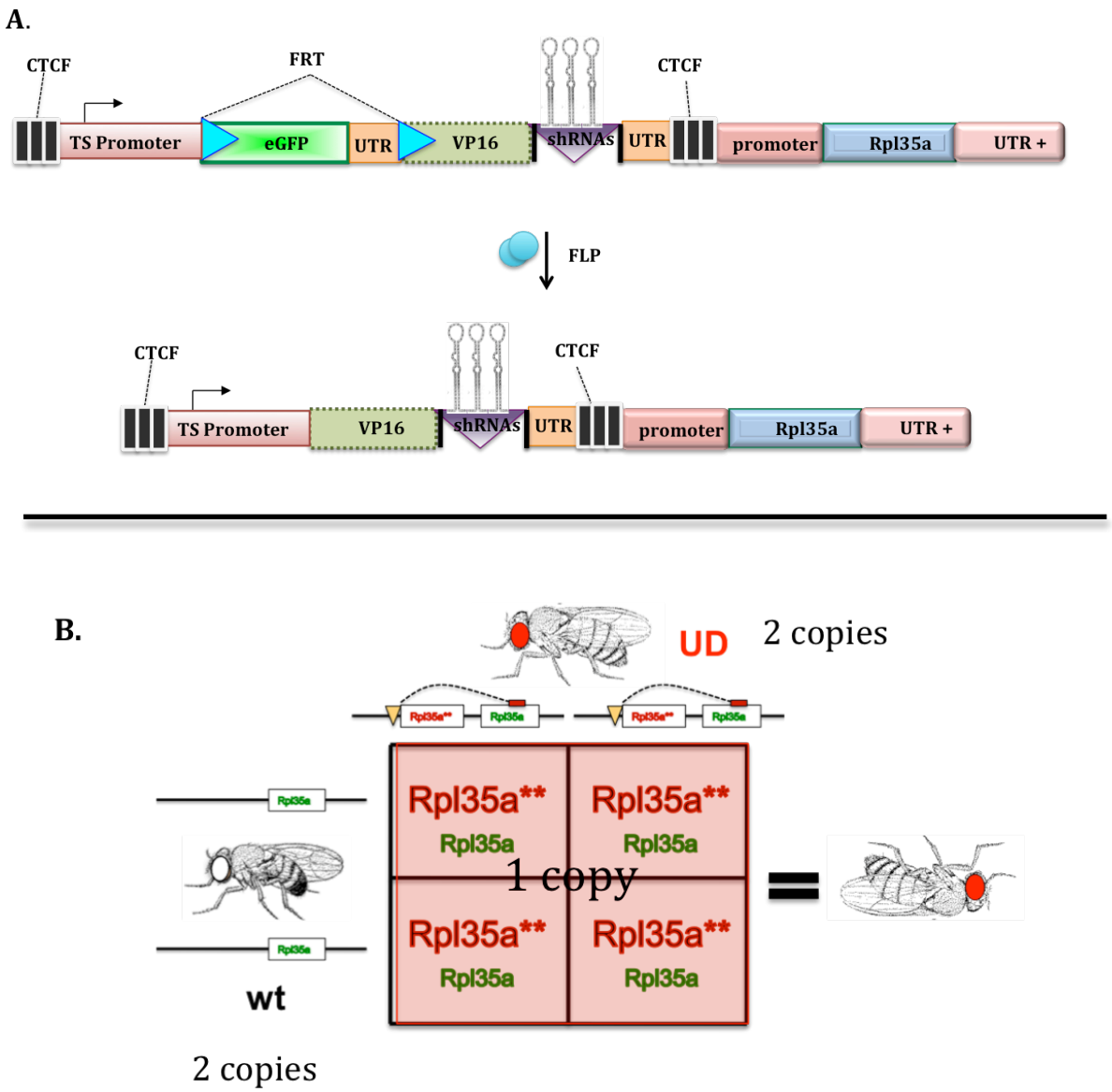

Figure 3.5: A. Schematic of the Rpl35a underdominant construct. The killer portion of the allele is on the 5' end, and consists of (5' to 3') a tissue-specific enhancer, a GFP blocker sequence, VP16, and shRNAs targeting Rpl35a in the 3'UTR of VP16. This entire cassette is flanked by CTCFs. The rescue portion of the construct contains the recoded Rpl35a gene with its endogenous 5' promoter and 3'UTR. After FLP catalyzes removal of the blocker cassette, the toxin is expressed. B. A Punnet square demonstrating the result of outcrossing Rpl35a homozygotes to wild type. Wild type flies have two endogenous copies of Rpl35a and are viable. Homozygotes for the Rpl35 underdominant allele have all of their endogenous Rpl35a silenced by shRNAs, but have two copies of the rescue allele and are also viable. Their progeny have all of their endogenous Rpl silenced by the shRNAs, but only have one copy of the rescue, and are thus all inviable.

gave 100% killing at 86Fa, we hypothesized that Rpl35a suppression was not particularly efficient, leaving enough endogenous transcript to effect rescue when combined with a single copy of the recoded rescue. A stronger kill was needed for the system to function as designed.

Experimentation with shRNA design in the lab had suggested that targeting the CDs rather than the UTRs of a gene resulted in more potent suppression, and work on Hid underdominance indicated that miR-1-based shRNAs may function better in the heart (Chapter 2). Based on these suppositions, I designed a new intronic set of six shRNAs that were based on the miR-1 backbone and targeted the CDs of Rpl35a, and placed them downstream of the existing set. However, even with these additional shRNAs, killing with handm remained very weak.

Oenocyte Promoter Work

At this point, we decided to move the Rpl35a system to a new (and hopefully more sensitive) tissue. Since oenocyte enhancers gave robust killing in the CG5266 experiments, I replaced handm in constructs that had the original and new miR-1-based shRNAs with either the dst1 or svp enhancer, and tested these two constructs at 86Fa. Dst1 produced 100% killing, but did not give any rescue. Svp gave strong but imperfect killing—less than 5% of heterozygotes survived, and those that did were short-lived, developmentally delayed (emerging days after wild type flies), and had fragile, buckled wings (they were, unfortunately, fertile). The svp construct also produced rescue, and homozygous individuals were generated. They had a slight wing phenotype, and did not appear 100% normal, suggesting that rescue was not complete (all further references to incomplete rescue imply the same phenotype). However, I was able to establish and maintain an underdominant fly stock for use in future drive experiments.

This was a good start. Ultimately, though, we wanted to engineer an extreme underdominant system with high homozygote fitness (i.e., 100% kill and robust rescue). To accomplish this, both kill and rescue functions would have to be improved. I first decided to experiment with existing shRNA sets. One hypothesis for why the rescue did not function at all in dst1 lines, and was not optimal in svp lines,

was insufficient recoding. As with CG5266, anti-Rpl35a shRNAs targeted the coding region of the gene, which is very highly conserved, and the amount of recoding I could do while preserving amino acid integrity was limited. Thus, it was possible that the shRNAs could suppress the recoded rescue allele to some extent.

To remedy this, I did a modest amount of further recoding of the rescue, focusing on the few sites that weren't previously recoded as thoroughly as possible, and in some instances forgoing Drosophila codon bias. I also cloned constructs that contained only the original shRNA set built by Kelly, since these targeted mostly UTR sites that were slightly more amenable to recoding. This effort gave me four new constructs—two with additional rescue recoding and two with one original shRNA set (one for each oenocyte enhancer) —all of which were injected into attP sites 86Fa and 51D.

The above set of experiments yielded some successful results. Increased recoding did not make a significant difference in antidote function: the dst1 construct still produced 100% killing and no rescue at both sites, and the svp construct at 86Fa gave the same strong killing and imperfect reversal of suppression. However, at 51D (which was not previously tested with oenocyte enhancer-Rpl35a constructs), the svp allele produced 100% killing and imperfect rescue, and an underdominant fly line that was reproductively isolated from conspecifics was established. Constructs with the original shRNA triplet at 51D effected complete killing and no rescue; at 86Fa, the dst1 version produced complete killing and imperfect rescue, and the svp version gave strong killing and incomplete rescue. From these, two more underdominant populations (one displaying extreme underdominant behavior) were established.

Design of New shRNAs

Our results with Rpl35a-based underdominance thus far were quite promising, so we continued optimizing this approach in hopes of generating an extreme underdominant system with high homozygote fitness. At this point, we decided to experiment with new shRNA sets that would silence potently yet be unable to target the recoded rescue.

To this end, I chose six new target sites in the CDs of Rpl35a that permitted more thorough recoding (i.e., mainly that contained codons in which more than just the third base pair could be changed without changing the corresponding amino acid). ShRNAs (based on the miR-1 backbone) targeting three of these sites at a time, plus a fourth 3'UTR-targeting shRNA from the original set (which targeted a non-conserved site that could be fully recoded), were cloned as intronic four-mers and assembled together with newly recoded rescues and either dst1 and svp. Additionally, I built a construct that had both of the above four-mers and a rescue modified to be resistant to all of the shRNAs. These six constructs (one with shRNAs targeting new sites one through three plus the UTR site, one with shRNAs targeting new sites four through six plus the UTR site, one with all eight shRNAs, for both dst1 and svp) were tested at 86Fa and 51D. All gave complete killing and no rescue.

I also built a construct with just the recoded rescue driven by its endogenous promoter region (exactly as in our underdominant alleles) to test whether expressing twice the normal amount of Rpl35a in tissues where there was no shRNA knockdown imposed some kind of fitness cost. The flies with just the rescue construct appeared completely normal, however, and had the same reproductive fitness as the empty attP lines used to generate them.

We suspected that, even with the above new constructs, insufficient recoding was the reason that rescue was not effected (even though the new set of target sites allowed for greater rescue manipulation, recoding was still far from complete). In order to circumvent this issue, we would have to select sites that could be changed significantly without disrupting proper protein expression, and this was very limiting, as the CDs and 5'UTR of Rpl35a are extremely highly conserved across all Drosophilids. However, I did find limited regions of the 3'UTR that could presumably tolerate significant change, as they were not conserved and didn't appear to have any significant regulatory sequences. Thus, I selected four such sites to target with yet another new set of miR-1-based shRNAs.

The latent concern with such a strategy is insufficient silencing, as targeting of UTRs does not seem to be as effective as targeting of CDs (Hay lab, unpublished data). As we brainstormed ways to avoid this, we found that, according to research

by De et al. 2013, mismatching the 3' end of a guide shRNA to its mRNA target significantly improved silencing of abundant genes, at least in mammalian cells. We had not previously tried this in the lab, and as a rule always made out guide strands 100% complementary to their mRNA target sites; however, since Rpl35a is very highly expressed, we decided to try this approach. Thus, in addition to cloning an intronic shRNA four-mer using our typical approach (with shRNA guide strands that are completely complementary to the mRNA sites they target), I cloned another four-mer set where the last four base pairs of each guide strand were the same as, instead of complementary to, the target mRNAs (since De et al. 2013 suggested that a mismatch of four bases was optimal for silencing). These two new shRNA sets were assembled together with appropriately recoded rescues and either dst1 or svp, and injected into 86Fa and 51D.

Constructs with completely complementary shRNAs produced very weak killing with both enhancers at both sites. However, the mismatching strategy worked: mismatched shRNA-containing lines with dst1 showed complete killing and no rescue at both sites, while the mismatched shRNA-containing svp construct gave strong but incomplete killing and robust rescue at 51D (at 86Fa, the killing was appreciably weaker). I established an underdominant fly stock—our first with a robust rescue—using the 51D line for use in future drive experiments.

Final Set of Constructs

Although, at this point, we had a number of underdominant populations with which to set up gene drive experiments, we wanted to undertake one final effort to build a system with complete kill and complete rescue. One confounding result from the previous set of experiments was that the dst1 allele with 3'UTR-targeting mismatched shRNAs failed to rescue, even though inadequate recoding was not a possibility. We went back and scrutinized our construct designs, and realized that the γ-tubulin 3'UTR following our shRNAs was minimal, and may not have contained all of the features necessary for proper transcription termination. This may have led to read-through into the rescue, compromising its expression. (Even though a CTCF

insulator separates the toxin and rescue, it is not clear that this insulator is effective at preventing read-through.)

To remedy this, I replaced the short t γ-ubulin UTR with a much longer version, one that should contain all of the necessary termination signals. I also replaced the VP16 fragment 5' of the shRNAs with dsRed, since VP16 can be slightly toxic on its own (Lin et al. 2006; Fu et al. 2010), even though this was never apparent in earlier work (e.g., Matzen 2012). Just in case read-through was still a possibility, I also assembled a version of the above construct that has the rescue segment inverted onto the bottom strand, with its 3'UTR abutting the γ-tubulin 3'UTR. These constructs have both been injected into 86Fa and 51D, and further analysis will show whether they can bring about 100% killing and rescue.

Drive Experiments

As mentioned in the introduction, many of our regularly used attP lines, including the ones used for most of the Rpl35a work, have intrinsic fitness costs that disadvantage them in competition with wild type. Consequently, all drive experiments involving transgenics created at these attP lines have to be done against the empty attP lines themselves, as these transgenics may be unable to compete with wild type for reasons unrelated to their underdominant alleles.

With this in mind, I set up initial drive experiments for all five underdominant populations created during the course of the above experiments against the empty attP sites used to generate them. For a single locus, single allele system with a low fitness cost, drive could theoretically be expected if the frequency of initial release was greater than 50% of the population (Davis et al. 2001; Sinkins and Gould 2006). To mitigate any potential fitness costs in my systems, I utilized significantly higher release frequencies (75% or 80%). Under these conditions, all of the underdominant lines produced some gene drive. Most quickly spread to fixation; however, one line did show both drive in and drive out (depending on the biological replicate), suggesting that its initial release threshold was right around the 80% frequency being used.

I then chose two of the underdominant populations—one of the reproductively isolated ones, as well as the one with the robust rescue—and set up more extensive drive in and drive out experiments at various thresholds. These experiments are still ongoing, but preliminary results of the drive trials for the population with the robust rescue are shown in Figure 3.6. As with the Hid system, this population is characterized by rapid spread of the underdominant alleles even when released below the hypothesized frequency of 50%. Instead, the threshold here seems closer to 40%, as at that release ratio some replicates drive in and some replicates drive out, while all releases below that frequency show decrease of underdominant alleles. The preliminary numbers for the reproductively isolated underdominant Rpl

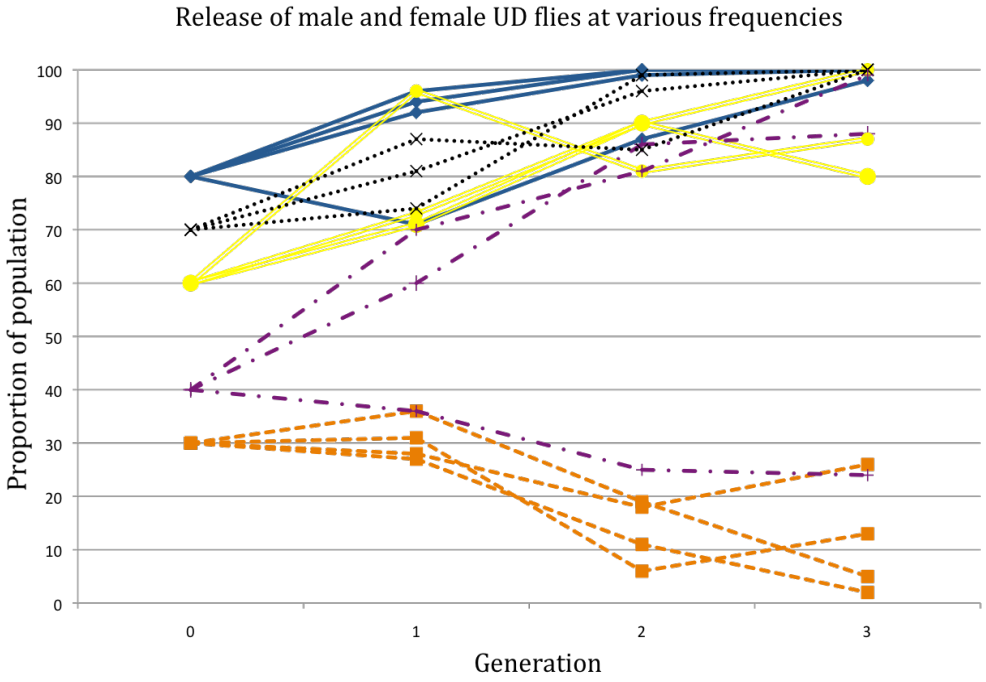

Figure 3.6: Drive experiments for the Rpl35a UD population with the robust rescue were set against the empty attP insertion line used to generate underdominant transgenics by releasing transheterozygous male and female individuals at various frequencies (from 30%-80%). Each line represents a biological replicate. Most replicates show spread of underdominant alleles in the population, even when released at frequencies below the predicted drive threshold of 50%. The threshold here seems closer to 40%, as at that release ratio some replicates drive in and some drive out, and releases below that frequency show decrease of underdominant alleles.

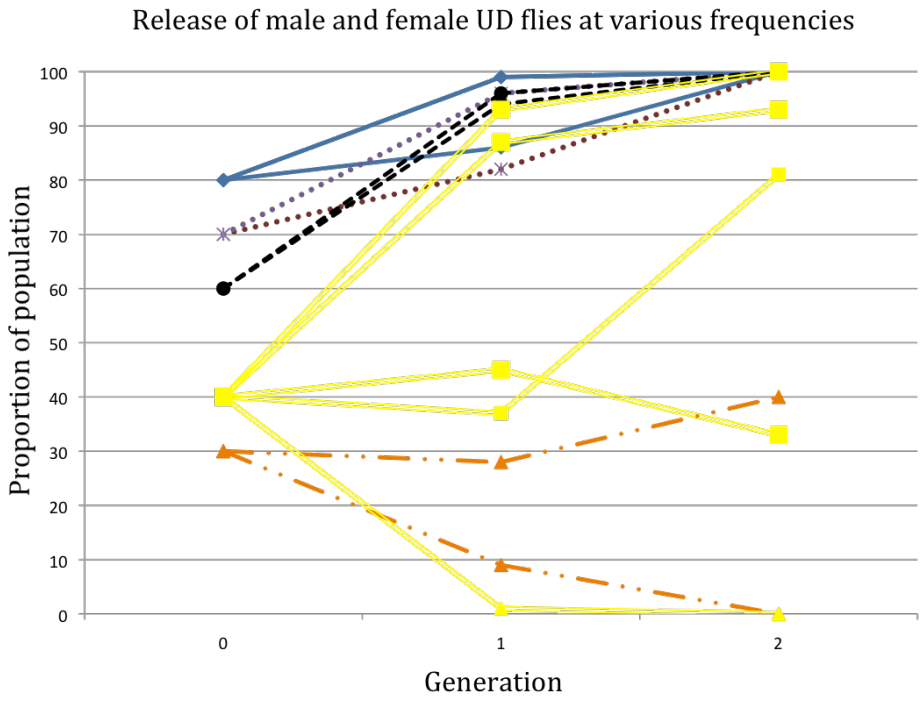

Figure 3.7: Drive experiments for the reproductively isolated Rpl35a UD population with imperfect rescue were set in the same manner as those for the Rpl35a UD population with robust rescue. Each line represents a biological replicate. As with the earlier Rpl example, most replicates are so far showing spread of underdominant alleles in the population, even when released at frequencies below the predicted drive threshold of 50%. The threshold here seems to be between 30% and 40%, but further data will be needed to ascertain this.

population appear similar (Figure 3.7), but the data set for those experiments is less complete and is still being collected.

As discussed in Chapter 2, these unexpected drive dynamics could be due to eye color differences between the underdominant flies and the empty attP insertion site individuals (the former have red eyes, which makes their vision superior to that of the latter, who have white eyes), and drive experiments in the dark will be set up to address this. It could also result from presence of recessive deleterious alleles in the attP insertion lines. The underdominant transgenic lines have been introgressed with white minus and other fly lines during generation of homozygous underdominant individuals, making recombination and selection of such alleles out of their genomes possible; the attP insertion lines were not at all introgressed, and might have lower fitness than the transgenics as a result. This can be mitigated by

outcrossing the attP line individuals to white minus for a few generations and then homozygosing them to create a comparatively introgressed stock for future drive experiments. However, as proof of principle, the experiments presented here clearly show the threshold-dependent property of underdominant gene drive.

3.3 Future Directions

Underdominance with CG5266

The first step to making either rescue scheme work is to assure that one copy of the recoded transcript is sufficient to counteract toxicity of both shRNA alleles, since they both will be necessarily present in transheterozygous animals. Hopefully, the ongoing experiments with novel UTR-targeting shRNAs and translationally enhanced rescues will achieve this goal. (If they do, and healthy killer-rescue double homozygotes can be generated, it might be worthwhile to perform drive experiments using this system, as its relative simplicity might make it easier to transfer to vector organisms than more complex underdominance schemes; Gould et al. 2008). If the current approach does not yield a rescue tolerant of two toxin alleles, it will be important to figure out why. It is possible that expression of CG5266 driven by either oenocyte enhancer is simply not high enough to restore endogenous transcript levels, even with translational enhancement. In this case, multimerization of the antidote gene should be considered, as well as use of its endogenous promoter for rescue expression.

Once single-component rescue is robust, approaches used to distribute it between two alleles can be revisited. For the intein-catalyzed protein splicing rescue, efficiency and speed of the splicing reaction will be critical to success. If this is thought to be limiting to rescue function, the system can be rebuilt using several newly characterized split inteins that possess greatly increased splicing rates and efficiencies (Carvajal-Vallejos 2012). Also, additional split rescue/intein pieces can be added to each construct to boost rescue protein levels.

Functionality of the flipping rescue will depend on whether FLP is able to efficiently reorient the inverted CG5266 fragment in oenocytes. If the rescue initially

fails, it will be prudent to come up with a method of easily measuring FLP efficiency in the target tissue, perhaps by using a fluorescent marker with an inverted portion (instead of CG5266) so that FLP activity produces a visible output. If FLP activity is found to be insufficient, translational enhancers can be added to boost expression, and another FLP copy can be appended to the original one via a self-cleaving 2A peptide (Szymczak et al. 2004).

It should be noted that for both rescue approaches, it is essential to move the toxin shRNAs out of the 3'UTR of the antidote. Recent experiments in our lab have shown that presence of shRNAs in the 3'UTR of a transcript, even if those shRNAs are intronic, significantly reduces the expression levels of the protein encoded by that transcript. Moving the shRNAs to the 3'UTR of some benign transcript under expression of the same promoter should resolve this issue.

Underdominance with Rpl35a

The haplolethal approach to shRNA toxin-based underdominance has proven quite fruitful: with this method, we have engineered multiple underdominant lines (two of which are reproductively isolated from conspecifics and are therefore, in effect, novel species), and have also shown gene drive with these lines. I am hopeful that the final set of constructs currently being tested will give us a complete kill-complete rescue underdominant population as the conclusive piece of this proof of principle study.

Of course, achieving gene drive into wild type populations using this system would be desirable, and to this end we are currently looking for attP sites that do not appear to harbor any fitness costs. However, it is not clear to what extent such efforts are worthwhile, since Drosophila melanogaster is not a pest species and there would be no utility for an actual drive mechanism in this species. If further work in Drosophila is pursued to optimize this haplolethal-based underdominance approach, it may be advisable to consider other haplolethal genes for targeting, since Drosophila has a number of them (Cook et al. 2012; Marygold et al. 2007). Specifically, it would be desirable to find genes with well-defined promoter and UTR regions, so that sequence fragments used for the rescue portion could be carefully

chosen to assure proper expression. Also, with shRNA designs that provide improved silencing and promoters that kill well, it should be possible to target UTRs exclusively, so it would be prudent to find haplolethal genes that possess UTR regions amenable to recoding (i.e., not so highly conserved).

The ultimate goal, however, would be to transfer this technology to a vector species of interest. This will likely be difficult: the system was not easy to optimize even in Drosophila, whose genetics are well-understood, and required careful selection and empirical testing of various components (promoters, shRNAs, target sites, etc.) Nevertheless, Rpl35a is highly conserved in insects, and the work described here can serve as a blueprint for attempting to engineer this scheme in species of interest.

3.4 Materials and Methods

Fly Culture and Strains

Fly husbandry and crosses were performed under standard conditions at 25°C. BestGene (Chino Hills, CA) and Rainbow Transgenics (Camarillo, CA) carried out all of the fly injections, with Rainbow Transgenics performing the vast majority of them with consistent reliability.

Additional fly strains utilized in this study were attP lines 86Fa (#24486), 51D(#24483), 51C (#24482), and #9724, as well as several germline FLP and B3 recombinase-expressing lines described in Chapter 2.

Fly crossing schemes used to introduce FLP into the germline and set up kill and rescue crosses are detailed in Appendix C.

To measure fecundity and larval competitiveness of flies with Rpl35a rescue-only constructs, equal numbers of pre-mated 86Fa-Rpl35a females and 86Fa females (with an empty attP site) were kept on grape plates at 25°C for several days, with fresh yeast and plates given approximately every 12 hours. Both sets of females were then allowed to lay eggs on fresh grape plates for several hours at a time, and their eggs were washed into both separate fly food bottles and into the same bottle. The number of each kind of progeny that emerged from each bottle was then

counted, and was roughly equal for both types of females.

The process of setting up drive experiments is described in Appendix C.

Cell Culture

Drosophila S2 cells were maintained in Schneider's medium with 10% FBS, 1% penicillin and streptomycin at 27.5°C, and passaged every ~ four days. Transfections were performed with the FuGENE6 reagent (Promega, Madison, WI), using a ratio of 2.5µl: 1µg of FuGENE to DNA. An RFP marker was always used as a transfection control, and usually made up 20% of the DNA transfected. DNA used for transfection was typically prepared using Zyppy™ Plasmid Miniprep Kit (Zymo Research, Irvine, CA) and eluted in endotoxin-free TE buffer.

CG5266 Constructs

Cell Culture Constructs

All components were cloned into the multiple cloning site (MCS) of pAc5.1-HisB by enzymatic assembly (EA) methods, and were expressed by the Actin 5 promoter located upstream of the MCS. The three sets of intronic shRNAs were PCRed from Kelly Matzen's plasmids; the mRNA sequences they targeted are listed in Appendix B. Ub-R-GFP was PCRed from Addgene plasmid #11939 (Addgene, Cambridge, MA); pd2EGFP was PCRed from Clontech plasmid pd2EGFP-N1 (Clontech, Mountain View, CA). The Nostoc DnaE split intein (Zettler et al. 2009, Dhar and Mootz 2011) was codon-optimized for Drosophila expression and gene synthesized by Genscript (Piscataway, NJ). The full intein sequence used is provided in Appendix B. K. Matzen originally recoded CG5266, and it was PCRed from her plasmid. Its full sequence is listed in Appendix B. It was either cloned whole into pAc5.1-HisB, or was put in with appropriate intein pieces, so that the N-terminal portion of the sequence was cloned upstream of the IntN intein part, and the C-terminal potion was separately cloned downstream of the IntC intein part.

The first and last 30 base pairs (bps) of all DNA fragment used in the described constructs are listed in Table 3.2.

Fly Constructs

All constructs, unless otherwise noted, were cloned using EA methods. All shRNA sets were assembled via Golden Gate cloning as described in Appendix A, and were flanked by the same 5' and 3' intronic consensus sequences utilized in the original set (5' site cagGTaagt, 3' site tactaattcttcttttccctttttttAGg, with splice donor and acceptor capitalized). The initial CG5266-intein constructs were cloned into the attB Drosophila vector using the MCS. The GMR promoter (which contained an hsp70 basal promoter fragment) was PCRed from K. Matzen's plasmids, and was cloned 5' of either the minimal GFP blocker or the Tet cassette (details about both fragments are discussed in Chapter 2). The CG5266-intein fragments were cloned downstream of the blocker, followed by intronic shRNAs and the γ-tubulin 3'UTR (PCRed from lab sources). CTCF insulators (Kyrchanova et al. 2008) PCRed from K. Matzen's plasmids flanked the entire insert.

In later construct versions, additional miR-1-based shRNAs were cloned 3' of the initial set by traditional cloning methods using PshAI and AvrII. The handm enhancer (Han and Olson 2005), tinman enhancer (tinC; Lo and Frasch 2001), desaturase 1 enhancer (dst1; Billeter et al. 2009), and seven-up enhancer (svp; Gutierrez et al. 2007), PCRed from Hid plasmids, were inserted into the original constructs in place of GMR, with a minimal hsp70 promoter downstream of each fragment. In subsequent versions the MHC16 IVS intron (Pfeiffer et al. 2012) was also cloned directly 5' of each CG5266-intein fragment, and the γ-tubulin 3'UTR was replaced by the p10 3'UTR (Pfeiffer et al. 2012).

The flipping rescue constructs were assembled into the MCS of the attB vector as above. Originally, handm was cloned upstream of a blocker surrounded by B3 sites (discussed in Chapter 2). FLP was PCRed from existing Tet cassette constructs (Chapter 2) and was cloned 3' of the blocker sequence, followed by intronic shRNAs and the γ-tubulin 3'UTR. The recoded CG5266 gene was PCRed from K. Matzen's plasmid in two pieces; the FLEX switch (Schnutgen and Ghyselinck 2007) arrangement of FRT and F3 sites was added on via multiple rounds of PCR, and the two pieces were then cloned 3' of the blocker sequence via EA (a partial sequence of this arrangement is provided in Appendix B). The dst1 and svp enhancers were

cloned in place of handm as described above. The entire insert was flanked by CTCFs.

The kill-only constructs were cloned using the strategies and fragments described above, and consisted of (5' to 3') an oenocyte enhancer with a minimal hsp70 promoter, the GFP blocking sequence, dsRed (PCRed from Tet cassette constructs, described in Chapter 2), intronic shRNAs, and the γ-tubulin 3'UTR, all flanked by CTCFs. The rescue-only constructs contained an oenocyte enhancer with a minimal hsp70 promoter, the entire recoded CG5266 sequence, and the γ-tubulin3'UTR, and were also flanked with CTCFs. In later versions, the recoded CG5266 was replaced by an endogenous version. Also, the MHC16 IVS intron was added directly 5' of CG5266, and the γ-tubulin UTR was replaced by p10. New miR-6.1-based shRNA sets targeting the 5' and 3'UTRs of CG5266 (one four-mer for the 5'UTR and one for the 3'UTR) were designed and assembled as described in Appendix A, and cloned in place of the original set by traditional subcloning using AscI and SpeI. Sequences of their target sites are provided in Appendix B.

The FRT site and F3 site sequences, as well as the first and last 30 base pairs (bps) of all other DNA fragment used in the described constructs, are listed in Table 3.2. All transcribed proteins had a Kozak sequence (CAACAAA) 5' of the first ATG codon added via PCR (unless they were directly preceded by the MHC16 IVS intron, which has its own Kozak sequence).

Rpl35a Constructs

All constructs, unless otherwise noted, were cloned using EA methods. All shRNA sets were assembled via Golden Gate cloning as described in Appendix A, and were flanked by the same 5' and 3' intronic consensus sequences utilized in the original set (5' site cagGTaagt, 3' site tactaattcttcttttccctttttttAGg, with splice donor and acceptor capitalized). The initial Rpl35a constructs were cloned into the attB Drosophila vector using the MCS. The 5' toxin portion of the construct consisted of (from 5' to 3') the handm enhancer with a minimal hsp70 promoter, the short GFP blocker, a VP16-shRNA fragment PCRed from plasmid KJD161 (provided by K. Matzen), and the γ-tubulin 3'UTR. This entire region was flanked by CTCFs. The

Rpl35a rescue was cloned 3' of the second CTCF. It consisted of, 5' to 3', 1200 bps upstream of the Rpl35a ATG (PCRed from genomic DNA), the recoded CDs with introns (PCRed from KJD161), and the annotated Rpl35a 3'UTR plus 500 bp of sequence past the UTR (PCRed from genomic DNA).

In later versions of these constructs, the handm enhancer was replaced by dst1 or svp with a minimal hsp70 promoter as previously described. Also, VP16 was replaced with dsRed that had a Kozak sequence (CAACAAA) 5' of the first ATG codon (PCRed from Tet cassette constructs), and the original 250 bp γ-tubulin promoter fragment was replaced by a 1500 bp fragment (PCRed from genomic DNA).

Subsequent shRNA sets were designed and assembled by Golden Gate cloning as detailed in Appendix A. The sequences of all shRNA target sites (the original miR-6-based triplet cloned by K. Matzen; the first miR-1-based six-mer targeting the CDs; the two subsequent miR-1-based four-mers targeting three CD sites each plus the original 3'UTR site; and the final miR-1-based four-mer targeting the 3'UTR) are provided in Appendix B. One of the miR-1-based four-mers targeting the 3'UTR was designed such that the last 4 bps of the shRNA guide strand were the same as the target mRNA instead of being complementary to it; all other guide strands were 100% complementary to their target sites. The shRNAs were subcloned using AscI and SpeI. Rpl35a rescue versions were recoded via PCR and enzymatic assembly. Sequences of the recoded versions used for each set of shRNAs are listed in Appendix B.

The first and last 30 bps of all DNA fragments used in the described constructs are listed in Table 3.2.

Name	30 bps of the 5' and 3' ends of each fragment, 5' to 3'
Hsp70 basal	AGCGCCGGAGTATAAATAGAGGCGCTTCGT AAGCGCAGCTGAACAAGCTAAACAATCTGC
Hand minimal	CTTCAAAACTTCAATATTTCCAAAAAGGCA CGCCCCAGTGCAGAAAAAGGGGGAGAAAAT
tinC	CATGAACAGCTTTCGATCGGCCACCGATCC CAATGCAACTGTAAAATGTTTCCCTGCCTC
Dst1	CACAGGTTCGTTTTCCACCTGGCCACTTGT TAAGCATGCCACCAATTCTGGCAATTATTT
Svp	AAGCTTGGAAAATGCGTTGGAGTAATAGCC GGAGCACAAACACATGGGCCCTGCCACTTT
GMR	ATCCCCGATCCCCCTAGAATCCCAAAACAA CTTTAAGTCGACTCTAGAGGATCCGCGGCC
P10 3'UTR	CTAGAATGAATCGTTTTTAAAATAACAAAT TGGCTTGGATAGCGATTCGAGTTAACGGCC
MHC16 IVS	AAAAGGTAGGTTCAACCACTGATGCCTAGG GAGGGTACCAACTTAAAAAAAAAAATCAAA
FLP	ATGCCACAATTTGATATATTATGTAAAACA TTATATGCGTCTATTTATGTAGGATGAAAG
GFP	ATGGTGAGCAAGGGCGAGGAGCTGTTCACC ACTCTCGGCATGGACGAGCTGTACAAGTAA
dsRed	ATGGTGCGCTCCTCCAAGAACGTCATCAAG CTACAGGAACAGGTGGTGGCGGCCCTCGGT
Ub-R-GFP	CACCATGCAGATCTTCGTGAAGACTCTGAC ACTCTCGGCATGGACGAGCTGTACAAGTAA
pd2eGFP	ATGGTGAGCAAGGGCGAGGAGCTGTTCACC GCCTGTGCTTCTGCTAGGATCAATGTGTAG
CG5266 C Fragment	CGTCCGTATCTCTACCAATCCGATCCTTCG ATCAAGGACTACTTGGCCAGCATCCCCTAG
CG5266 N Fragment	ATGGCTACCGAACGATACAGCTTTTCGTTG TCCCTACTGATCTGCGGCTGGGACAATGAT
CTCF	CCTTGCAGCGCCACCTGGCCGCGAAGAGTT GGCCAGGTGGCGCTGCAAGGTGTGAGTTGT
VP16	GCGGCTCCGGCTGCCAAAAAGAAGAAACTG TTACAGTGGGCCGTATCGTGTCCGATACTG
Rpl35a CDs	ATGGCTGACACACAAGCCAAGTCCACTACT TTGCAGATGCTGTACCCATCAAGGATTTAA
Rpl35a 5' Region	AATCTTGTATACCCCATTACTCTGCAGATA AGATTATCTTTTCTCATCCACAGCTAAGCC
Rpl35a 3' Region	GTTAATATCCGACTTGAATTACTGACCTGC CGAATATACCACCTAACATATTTTTAGAAG
γ-tubulin 3'UTR	TCGCTTGTGCCAGAAGAAATGCGTGCCATG GACACATTAAATAAAAAACAGTTTTTAAAT
γ-tubulin extended 3'UTR	TCGCTTGTGCCAGAAGAAATGCGTGCCATG TATTCTGTTTCGAGGTCCATGATTTTAAAT

Table 3.2: The first and last 30 bps of longer fragments used in CG5266 and Rpl35a underdominance constructs.

Chapter 4
Translocations

4.1 Introduction

The potential caveat of underdominance systems described in the preceding chapters is that they can break down after release. Even with utilization of evolutionarily robust shRNA toxins, failure of the system can occur in several ways: the protein or shRNA toxin may mutate and cease to function, which will result in selection against the intact underdominance allele, because the mutant version of the chromosome will always survive; the rescue may also become unlinked from the toxin, which would also cause the drive mechanism to break down for similar reasons; finally, the genes of interest (e.g., disease refractory genes) could become unlinked from the drive mechanism, and any achieved strain replacement would be of no practical use. Since any large-scale insect release is a time and resource-intensive undertaking, it would be desirable to engineer a drive system with very little chance of breakdown, so that future releases to replace or eradicate non-functioning alleles can be minimized.

This chapter details our development of an underdominance-like drive mechanism that is inherently robust on evolutionary time scales.

Translocations: Conceptual Framework

A reciprocal chromosomal translocation is an exchange of chromosome pieces between two non-homologous chromosomes (Foster et al. 1972; Robinson 1976); a balanced translocation simply means that the exchange is even and mutual, with no excess or missing genes. Reciprocal translocations can occur naturally (for example, their incidence is ~ 1/625 in the general human population; Van Dyke et al. 1983) or can be induced by mutagens (e.g., Lyon et al. 1972). A heterozygote for a reciprocal translocation should theoretically be normal, since it has a balanced chromosome set (Figure 4.1A; of course, translocations that disrupt gene function can impose fitness costs on their carriers). However, when out crossed to wild type, only 50%

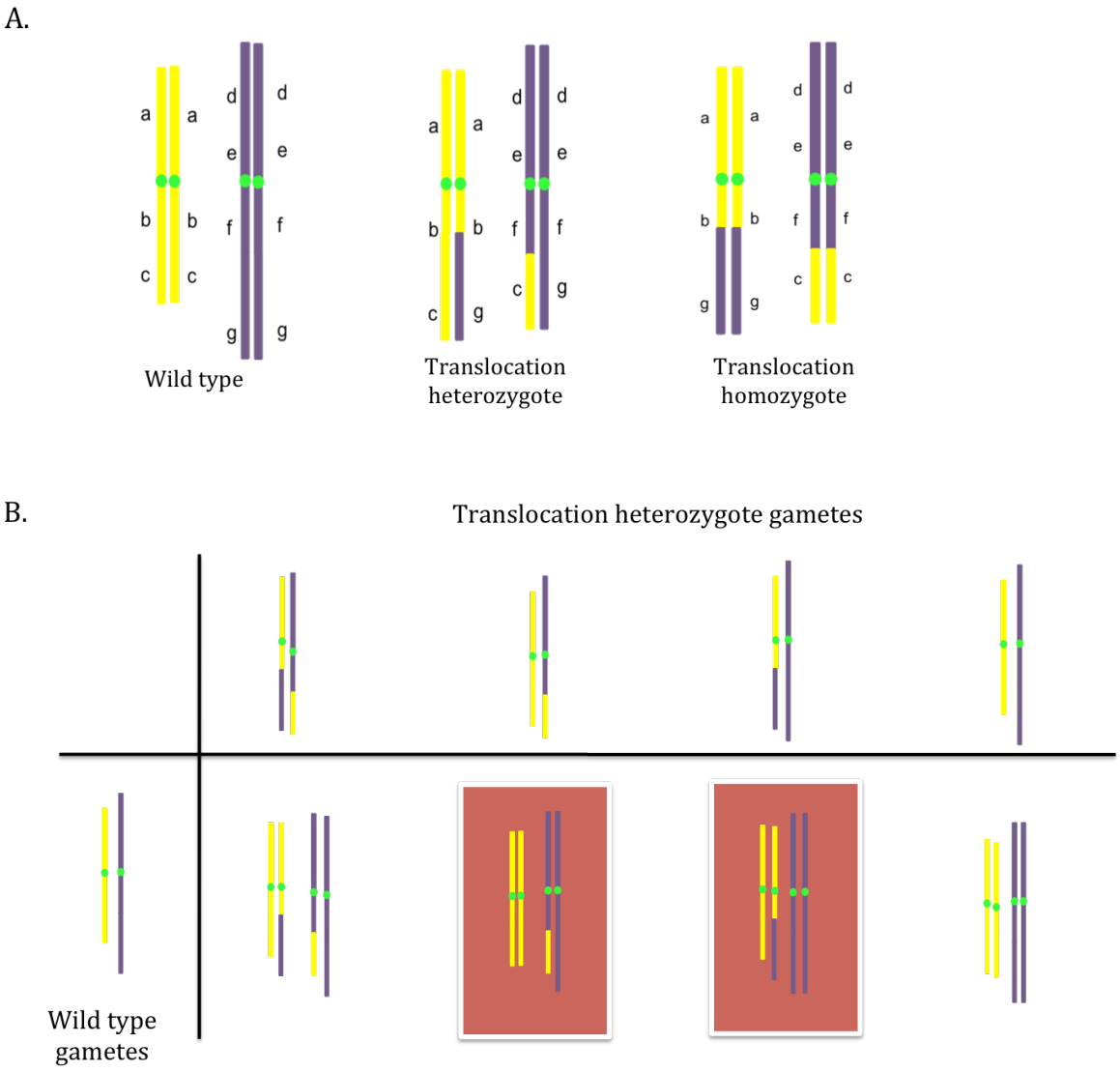

Figure 4.1: A. A set of wild type (left), translocation heterozygous (center), and translocation homozygous (right) chromosomes. Although translocation-bearing individuals will have rearranged chromosomes, they will have equal gene copy number and thus be viable. B. When a translocation heterozygote mates with a wild type, 50% of the resulting offspring will be aneuploid and inviable (shaded in red), 25% will be wild type, and 25% will be translocation heterozygotes.

of the translocation heterozygote's progeny will be viable (Figure 4.1B). This can be visualized as follows: if the genotype of a given translocation heterozygote is T_1/+; T_2/+ (where "+" denotes wild type), it will produce four kinds of gametes in equal proportion – (T_1,+), (T_1, T_2), (+,+), and (+,T_2). When combined with wild type gametes – (+,+) – only the (T_1, T_2) and (+,+) will give rise to viable progeny (translocation heterozygotes and wild types, respectively). The other gamete types will perish in progeny that have unbalanced chromosome sets. If the starting translocation individual is homozygous (T_1/T_1; T_2/T_2), all first-generation progeny in a wild type outcross will survive, but further mating of these progeny to wild type would result in F_2 generation lethality.

Drive Characteristics

Unlike the underdominant schemes described earlier in this work, the basis of translocation-mediated gene drive is not lethality of the heterozygotes themselves, but rather death of some of their progeny. The system is similar to classical underdominance, however, because it functions like a bi-stable switch: translocations either spread to fixation or disappear from the population depending on their initial frequency and associated fitness costs (as compared to wild type).

When the frequency of translocation-bearing individuals (T_1/+; T_2/+ and T_1/T_1; T_2/T_2) is low, they are more likely to mate with wild types (+/+; +/+) than with each other, and many of the translocation chromosomes end up being lost in inviable progeny or grandchildren (ones with an unbalanced chromosome set, such as T_1/+; +/+). Wild type chromosomes are lost in these progeny, too, but since wild type individuals are more abundant, they mostly mate with each other, producing more viable wild types (+/+; +/+), and eventually T chromosomes are lost from the population. Conversely, if TT animals are more abundant, they largely mate with each other, producing a high frequency of T_1/T_1;/T_2/T_2 offspring, while wild type chromosomes are regularly lost in inviable individuals with unbalanced chromosome sets (such as +/T_1; +/+). In this scenario, T alleles become fixed in the population, while + alleles are lost.

For a translocation with no fitness cost, a 50% population frequency represents an unstable equilibrium (Curtis 1968); if the translocation's frequency is greater than 50%, it will spread to fixation, and if it's less than 50%, it will disappear from the population (Curtis 1968). Thus, releases of high fitness translocation homozygotes (with fitness ~ 1) at frequencies above 50% should lead to gene drive, while releases below that threshold should lead to loss of translocation alleles. If there is a fitness cost associated with the translocation, the release frequency will be correspondingly higher (Curtis 1968).

Advantages

As discussed in Chapter 2, threshold-dependent gene drive can be of considerable advantage in initial tests of insect population replacement or when public approval of transgenic releases is limited (Marshall et al. 2010a; Marshall and Hay 2012). A system with a substantial introduction threshold is easier to confine to targeted release areas—for example, theoretical analyses of translocation containment (Marshall and Hay 2012) suggests that even as a translocation with a modest fitness cost spreads to fixation in one population, it would not rise to greater than 4% frequency in a neighboring population given a 1% migration rate. It is also removable, since diluting translocation allele frequency below the release threshold through continued release of wild type individuals will cause the allele to be lost from the population (Marshall and Hay 2012). The release threshold for translocations, while high, is still much lower than some of the release frequencies required by sterile insect technique (SIT; Krafsur 1998; Gould and Schiekelman 2004). And, unlike SIT, translocations are self-propagating from generation to generation, while SIT is an inundative approach that must be repeated on a regular basis (Baker 1984; Robinson 1976). In short, while the initial costs associated with release of a high threshold gene drive mechanism are greater than those associated with low-threshold gene drive mechanisms such as Medea, the increased level of control over population fate and spread are likely to make it attractive in many real world settings.

Another major benefit of translocation-dependent drive is its tremendous robustness in the face of mutation. Unlike the toxin-antidote schemes described in previous chapters, a translocation is itself both the toxin and the antidote, since the presence of one copy results in lethality for some progeny, while the presence of two copies in homozygotes guarantees survival of all progeny (when outcrossed to wild type). Therefore toxin and rescue functions cannot be unlinked, and so breakdown of the system due to toxin inactivation is not possible; and in any case, reversion of the translocation chromosome to its original arrangement is exceedingly improbable. Even if reversion did occur (which would necessarily be in one individual), this chromosome would find itself in a sea of translocation homozygotes and would rapidly be eliminated from the population.

Additionally, since the gene of interest can be inserted at the translocation breakpoints, where meiotic recombination is inhibited (Sherizen et al. 2005), it is not likely to become unlinked from the translocation (the drive mechanism). If extra protections are desired, the translocation (and the cargo gene of interest) can be created so that one of its breakpoints is in an inversion, essentially eliminating the possibility of recombination with a wild type chromosome (Curtis 1968; Egli et al. 2004).

It is important to acknowledge that we can never prevent mutation to inactivity of the gene of interest, which will occur at a frequency dependent on gene size and the organism-specific mutation rate. The fate of a chromosome carrying such a mutation will depend on its fitness relative to that of chromosomes carrying intact copies of these genes. As discussed in the introduction, we generally assume that any gene conferring disease refractoriness will result in a fitness cost to carriers. The critical implication of this is that mutant-bearing translocation chromosomes will spread over time. However, the epidemiological consequences of this for disease resurgence can be minimized in translocation-based gene drive because translocation homozygotes (the state to which the population is driven) will carry at least two copies of each gene of interest.

Finally, it may be easier to achieve public acceptance of gene drive based on translocations than with previously discussed mechanisms. The considerable public

discomfort with genetically engineered organisms is likely to extend to any planned releases of genetically modified insects (Gould et al. 2006), and even those most likely to benefit from disease eradication will likely have concerns and hesitations (Marshall et al. 2010a). A translocation-based approach is, in my opinion, more likely to win general approval as a drive system for several reasons: translocations are a naturally occurring genetic aberration present at significant frequencies in populations of many different organisms; the words "toxin" and "antidote" need not be used; and finally, they can be removed from the population, restoring the pre-transgenic state, through dilution of the replaced population with wild type males.

Historical Perspective

Use of translocations to suppress or manipulate insect pest populations had been deliberated and modeled for much of the 20th century (reviewed in Gould and Schliekelman 2004). Initially suggested for purposes of population suppression (Serebrovskii 1940), utilization of translocations for gene drive was proposed in 1968 (Curtis) and generated a significant amount of interest through the 1970's and 1980's (Gould and Schliekelman 2004). Computer models simulating dynamics of translocation releases were developed (Cutis and Robinson 1971; Whitten 1971), and a number of lab and field cage experiments in various species were carried out (Robinson and Curtis 1973; reviewed in Asman et al. 1981 and Baker 1984). However, despite some progress, this area of research was ultimately abandoned (Gould et al. 2006).

The reasons for this were twofold: firstly, translocation individuals typically had lower fitness than their wild type counterparts (Gould and Schliekelman 2004; Gould et al. 2006), likely because they were generated using X-rays and other mutagens that can reduce robustness, survival, and mating competitiveness (Jacobs-Lorena 2004). Secondly, because of the random nature of the above methods for creating chromosomal rearrangements, there was no way to link a gene of interest to the translocation. These factors greatly limited the potential for generating fit chromosomal translocations, and for using them to drive anything useful into wild populations.

Advances in molecular genetics over the last few decades, especially the ability to carry out germline transformation of Drosophila (Spradling and Rubin 1982) and other insects (reviewed in Robinson et al. 2004), have led to the creation of novel ways of manipulating pest genomes in a precise, controlled fashion. These developments have allowed researchers to accomplish a number of objectives related to genetic pest control, from creation of a potent drive mechanism in fruit flies (Chen et al. 2007) to implementation of genetic alternatives for sterile insect production (Thomas et al. 2000), to generation of mosquitoes with greatly diminished potential to transmit disease (Ito et al. 2002; Moreira et al. 2002; Kim et al. 2004). They have also permitted the development of several techniques—based on FLP/FRT recombination (Beumer et al. 1998), Cre/*loxP* recombination (Yu and Bradley 2001; Egli et al. 2004), or homologous recombination following double-stranded breaks (Egli et al. 2004; Uemura et al. 2010) —to engineer translocations with sequence-specific breakpoints that can contain any gene of interest.

In the following work, I describe the generation of site-specific translocations in Drosophila, and the analysis of those translocations as mediators of gene drive.

4.2 Results and Discussion

We decided to employ the homologous recombination (HR) technique for generating translocations described in Egli et al. 2004, largely because those researchers determined it to be more efficient than other approaches involving the use of site-specific recombinases such as Cre or FLP. Briefly, this method involves generating double-strand breaks (DSB) in transgenes on two different chromosomes. The created broken ends are designed so that they that have homology with each other, facilitating the formation of a translocation chromosome through homology-directed repair. Individuals carrying recombinant chromosomes are then identified based on recombination-dependent creation of a visible marker. Our adaptation of this method is detailed below.

Construct Architecture

The basic structure of the two alleles used to generate translocations is shown in Figure 4.2. Each construct has a promoter driving a fluorescent marker: construct A has the ubiquitous baculovirus opie2 promoter (provided by O. Akbari; adapted from Theilmann and Stewart 1992) driving dsRed expression, while construct B has the oenocyte-specific svp enhancer (Gutierrez et al. 2007) driving GFP. In between each promoter and fluorescent marker is a large intronic stuffer containing a piece of DNA—fragments of the IgG variable sequence and the mouse IgG heavy chain constant region (provided by J. Li) that are foreign to Drosophila—surrounded by splice donor and acceptor sites. The stuffer region is broken up into two fragments of equal length (randomly labeled UVW and XYZ, for clarity), and has two sites recognized by the rare-cutting I-SceI restriction endonuclease positioned in the middle. The intronic stuffer is the same in both constructs, except that in construct A, fragment XYZ is on the left and UVW is on the right, and in construct B this order is reversed. Both alleles are flanked by artificial CTCF insulators (Kyrchanova et al. 2008) to minimize opportunities for the local genomic environment to influence gene expression.

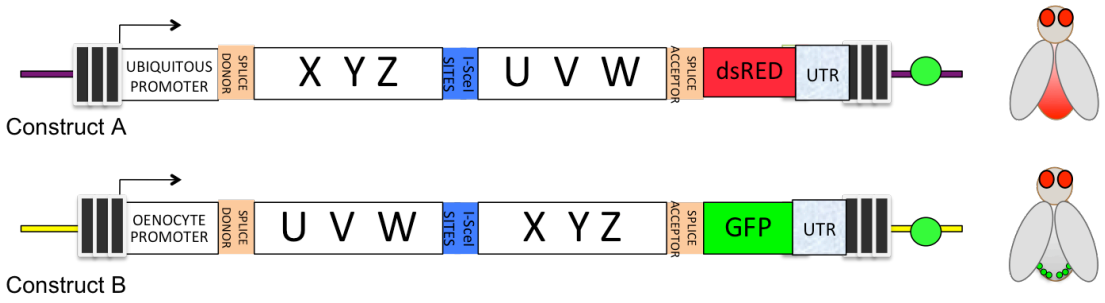

Figure 4.2: General architecture of the two translocation constructs. Each allele has a promoter driving a reporter, with an intronic stuffer region of homology (XYZ-UVW) with I-SceI restriction sites in the middle. The stuffer regions are identical between the two constructs, except that their arms are switched sides. Both constructs are flanked by CTCFs (black bars). The red body of the fly with construct A represents ubiquitous opie2-driven expression of dsRed, while the green dots on the torso of the fly bearing construct B represent svp-driven expression of GFP in the oenocytes.

If both of these constructs are introduced into the same fly on non-homologous chromosomes, the animal should have ubiquitous expression of dsRed and oenocyte-specific expression of GFP (Figure 4.3). Adding a source of I-SceI should produce double-stranded breaks in both transgenes on both chromosomes. Cells seek to quickly repair DSBs, as they pose a threat to the integrity of genetic information, and often do so by finding regions of homology to use as a template, so that the original sequence is preserved (Egli et al. 2004). In Drosophila, the sister chromatid and the homologous chromosome are the favored templates for repair (Rong and Golic 2003); however, ectopic sources are also sometimes used, as DSBs are capable of finding repair templates anywhere in the genome (Gong and Golic 2003; Egli et al. 2004).

Given the above construct configuration, I-SceI-induced cleavage will, in a small percentage of cases, lead to recombination between the UVW-bearing fragments to generate one translocation chromosome, and between the XYZ-bearing fragments to generate another translocation chromosome (Figure 4.3). Translocation-bearing individuals can be recognized by virtue of a color switch, since each promoter will now drive a novel reporter: they will have ubiquitous expression of GFP and oenocyte-specific expression of dsRed (Figure 4.3). Isolated individual translocation heterozygotes can then be crossed out to wild type to generate more translocation heterozygotes, and these can be further crossed to produce homozygous translocation stocks.

Selection of Insertion Sites

Selecting appropriate genomic insertion sites for the described translocation alleles is crucial, as not all site combinations will permit generation of translocation individuals. Besides producing reliable transformation efficiencies and robust expression levels, selected sites have to meet two criteria: they have to be located in gene deserts, and oriented in a specific direction with respect to each other.

Ideally, insertion sites will be located as far away from genes as possible so that the translocation event will not disrupt any gene expression or function (if essential gene function is perturbed, translocation-bearing individuals may be unfit or even

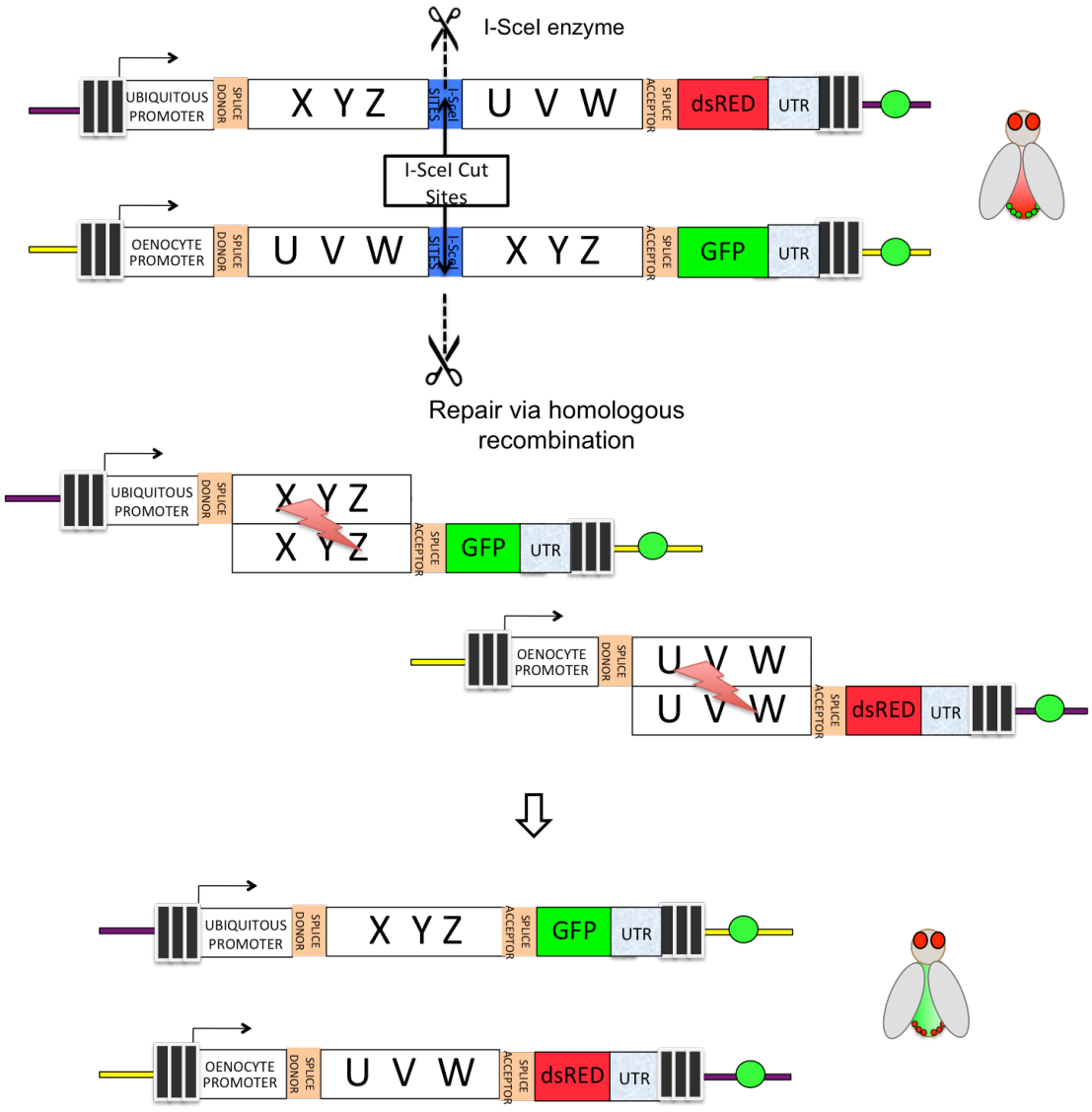

Figure 4.3: When both translocation constructs are in the same fly, it will have ubiquitous dsRed expression and oenocyte-specific GFP expression. When a source of I-SceI is introduced, it will cleave in the center on the region of homology, creating a double-stranded break (DSB). In a small percentage of cases, the DSB will be repaired by homologous recombination between stuffer fragments on different chromosomes, creating a reciprocal translocation. Flies bearing the translocation chromosomes will now have ubiquitous GFP expression and oenocyte-specific dsRed expression.

inviable). As far as their directionality, the two translocation alleles have to be oriented in the same direction (with respect to their centromeres) if they are on the same arms of the (nonhomologous) insertion chromosomes, and in opposite orientation if they are on different arms of the chromosomes. Only these construct orientations will allow for the creation of balanced translocations (Figure 4.4). In the other two possible orientations, HR would produce one acentric chromosome and one with two centromeres, both of which would result in the creation of inviable individuals.

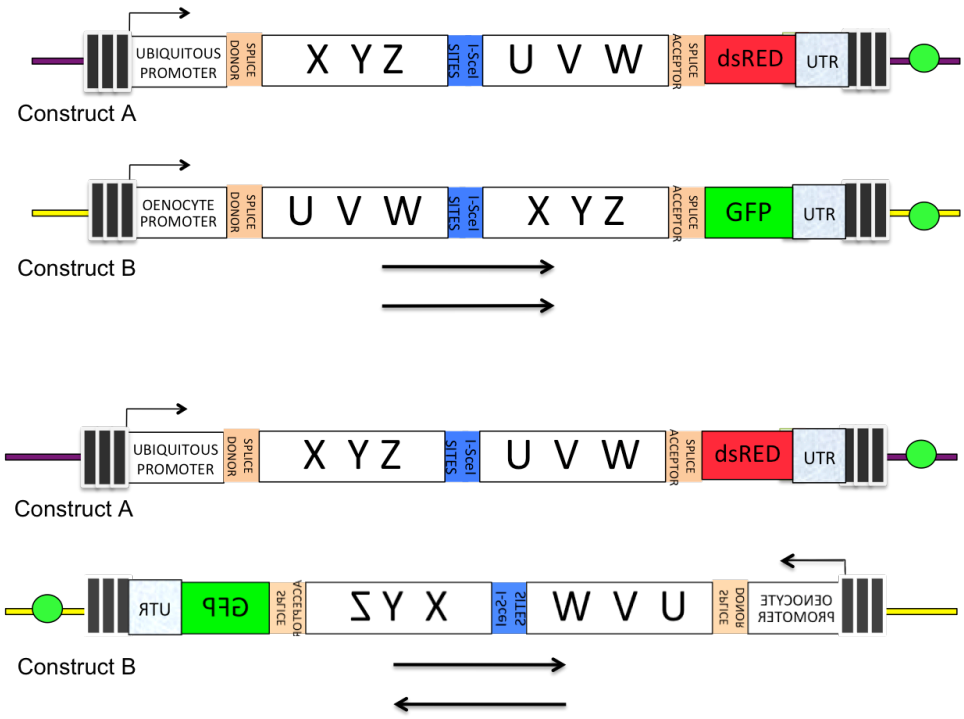

Figure 4.4: In order to generate viable reciprocal translocations, the two translocation constructs have to either be on the same arm of non-homologous chromosomes in the same orientation, or on opposite arms of non-homologous chromosomes facing in opposite directions with respect to each other. Any other arrangement of the constructs will cause homologous recombination to produce one acentric chromosome and one with two centromeres, and individuals bearing these will not be viable.

To find suitable insertion sites, the available attP lines reported to give acceptable transformation rates were surveyed to find ones located in gene deserts. After five potential candidate lines were identified, I obtained fly stocks for each line and extracted genomic DNA so that the orientation of the attP sites (and thus the orientation our constructs would be inserted) could be ascertained. Ultimately, three attP lines were chosen for transformant generation—one with the insertion site on the second chromosome and two with the site on the third chromosome—so that two distinct types of translocation individuals could be produced.

Fly Results

In the initial set of translocation constructs, I utilized the 5' and 3' splice sites from an intron of Rpl35a, since I was working with this gene already, it was highly expressed, and the chosen splice sites were empirically predicted to be quite strong. However, transformants for these constructs at all three chosen attP sites showed no visible GFP or dsRed expression. Since we knew that for at least one of these sites, which we've previously worked with, expression of fluorescent reporters was typically visible, we suspected that inefficient splicing was responsible for lack of marker expression. To verify this, I tested the translocation constructs (with the Actin5 promoter replacing the tissue-specific ones) in Drosophila S2 cell culture, and found that there was, indeed, no expression of the fluorescent markers. I then replaced the Rpl35a intron with an MHC16 intron, which was reported to function well in Drosophila (Pfeiffer et al. 2010) and which I had tested successfully in S2 cell culture, to produce a second generation of translocation constructs.

Transformants bearing the second-generation opie2-driven dsRed translocation allele did show ubiquitous expression of red fluorescence, as expected. However, the svp-GFP allele still produced no detectable GFP signal in the fly oenocytes. We conjectured that, although the splice sites were now functional, something about the particular orientation of the UVW-XYZ stuffer fragments in the svp construct (since that orientation differed from the orientation in the opie2 construct) still prevented proper splicing. (Our suspicions were later largely confirmed when, during post-I-SceI cleavage, we found a fly line that expressed both opie2-dsRed and svp-GFP; the

cleavage must have led to removal of part or all of the stuffer fragment, which permitted for proper splicing.) Nevertheless, since we only needed one reporter—because in principle, individuals with one translocation chromosome must carry the other one to survive—we decided to proceed with fly experiments.

Transformants bearing each of the translocation alleles were crossed with each other and with a heat shock driven I-SceI line from the Bloomington Stock Center (Bloomington, IL) to create a stock with both translocation alleles and a source of I-SceI. Adult flies from this stock were then heat shocked repeatedly as they were mating and producing eggs, and progeny received numerous heat shocks during early larval development. These progeny, a number of which had clonal groups of cells that showed the translocation phenotype (i.e., spots of GFP expression), were further out crossed to each other, and their offspring were screened for ubiquitous GFP expression.

For the first combination of insertion sites, numerous suspected translocation-bearing individuals were identified by opie2-GFP expression (as expected, they did not have svp-driven dsRed expression, since svp-GFP was not visible in parent flies). These individuals were first outcrossed to white minus flies for several generations, and then balanced and crossed to each other to generate homozygous translocation individuals. Homozygotes were obtained at expected ratios and appeared quite healthy and robust (although some of them did display a slight curling of the wings), and a homozygous stock was established. The type of genomic PCR confirmation I could do to verify that reciprocal translocations were actually present in these flies was limited: both translocation constructs had large areas of sequence flanking the specific allele components in such a way that PCRing across the entire insertion, or even from a specific promoter to its novel 3' chromosome end, was not possible. I was, however, able to confirm that each promoter was now associated with a different reporter, and that flies had insertions at both non-homologous attP sites (which wasn't visually clear, since only one of the translocation allele reporters was visible) via PCR. Outcrosses of translocation heterozygotes to white minus also showed that roughly half of the resulting embryos died, which would be expected from a true translocation heterozygote.

For the second combination of sites, numerous opie2-GFP expressing individuals were also found, and the presence of translocation chromosomes confirmed by PCR analysis and heterozygous outcrosses as above. To save time, I outcrossed all of these to balancer lines without carrying out multiple generations of outcrossing to white minus (which was done for the translocation discussed previously), and then to each other to obtain translocation homozygotes. In this case, however, very few homozygous individuals were recovered, and most of them appeared sickly. It is possible that changing chromosome structure at these particular sites imposes a significant fitness cost on translocation-bearing individuals, and some insect translocation homozygotes have been previously reported to have very low fitness (Robinson 1976; Asman et al. 1981). However, I am currently crossing these translocation individuals to white minus to see whether introgression with a wild type background for a few generations will allow us to recombine and/or select out any possible deleterious recessive mutations, and thereby increase translocation homozygote vigor (Asman et al. 1981).

Drive Experiments

As previously noted, many attP lines have intrinsic fitness costs that disadvantage them in competition with wild type, and therefore drive experiments involving transgenics created at these attP lines have to be done against the empty attP lines themselves.

With this in mind, I set up initial drive experiments for the translocation line that was healthy in the homozygous state against one of the empty attP lines used in generating that translocation. For translocation homozygotes with no fitness cost, drive could theoretically be expected if the frequency of translocation individuals was greater than 50% of the population (Curtis 1968). Since translocation homozygotes do frequently have fitness costs (Asman et al. 1981), I utilized significantly higher initial release frequencies (75% or 80%) and did observe gene drive (Figure 4.5). This drive was not as forceful as that observed with the Hid and Rpl35a underdominant systems—by the fourth generation, only three of the eight biological replicates showed complete phenotypic fixation of the translocation allele

—but this is expected given the different dynamics of the two systems (Marshall and Hay 2012).

I then set up further drive experiments above and below the conjectured 50% threshold. The experiments are currently ongoing, but preliminary data (Figure 4.5) show that translocation alleles are spreading into populations even when introduced below the frequency notionally required for drive. It is not presently clear whether loss of translocation alleles from the population will be demonstrated with the sub-threshold releases. The translocation-associated phenotype is expected to initially increase upon release, since matings of translocation homozygotes to

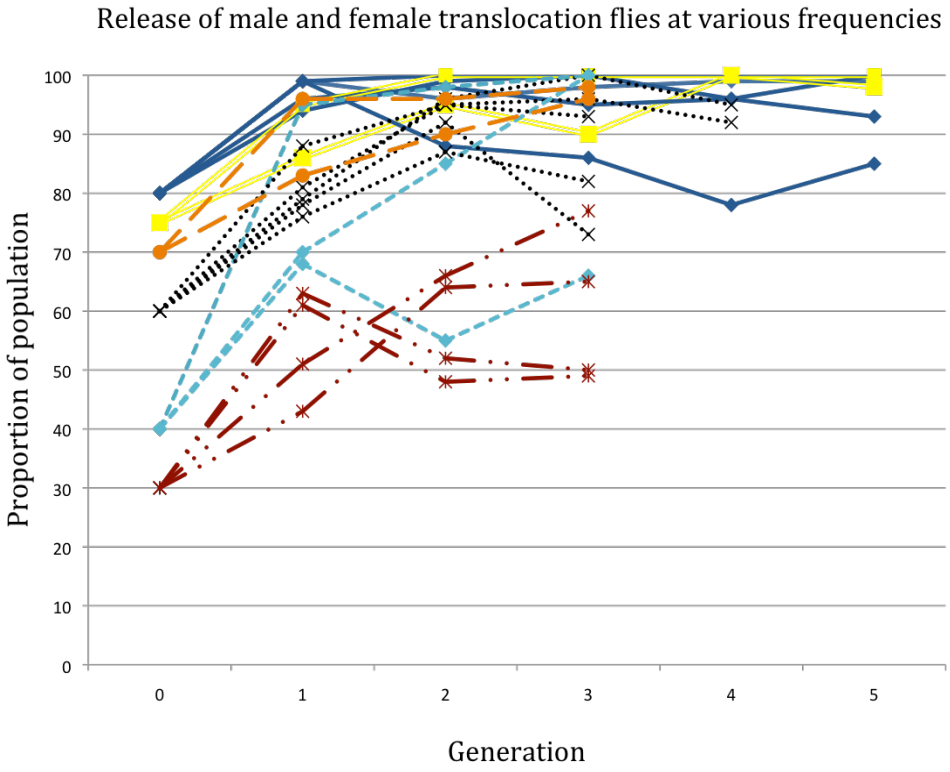

Figure 4.5: Drive experiments were set against one of the empty attP insertion lines (the one that appeared healthier) used to generate translocation transgenics by releasing homozygous male and female translocation individuals at various frequencies (from 30%-80%). Each line represents a biological replicate. Most replicates are so far showing spread of translocation alleles in the population, even when released at frequencies below the predicted drive threshold of 50%. The experiments are still ongoing, however, and it is still unclear whether drive out as well as drive in of translocation alleles will be achieved.

wild type first produce viable heterozygotes and lethality isn't observed until these individuals mate further, and a few replicates do seem to be showing a decrease in translocation frequency following an initial increase. In any case, however, it is clear that the translocation system produces stronger than expected gene drive.

This is not likely due to the type of eye color-related advantage discussed for Hid and Rpl underdominant systems (Chapters 2 and 3), as the translocation flies have light orange eyes (due to weaker expression of the *white* gene at the specific attP insertion sites) and are not expected to have significantly better vision than the white-eyed attP lines. However, since the translocation individuals had been outcrossed to white minus flies for a number of generations prior to homozygosing, it is quite possible that they might have had some of deleterious recessive alleles associated with the attP lines recombined and/or selected out, giving them a fitness advantage over the attP insertion site flies (as discussed in Chapter 2). Indeed, the homozygous translocation stock does seem somewhat more robust than either of the attP lines used in its generation. Outcrossing attP line individuals to white minus for several generations and then homozygosing them again (or, ideally, creating a double-homozygous stock containing both of the attP insertion sites present in the translocation homozygotes) should generate a comparable introgressed stock that can be used for future drive experiments.

4.3 Future Directions

The efforts described here serve two purposes: to demonstrate, as proof of concept, that translocations can be specifically generated and used as modulators of gene drive in insects; and to inform future efforts to employ this strategy in species of interest by identifying potential pitfalls and difficulties.

This research clearly shows that choice of insertion sites is crucial to generating utilizable translocation individuals. In Drosophila, it might be worthwhile to attempt creating translocation stocks using completely fit insertion sites, so that drive experiments can be conducted against wild types in an unambiguous proof of principle experiment that conforms to predicted thresholds. If there is a dearth of

existing fit insertion lines, such lines can be created using one of the emerging genome editing technologies, such as the CRISPR-Cas9 system (reviewed in Ran et al. 2013). In fact, this might be preferable to using pre-existing lines, as insertion sites can be specifically created in locations deemed optimal in terms of minimizing effects on local gene expression.

The ability to generate appropriate insertion sites will likely be the greatest limiting factor in transferring this technology to vector species, largely because it is time-intensive to generate numerous insertion lines and test them for general fitness. However, that should not discourage researchers from attempting it. Given the simplicity of the described translocation-generating system and its components, I personally believe it will be easier to engineer in mosquitoes and other vectors than the various forms of underdominance described in earlier chapters (a lot of the fragments used here, including the opie2 promoter and fluorescent markers, are directly transferable to Aedes aegypti, for example). Insertion sites must be chosen carefully, the various components (promoters, introns, etc.) must be tested empirically to make sure they function properly, and extensive crossing and screening must be done to isolate translocation individuals, since many vector species do not have balancer chromosomes to simplify crossing schemes. However, if the appropriate components are generated, it should be entirely possible to build transgenic translocation-bearing vector populations capable of driving genes of interest into wild populations.

4.4 Materials and Methods

Fly Culture and Strains

Fly husbandry and crosses were performed under standard conditions at 25°C. BestGene (Chino Hills, CA) and Rainbow Transgenics (Camarillo, CA) carried out all of the fly injections, with Rainbow Transgenics performing the vast majority of them with consistent reliability.

Additional fly strains utilized in this study were attP lines 22A (Bloomington Stock Center #24481), 68E (#24485), 96E (#24487), 51C (#24482), and #9741.

Stock #6935 was used as the source of heat shock I-SceI. Stocks #39631 #2555 were used for all balancing.

The translocation constructs were inserted into three sites—51C on the second chromosome and 68E and 9741 on the third chromosome—so that two distinct translocation types could be generated (one from the 51C/68E combination, the other from the 51C/9741 combination). Healthy homozygous stocks were generated for the 51C/68E site combination, and drive experiments for these stocks were set up against the 51C attP line, as it was the healthier of the two. The 51C/9741 translocation homozygotes did not appear to be 100% healthy (as discussed in more detail in the previous section).

Heat shocks were initially carried out by placing fly vials in a fly incubator set at 38°C for one hour. However, after only one translocation individual was recovered, heat shocks were subsequently performed by placing fly vials in a 38°C water bath for one hour, in hopes of increasing the efficiency of heat shock. Larvae were heat shocked ~ five times during early development. Offspring of heat-shocked larvae were screened for ubiquitous expression of GFP.

The crossing scheme used to bring both translocation alleles and the I-SceI source into a single fly line is detailed in Appendix C. The process of setting up drive experiments is also described in Appendix C.

Cell Culture

Drosophila S2 cells were maintained in Schneider's medium with 10% FBS, 1% penicillin and streptomycin at 27.5°C, and passaged every ~ four days. Transfections were performed with the FuGENE6 reagent (Promega, Madison, WI), using a ratio of 2.5µl: 1µg of FuGENE to DNA. A GFP or RFP marker was always used as a transfection control, and usually made up 20% of the DNA transfected. DNA used for transfection was typically prepared using Zyppy™ Plasmid Miniprep Kit (Zymo Research, Irvine, CA) and eluted in endotoxin-free TE buffer.

Construct Assembly

Translocation allele components were cloned into a multiple cloning site (MCS)-containing Drosophila attB backbone using the enzymatic assembly (EA) cloning method described by Gibson et al. 2009 (detailed in Appendix A). For construct A (Figure 4.2), the opie2 promoter fragment was PCRed from plasmid OA791 (provided by O. Akbari), and dsRed and SV40 3'UTR were amplified from various constructs described in Chapter 2. The UVW and XYZ stuffer fragments were PCRed from a plasmid provided by J. Li, which she generated by cloning an IgG variable region (Nath 2003) upstream of the mouse IgG heavy chain constant region contained in plasmid pFUSEss-CHIg-mG1 (Invivogen, San Diego, CA). Two I-SceI recognition sites were added between the fragments via PCR, 5' and 3' splice sites from an Rpl35a intron were also PCRed onto the fragment ends, and the resulting stuffer region was then inserted between opie2 and dsRed. For construct B, this process was identical except that the UVW and XYZ fragments were arranged on opposite sides (UVW on the left, XYZ on the right). The svp enhancer plus hsp70 basal promoter, GFP, and SV40 3'UTR for construct B were PCRed from plasmids described in Chapter 3. Both GFP and dsRed had a Kozak sequence (CAACAAA) directly 5' of the start codon. Both translocation alleles were flanked by CTCFs, which were PCRed from constructs described in Chapter 3.

To test splicing of the above constructs in S2 culture, the tissue-specific promoters were replaced by Actin5, which was PCRed from pAc5.1-HisB. To replace the Rpl35a splicing signals, 5' and 3' splice sites from the MHC16 intron (Pfeiffer et al. 2010) were PCRed onto each stuffer region, and these regions were inserted in place of the original ones by EA. After these were shown to splice properly in S2 culture, the Actin5 promoter was replaced by the original tissue-specific promoters. All sequences were analyzed with NNSPLICE 0.9 (available at http://www.fruitfly.org/seq_tools/splice.html) to confirm strength of splice signals and to check for cryptic splice sites.

The full sequences of the I-SceI recognition site, Rpl35a intron 5' and 3' splice signals, and MHC16 intron 5' and 3' signals, along with the first and last 30 base

Name	Sequence
I-SceI Site	ATTACCCTGTTATCCCTA
Rpl35a Intron 5'	CCATCCTCAAG**GT**ATGTCTATACTTCAATGTGA TGGGTCCGGACTTCACAGAGTTTTTCAAATAA TAATTAATA[1]
Rpl35a Intron 3'	CATTCCAGTTCTGTAAAAACATTTTAGTAATGT AATTGATTAACCAACATATTACATTGC**AG**ATTG AGGGC[1]
MHC16 Intron 5'	CACGCCAACAAG**GT**AGGTTCAACCACTGATGC TTAGGAATAGG[1]
MHC16 Intron 3'	CTAGGCACACCGAAACGACTAACCCTAATTCTT ATCCTTTCCTTT**AG** [1]
Name	**30 bps of the 5' and 3' ends of each fragment, 5' to 3'**
Opie2	CACCAACTTTTTTGCACTGCAAAAAAACAC ATCCAACCGCCGCCGCAACCTGTCTCTGGT
svp	AAGCTTGGAAAATGCGTTGGAGTAATAGCC GGAGCACAAACACATGGGCCCTGCCACTTT
Hsp70 basal	AGCGCCGGAGTATAAATAGAGGCGCTTCGT AAGCGCAGCTGAACAAGCTAAACAATCTGC
Actin5	TAAAAAAAATCATGAATGGCATCAACTCTG CATCAGCCAGCAGTCGTCTAATCCAGAGAC
XYZ	CATGTGAAGCTGAAGGAATCTGGCCCTGGG AGGTGGACAAGAAAATTGTGCCCAGGGATT
UVW	GTGGTTGTAAGCCTTGCATATGTACAGTCC GAGAAGAGCCTCTCCCACTCTCCTGGTAAA
GFP	ATGGTGAGCAAGGGCGAGGAGCTGTTCACC ACTCTCGGCATGGACGAGCTGTACAAGTAA
dsRed	ATGGTGCGCTCCTCCAAGAACGTCATCAAG CTACAGGAACAGGTGGTGGCGGCCCTCGGT
SV40 3'UTR	ACATTGATGAGTTTGGACAAACCACAACTA TGTGGTATGGCTGATTATGATCAGTCGACC
CTCF	CCTTGCAGCGCCACCTGGCCGCGAAGAGTT GGCCAGGTGGCGCTGCAAGGTGTGAGTTGT

Table 4.1: Full sequences of short fragments, and first and last 30 bps of longer fragments, used in the translocation constructs.
[1] Splice donors and acceptors are in bold and underlined

pairs (bps) of all longer DNA fragments used in the described constructs, are listed in Table 4.1.

Genomic PCR

Genomic DNA was extracted from whole flies using Qiagen's DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA). To determine the orientation of attP landing sites in each insertion line, a forward or reverse oligonucleotide primer inside the attP site was paired with a forward or reverse primer from the genomic region surrounding the insertion site (for a total of four PCR reactions per site), and PCR products from successful reactions were sequenced to confirm site orientation. As mentioned previously, three sites—51C on the second chromosome and 68E and 9741 on the third chromosome—were chosen for generating transformants. Sequences of the forward and reversed primers used above are listed in Table 4.2.

To verify that isolated translocation-bearing individuals had expected construct architecture, a forward primer from each promoter was tested with reverse primers from both fluorescent reporters to check whether the promoters were now associated with novel reporters (i.e., opie2 with GFP and svp with dsRed), and obtained PCR products were sequenced to confirm the expected loss of one stuffer fragment (XYZ or UVW). A forward primer at the very 3'end of the constructs was used with a reverse primer from the neighboring genomic region to confirm that translocation individuals had insertions at both attP sites, since only one of the translocation-associated markers (opie2-GFP) could be seen. Sequences of the primers used above are listed in Table 4.3.

Name	Sequence 5' to 3'
attP1-F[1]	AGAGTCGTCGACGTCAAAATCACCAC
attP1-R[1]	GCATACATTATACGAAGTTATGAG
attP2-F[2]	AGGTTACCCCAGTTGGGGCACTACTC
attP2-R[2]	TAACCTTTGAGTTCTCTCAGTTGGGGGC
22A-F	AATGGATTCGTGCTCATCTTCTGG
22A-R	AGTGAAGTCAAACTTCTTGTGAGTC
51C-F	CTCGCAAATGCCAGCAGGGTAATG
51C-R	TAGCGAATGAAAACTGCGAAGAAG
68E-F	CAATTACATTTCGATTGATTTTCA
688-R	GCAAACATGACGTATGGAAAATATC
96E-F	GGTGCCGTGTGTCAAATGTGTCGC
96E-R	GATTAACGTGCTGCACGGCTCACG
9741-F	TACTTTTCGTAAACCATATTGAGATAC
9741-R	TAAAATGGACCTGTAGGAATTTACTTAC

Table 4.2: Oligonucleotide primers used to determine the orientation of the attP insertion sites of various candidate attP lines.
[1] These attP primers were used in PCRs with all lines except 9741
[2] These attP primers were used in PCRs with the 9741 line

Name	Sequence 5' to 3'
Opie2-F[1]	CACCAACTTTTTTGCACTGCAAAAAAACAC
GFP-R[1]	ACTCTCGGCATGGACGAGCTGTACAAGTAA
Hsp70-F[2]	TCAAACAAGCAAAGTGAACACATCG
dsRed-R[2]	CTACAGGAACAGGTGGTGGCGGCCCTCGGT
SV40-F[3]	ACATTGATGAGTTTGGACAAACCACAACTA
51C-R[3]	TAGCGAATGAAAACTGCGAAGAAG
68E-F[3]	CAATTACATTTCGATTGATTTTCA
9741-F[3]	TACTTTTCGTAAACCATATTGAGATAC

Table 4.3: Oligonucleotide primers used to confirm the novel association of promoter and reporter, and the presence of alleles in both attP sites, in translocation individuals.
[1] These primers were used to PCR opie2-GFP
[2] These primers were used to PCRs svp-dsRed
[3] These primers were used to confirm presence of constructs in both attP sites; assignment of forward and reverse status is random

Chapter 5
Conclusions

Summary

This work described the development of two types of high-threshold drive mechanisms in Drosophila: engineered underdominance and translocations. Although one mode of engineered underdominance has previously been constructed in Drosophila (Akbari et al. 2013), the systems detailed herein are genetically different from that approach, and may be easier to transfer to vector species due to their simplicity. Additionally, because several methods of engineering underdominance are described—a protein toxin/shRNA rescue method, and two shRNA toxin/two-part rescue approaches—these proof of concept studies should give researchers interested in developing underdominant systems in vector species different options to test in their engineering endeavors.

The use of modern molecular techniques has not, to my knowledge, been applied to creation of specific translocations for use as modulators of gene drive. The method of creating translocations described here should be transferrable to other insect species with only minor modifications, and can serve as a template for engineering viable and fit translocations in insect vectors of disease.

Lessons Learned

I resonate very strongly with the general conclusion that Kelly Matzen, the Hay lab graduate student that worked on underdominance prior to me, reached regarding this project (Matzen 2012)—engineering underdominance in an insect is exceptionally difficult. Despite working with Drosophila, the genetics of which are rather well understood, and despite utilizing well-characterized components (enhancers, shRNAs, etc.), it took our lab over five years to generate underdominant systems capable of genetic drive, and even those are not yet optimized in terms of fitness costs and toxin and antidote potencies. It is also telling that, despite interest

in the subject, no one outside of our research group has engineered a functional underdominant system in Drosophila or in any other species.

This difficulty is, in large part, due to the sheer complexity of interactions within any genetic system, and the impossibility of predicting exactly how any introduced component will behave in any tissue. Even though the CG5266 intein-based system worked in the fly eye, we got very poor to no rescue in essential tissues; in spite of data showing complete lethality when CG5266 and Rpl35a were knocked down in the heart (Neely et al. 2010), we did not get particularly strong killing with the well-characterized heart-specific enhancers (Lo and Frasch 2001; Han and Olson 2005) we tried; and despite targeting sites of Rpl35a that allowed for greater recoding and thus better predicted antidote function, we actually went from having some rescue to no rescue. This demonstrates that it can be impossible to foretell how any given component in such a complex engineering scheme will function, despite what others have reported, or how it behaved in a different tissue, or how theory suggests it will behave. It also underscores how important it is to test each element individually, empirically, and *in vivo*—it may be time-consuming, but I believe it is the only way to really make progress in this kind complex engineering endeavor.

Finally, this work shows what a significant role the specific genomic insertion position plays in the fitness of transgenic animals, and on the ability of an engineered drive system to perform as expected. To spread into a wild type population, underdominant animals must not be significantly unfit as compared with wild type, and since the underdominant alleles themselves often have associated fitness costs, minimizing any negative fitness contribution of insertion sites is essential. This can be done by carefully evaluating the fitness of existing insertion site lines, or, as mentioned in Chapter 4, by using emerging genome editing technologies to create sites in locations deemed optimal in terms of minimizing effects on local gene expression. Both of these approaches are laborious and time-consuming, but ultimately, both should yield fitness-neutral landing sites.

Future Perspectives

Despite the above difficulties, work on underdominant drive mechanisms needs to continue, especially in pertinent vectors of disease, and the research described here can significantly inform any such future efforts.

To create any of the engineered underdominance systems described here in other species, it will be necessary to rediscover every component of the system anew in each target vector, as every organism will likely have unique essential tissue sensitivity, haplolethal genes, shRNA design parameters, etc., and this will certainly involve a considerable amount of effort. Selection of effective tissue-specific enhancers that produce strong killing will be essential. This can be challenging, as my work in Drosophila has shown; however, with the increasing availability of transcriptional profiling data, it should be possible to identify and test multiple suitable enhancers independently to find one that produces strong, tissue-specific lethality. Selection of haplolethal genes for targeting seems most promising, both because a lower level of suppression is required to achieve lethality, and because a mode of rescue that does not rely on protein splicing or recombinase function is more likely to work. All complex animals have haplolethal genes, and methodical analysis of carefully selected haplolethal candidates should reveal some that are sufficiently sensitive to knockdown and capable of rescuing in two recoded copies.

As discussed in Chapter 4, the translocation-based system will likely be the easiest to transfer to disease vectors of interest: it is simple, and does not rely on components that may not be available in other species, such as well-studied tissue-specific enhancers or working shRNAs. The biggest challenge to transferring this technology will be the selection of appropriate breakpoint sites and the generation of fit translocation homozygotes. However, with enough effort and empirical testing, this difficulty will be resolvable. We know this in part because naturally occurring translocations are found in otherwise apparently healthy animals of a number of different species, including humans.

And the goal is certainly worthy of such considerable effort. Despite recent advances in reducing malaria mortality (World Health Organization 2014a), the disease still imposes a tremendous global cost, in terms of both human health and

economic loss (Gallup and Sachs 2001; Crompton et al. 2014; World Health Organization 2014a). The incidence of dengue has increased thirtyfold in the last 50 years, and the disease is rapidly spreading to new areas (e.g., Europe is now at risk for a dengue outbreak) and causing explosive outbreaks in a number of locations (World Health Organization 2014b). Climate change is threatening to increase the range and occurrence of many insect-borne diseases (Githeko et al. 2000; Lemon et al. 2009; Carvalho et al. 2014), numerous previously controlled diseases are resurging (Gubler 1998; Lemon et al. 2009), and insecticide and drug resistance is still a significant concern in many disease endemic areas (Miller and Pierce 2009; Enayati and Hemingway 2010; Ndiath et al. 2012; Wang and Jacobs-Lorena 2013).

The eradication of insect-borne pathogens like malaria will likely require use of insects refractory to disease transmission (Miller and Pierce 2009), and recent releases of genetically modified mosquitoes for SIT programs (Thomas et al. 2000) show that, despite regulatory difficulties and public mistrust, use of transgenic organisms for vector suppression can be carried out (Harris et al 2011; Harris et al 2012; de Jesus and Rego 2013). The initial input of effort and funding required to develop disease-refractory insects capable of replacing wild populations is certainly considerable. However, given the current toll of vector-borne diseases and the lack of other compelling options, it is necessary.

Bibliography

Akbari, O. S., Matzen, K. D., Marshall, J. M., Huang, H., Ward, C. M., & Hay, B. A. (2013). A synthetic gene drive system for local, reversible modification and suppression of insect populations. *Current Biology*, *23*(8), 671–77.

Alphey, L., Beard, C.B., Billingsley, P., et al. (2002). Malaria Control with Genetically Manipulated Insect. *Nature*, *415*, 702.

Altrock, P. M., Traulsen, A., Reeves, R. G., & Reed, F. A. (2010). Using underdominance to bi-stably transform local populations. *J Theor Biol, 267*(1), 62–75.

Altrock, P. M., Traulsen, A., & Reed, F. A. (2011). Stability properties of underdominance in finite subdivided populations. *PLoS Comput Biol, 7*(11), e1002260.

Arndt, K. M., MuÈller, K. M., & PluÈckthun, A. (2001). Helix-stabilized Fv (hsFv) antibody fragments: substituting the constant domains of a Fab fragment for a heterodimeric coiled-coil domain. *Journal of molecular biology*, *312*(1), 221–228.

Asman, S. M., McDonald, P. T., & Prout, T. (1981). Field studies of genetic control systems for mosquitoes. *Annual Review of Entomology*, *26*(1), 289–318.

Baker, R. H. (1984). Chromosome rearrangements in the control of mosquitoes. *Preventive Veterinary Medicine*, *2*(1), 529–540.

Bergmann, A., Agapite, J., McCall, K., & Steller, H. (1998). The Drosophila gene hid is a direct molecular target of Ras-dependent survival signaling. *Cell, 95*(3), 331–341.

Beumer, K. J., Pimpinelli, S., & Golic, K. G. (1998). Induced chromosomal exchange directs the segregation of recombinant chromatids in mitosis of Drosophila. *Genetics*, *150*(1), 173–188.

Billeter, J. C., Atallah, J., Krupp, J. J., Millar, J. G., & Levine, J. D. (2009). Specialized cells tag sexual and species identity in Drosophila melanogaster. *Nature*, *461*(7266), 987–991.

Boete, C., & Koella, J. C. (2002). A theoretical approach to predicting the success of genetic manipulation of malaria mosquitoes in malaria control. *Malar J, 1*, 3.

Boete, C., & Koella, J. C. (2003). Evolutionary ideas about genetically manipulated mosquitoes and malaria control. *Trends Parasitol, 19*(1), 32–38.

Bohannon, J. (2002). Zambia rejects GM corn on scientists' advice. *Science*,

298(5596): 1153–1154.

Borycz, J., Borycz, J. A., Kubow, A., Lloyd, V., & Meinertzhagen, I. A. (2008). Drosophila ABC transporter mutants white, brown and scarlet have altered contents and distribution of biogenic amines in the brain. *Journal of Experimental Biology*, *211*(21), 3454-3466.

Braig, H. R., & Yan, G. (2001). The spread of genetic constructs in natural insect populations. In D. K. Letourneau & B. E. Burrows (Eds.) *Genetically Engineered Organisms: Assessing Environmental and Human Health Effects* (pp. 251–314). Cleveland, OH/Boca Raton, FL: CRC Press.

Carvajal-Vallejos, P., Pallisse, R., Mootz, H. D., & Schmidt, S. R. (2012). Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J Biol Chem, 287*(34), 28686-28696.

Carvalho, D. O., Costa-da-Silva, A. L., Lees, R. S., & Capurro, M. L. (2013). Two step male release strategy using transgenic mosquito lines to control transmission of vector-borne diseases. *Acta Tropica 132S*, S170-S177.

Centers for Disease Control and Prevention (2014). About malaria. Retrieved April 30, 2014, from http://www.cdc.gov/malaria/about/facts.html.

Centers for Disease Control and Prevention (2012). Dengue fact sheet. Retrieved April 30, 2014, from http://www.cdc.gov/Dengue/faqFacts/fact.html.

Chen, C. H., Huang, H., Ward, C. M., Su, J. T., Schaeffer, L. V., Guo, M., & Hay, B. A. (2007). A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. *Science*, *316*(5824), 597-600.

Cheriyan, M., Pedamallu, C. S., Tori, K., & Perler, F. (2013). Faster protein splicing with the Nostoc punctiforme DnaE intein using non-native extein residues. *Journal of Biological Chemistry*, *288*(9), 6202-6211.

Cook, R. K., Christensen, S. J., Deal, J. A., Coburn, R. A., Deal, M. E., Gresens, J. M., . . . Cook, K. R. (2012). The generation of chromosomal deletions to provide extensive coverage and subdivision of the Drosophila melanogaster genome. *Genome Biol, 13*(3), R21.

Corby-Harris, V., Drexler, A., Watkins de Jong, L., et al. (2010). Activation of Akt signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in Anopheles stephensi mosquitoes. *PLoS Pathog, 6*(7), e1001003.

Crompton, P. D., Moebius, J., Portugal, S., Waisberg, M., Hart, G., Garver, L. S., ... & Pierce, S. K. (2014). Malaria immunity in man and mosquito: insights into

unsolved mysteries of a deadly infectious disease. *Annual Review of Immunology*, *32*(1), 157-187.

Curtis, C. F. (1968). Possible use of translocations to fix desirable genes in insect pest populations. *Nature, 218*(5139), 368-369.

Curtis C.F., & Robinson, A.S. (1971). Computer simulation of the use of double translocations for pest control. *Genetics* (69), 97–113.

Dantuma, N. P., Lindsten, K., Glas, R., Jellne, M., & Masucci, M. G. (2000). Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. *Nature biotechnology*, *18*(5), 538-543.

Davis, S., Bax, N., & Grewe, P. (2001). Engineered underdominance allows efficient and economical introgression of traits into pest populations. *J Theor Biol, 212*(1), 83-98.

De, N., Young, L., Lau, P. W., Meisner, N. C., Morrissey, D. V., & MacRae, I. J. (2013). Highly complementary target RNAs promote release of guide RNAs from human Argonaute2. *Molecular cell*, *50*(3), 344-355.

de Jesus, C., & Rego, T.M. (2013). Use of genetic modified mosquitoes to fight dengue in Brazil. *International Journal of Research in Pharmaceutical and Nano Sciences. 2*(6), 811 – 816.

de Lara Capurro, M., Coleman, J., Beerntsen, B. T., et al.. (2000). Virus-expressed, recombinant single-chain antibody blocks sporozoite infection of salivary glands in Plasmodium gallinaceum-infected Aedes aegypti. *Am J Trop Med Hyg, 62*(4), 427-433.

Deredec, A., Burt, A., & Godfray, H. C. J. (2008). The population genetics of using homing endonuclease genes in vector and pest management. *Genetics*, *179*(4), 2013-2026.

Dhar, T., & Mootz, H. D. (2011). Modification of transmembrane and GPI-anchored proteins on living cells by efficient protein trans-splicing using the Npu DnaE intein. *Chem Commun (Camb), 47*(11), 3063-3065.

Egli, D., Hafen, E., & Schaffner, W. (2004). An efficient method to generate chromosomal rearrangements by targeted DNA double-strand breaks in Drosophila melanogaster. *Genome research*, *14*(7), 1382-1393.

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS one*, *3*(11), e3647.

Engler, C., Gruetzner, R., Kandzia, R., & Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS one*, *4*(5), e5553.

Enayati, A., & Hemingway, J. (2010). Malaria management: past, present, and future. *Annual review of entomology*, *55*, 569-591.

Eppstein, M. J., Payne, J. L., & Goodnight, C. J. (2009). Underdominance, multiscale interactions, and self-organizing barriers to gene flow. *Journal of Artificial Evolution and Applications 5*, 1-13.

Fields, S., & Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature, 340*(6230), 245-246.

Filipowicz, W., Jaskiewicz, L., Kolb, F. A., & Pillai, R. S. (2005). Post-transcriptional gene silencing by siRNAs and miRNAs. *Current opinion in structural biology*, *15*(3), 331-341.

Foster, G. G., Whitten, M. J., Prout, T., & Gill, R. (1972). Chromosome rearrangements for the control of insect pests. *Science*, *176*(4037), 875-880.

Franz, A. W., Sanchez-Vargas, I., Adelman, Z. N., Blair, C. D., Beaty, B. J., James, A. A., & Olson, K. E. (2006). Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. *Proc Natl Acad Sci U S A, 103*(11), 4198-4203.

Fu, G., Lees, R. S., Nimmo, D., Aw, D., Jin, L., Gray, P., et al. (2010). Female-specific flightless phenotype for mosquito control. *Proceedings of the National Academy of Sciences*, *107*(10), 4550-4554.

Gallup, J.L., & Sachs, J.D. (2001). The economic burden of malaria. *Am J Trop Med Hyg 64*(1-2 Suppl), 85-96.

Gdula, D.A., Gerasimova, T.I., & Corces, V.G. (1996). Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. *Proc Natl Acad Sci USA 93*(18), 9378-83.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, *6*(5), 343-345.

Githeko, A. K., Lindsay, S. W., Confalonieri, U. E., & Patz, J. A. (2000). Climate change and vector-borne diseases: a regional analysis. *Bulletin of the World Health Organization*, *78*(9), 1136-1147.

Gong, W. J., & Golic, K. G. (2003). Ends-out, or replacement, gene targeting in Drosophila. *Proceedings of the National Academy of Sciences*, *100*(5), 2556-2561.

Gould, F., & Schliekelman, P. (2004). Population genetics of autocidal control and strain replacement. *Annu Rev Entomol, 49*, 193-217.

Gould, F., Magori, K., & Huang, Y. (2006). Genetic strategies for controlling mosquito-borne diseases: engineered genes that block the transmission of malaria and dengue can hitch a ride on selfish DNA and spread into wild populations. *American scientist*, 238-246.

Gould, F., Huang, Y., Legros, M., & Lloyd, A. L. (2008). A Killer–Rescue system for self-limiting gene drive of anti-pathogen constructs. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1653), 2823-2829.

Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage φC31. *Genetics*, *166*(4), 1775-1782.

Gubler, D. J. (1998). Resurgent vector-borne diseases as a global health problem. *Emerging infectious diseases*, *4*(3), 442.

Gubler, D. J., Reiter, P., Ebi, K. L., Yap, W., Nasci, R., & Patz, J. A. (2001). Climate variability and change in the United States: potential impacts on vector-and rodent-borne diseases. *Environmental health perspectives*, *109* (Suppl 2), 223.

Gutierrez, E., Wiggins, D., Fielding, B., & Gould, A. P. (2007). Specialized hepatocyte-like cells regulate Drosophila lipid metabolism. *Nature*, *445*(7125), 275-280.

Han, Z., & Olson, E. N. (2005). Hand is a direct target of Tinman and GATA factors during Drosophila cardiogenesis and hematopoiesis. *Development, 132*(15), 3525-3536.

Harris, A. F., Nimmo, D., McKemey, A. R., et al. (2011). Field performance of engineered male mosquitoes. *Nature biotechnology*, *29*(11), 1034-1037.

Harris, A. F., McKemey, A. R., Nimmo, D., Curtis, Z., Black, I., Morgan, S. A., ... & Alphey, L. (2012). Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nature biotechnology*, *30*(9), 828-830.

Hartl, D.L., & Clark, A.G. (1997). *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates, Inc.

Hay, B. A., Chen, C. H., Ward, C. M., Huang, H., Su, J. T., & Guo, M. (2010). Engineering the genomes of wild insect populations: challenges, and opportunities provided by synthetic Medea selfish genetic elements. *J Insect Physiol, 56*(10), 1402-1413.

Hoffmann, A. A., Montgomery, B. L., Popovici, J., et al. (2011). Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature*, *476*(7361), 454-457.

Ito, J., Ghosh, A., Moreira, L. A., Wimmer, E. A., & Jacobs-Lorena, M. (2002). Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature, 417*(6887), 452-455.

Jacobs-Lorena, M. (2004). Genetic approached for malaria control. In Bogers, R.J. (ed.), *Bridging Laboratory and Field Research for Genetic Control of Disease Vectors* (pp. 52-65). Retrieved from http://library.wur.nl/frontis/

James, A. A. (2005). Gene drive systems in mosquitoes: rules of the road. *Trends Parasitol, 21*(2), 64-67.

Kim, W., Koo, H., Richman, A. M., Seeley, D., Vizioli, J., Klocko, A. D., & O'Brochta, D. A. (2004). Ectopic expression of a cecropin transgene in the human malaria vector mosquito Anopheles gambiae (Diptera: Culicidae): effects on susceptibility to Plasmodium. *Journal of medical entomology*, *41*(3), 447-455.

Knols, B. G., Bossin, H. C., Mukabana, W. R., & Robinson, A. S. (2007). Transgenic mosquitoes and the fight against malaria: managing technology push in a turbulent GMO world. *Am J Trop Med Hyg*. 77(6 Suppl), 232-42.

Krafsur, E. S. (1998). Sterile insect technique for suppressing and eradicating insect populations: 55 years and counting. *J. Agric. Entomol. 15*, 303– 317.

Krstic, D., Boll, W., & Noll, M. (2013). Influence of the White Locus on the Courtship Behavior of Drosophila Males. *PLoS one*, *8*(10), e77904.

Kwit, C., Moon, H. S., Warwick, S. I., & Stewart, C. N., Jr. (2011). Transgene introgression in crop relatives: molecular evidence and mitigation strategies. *Trends Biotechnol, 29*(6), 284-293.

Kyrchanova, O., Chetverina, D., Maksimenko, O., Kullyev, A., & Georgiev, P. (2008). Orientation-dependent interaction between Drosophila insulators is a property of this class of regulatory elements. *Nucleic acids research*, *36*(22), 7019-7028.

Lemon, S. M., Sparling, P. F., Hamburg, M. A., Relman, D. A., Choffnes, E. R., & Mack, A. (2008). *Vector-Borne Diseases: Understanding the Environmental, Human Health, and Ecological Connections, Workshop Summary (Forum on Microbial Threats)*. National Academies Press.

Lin, H., McGrath, J., Wang, P., & Lee, T. (2007). Cellular toxicity induced by SRF-mediated transcriptional squelching. *Toxicological sciences*, *96*(1), 83-91.

Lo, P. C., & Frasch, M. (2001). A role for the COUP-TF-related gene seven-up in the diversification of cardioblast identities in the dorsal vessel of Drosophila. *Mech Dev, 104*(1-2), 49-60.

Lockless, S. W., & Muir, T. W. (2009). Traceless protein splicing utilizing evolved split inteins. *Proc Natl Acad Sci U S A, 106*(27), 10999-11004.

Luan, H., Peabody, N. C., Vinson, C. R., & White, B. H. (2006). Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. *Neuron, 52*(3), 425-436.

Lyon, M. F., Phillips, R. J., & Bailey, H. J. (1972). Mutagenic effects of repeated small radiation doses to mouse spermatogonia I. Specific-locus mutation rates. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *15*(2), 185-190.

Magori, K., & Gould, F. (2006). Genetically engineered underdominance for manipulation of pest populations: a deterministic model. *Genetics, 172*(4), 2613-2620.

Marris, E. (2010). Transgenic fish go large. *Nature, 467*(7313), 259.

Marshall, J. M. (2009). The effect of gene drive on containment of transgenic mosquitoes. *Journal of theoretical biology*, *258*(2), 250-265.

Marshall, J. M. (2010). The Cartagena Protocol and genetically modified mosquitoes. *Nature biotechnology*, *28*(9), 896-897.

Marshall, J. M., & Hay, B. A. (2011). Inverse Medea as a novel gene drive system for local population replacement: a theoretical analysis. *J Hered, 102*(3), 336-341.

Marshall, J. M., & Hay, B. A. (2012). Confinement of gene drive systems to local populations: a comparative analysis. *J Theor Biol, 294*, 153-171.

Marshall, J. M., Touré, M. B., Traore, M. M., Famenini, S., & Taylor, C. E. (2010a). Perspectives of people in Mali toward genetically-modified mosquitoes for malaria control. *Malar J, 9*, 128.

Marshall, J. M., Touré, M. B., Traore, M. M., & Taylor, C. E. (2010b). Towards a quantitative assessment of public attitudes to transgenic mosquitoes: Questions based on a qualitative survey in Mali. *Asia Pacific J. Mol. Biol. Biotechnol*, *18*, 251-273.

Marygold, S. J., Roote, J., Reuter, G., et al. (2007). The ribosomal protein genes and Minute loci of Drosophila melanogaster. *Genome Biol, 8*(10), R216.

Matzen, K.J. (2012). Engineering of Dengue virus refractoriness in Aedes aegypti and development of an underdominant gene drive system (Doctoral dissertation). California Institute of Technology, Pasadena, CA.

McManus, M. T., Petersen, C. P., Haines, B. B., Chen, J., & Sharp, P. A. (2002). Gene silencing using micro-RNA designed hairpins. *RNA*, *8*(6), 842-850.

Miller, T. A. (2011). Let high-tech genetically modified insects counter dengue. *BioScience*, *61*(8), 586-587.

Miller, L. H., & Pierce, S. K. (2009). Perspective on malaria eradication: is eradication possible without modifying the mosquito? *Journal of Infectious Diseases*, *200*(11), 1644-1645.

Moreira, L. A., Ito, J., Ghosh, A., Devenport, M., Zieler, H., Abraham, E. G., . . . Jacobs-Lorena, M. (2002). Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. *J Biol Chem, 277*(43), 40839-40843.

Moreno, E. (2012). Design and construction of "synthetic species". *PLoS One, 7*(7), e39054.

Morrison, N. I., Franz, G., Koukidou, M., et al. (2010). Genetic improvements to the sterile insect technique for agricultural pests. *Asia-Pacific Journal of Molecular Biology and Biotechnology*, *18*(2), 275-295.

Mumford, J. D. (2012). Science, regulation, and precedent for genetically modified insects. *PLoS neglected tropical diseases*, *6*(1), e1504.

Murray, C. J., Rosenfeld, L. C., Lim, S. S., et al. (2012). Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet*, *379*(9814), 413-431.

Nath, R. (2003). Generation and characterisation of plant produced recombinant antibodies specific to LHRH for treatment of sex hormone dependent diseases. (MS thesis). Fachhochschule Aachen, Aachen, Germany.

Ndiath, M. O., Sougoufara, S., Gaye, A., et al. (2012). Resistance to DDT and pyrethroids and increased kdr mutation frequency in An. gambiae after the implementation of permethrin-treated nets in Senegal. *PloS one*, *7*(2), e31943.

Neely, G. G., Kuba, K., Cammarato, A., et al. (2010). A Global In Vivo Drosophila RNAi Screen Identifies NOT3 as a Conserved Regulator of Heart Function. *Cell*, *141*(1), 142-153.

Nern, A., Pfeiffer, B. D., Svoboda, K., & Rubin, G. M. (2011). Multiple new site-specific recombinases for use in manipulating animal genomes. *Proceedings of the National Academy of Sciences*, *108*(34), 14198-14203.

Ni, J. Q., Zhou, R., Czech, B., et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Methods, 8*(5), 405-407.

Pardo, R., Engelhard, M., Hagen, K., et al. (2009). The role of means and goals in technology acceptance. A differentiated landscape of public perceptions of pharming. *EMBO Rep, 10*(10), 1069-1075.

Parvy, J. P., Napal, L., Rubin, T., Poidevin, M., Perrin, L., Wicker-Thomas, C., & Montagne, J. (2012). Drosophila melanogaster Acetyl-CoA-Carboxylase Sustains a Fatty Acid–Dependent Remote Signal to Waterproof the Respiratory System. *PLoS genetics*, *8*(8), e1002925.

Perrimon, N., Ni, J. Q., & Perkins, L. (2010). In vivo RNAi: today and tomorrow. *Cold Spring Harbor perspectives in biology*, *2*(8), a003640.

Pfeiffer, B. D., Ngo, T. T. B., Hibbard, K. L., Murphy, C., Jenett, A., Truman, J. W., & Rubin, G. M. (2010). Refinement of tools for targeted gene expression in Drosophila. *Genetics*, *186*(2), 735-755.

Pfeiffer, B. D., Truman, J. W., & Rubin, G. M. (2012). Using translational enhancers to increase transgene expression in Drosophila. *Proc Natl Acad Sci U S A, 109*(17), 6626-6631.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, *8*(11), 2281-2308.

Riehle, M. M., Xu, J., Lazzaro, B. P., et al. (2008). Anopheles gambiae APL1 is a family of variable LRR proteins required for Rel1-mediated protection from the malaria parasite, Plasmodium berghei. *PLoS One, 3*(11), e3672.

Ringrose, L., Chabanis, S., Angrand, P. O., Woodroofe, C., & Stewart, A. F. (1999). Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *The EMBO Journal*, *18*(23), 6630-6641.

Robinson A.S. (1975). A reassessment of the use of chromosome inversions for insect control. *Journal of Heredity* (66):35-37.

Robinson, A.S. (1976). Progress in the use of chromosomal translocations for the control of insect pests. *Biological Reviews*, *51*(1), 1-24.

Robinson, A. S., Franz, G., & Atkinson, P. W. (2004). Insect transgenesis and its potential role in agriculture and human health. *Insect biochemistry and molecular biology*, *34*(2), 113-120.

Rong, Y. S., & Golic, K. G. (2003). The homologous chromosome is an effective template for the repair of mitotic DNA double-strand breaks in Drosophila. *Genetics*, *165*(4), 1831-1842.

Rørth, P. (1998). Gal4 in the Drosophila female germline. *Mechanisms of development*, *78*(1), 113-118.

Serebrovskii, A.S. (1940). On the possibility of a new method for the control of insect pests. *Zool. Zh.* 19:618–90.

Schmid-Hempel, P. (2005). Evolutionary ecology of insect immune defenses. *Annu Rev Entomol, 50*, 529-551.

Schnütgen, F., & Ghyselinck, N. B. (2007). Adopting the good reFLEXes when generating conditional alterations in the mouse genome. *Transgenic research*, *16*(4), 405-413.

Schwartz, E. C., Saez, L., Young, M. W., & Muir, T. W. (2007). Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat Chem Biol, 3*(1), 50-54.

Sellin, J., Albrecht, S., Kolsch, V., & Paululat, A. (2006). Dynamics of heart differentiation, visualized utilizing heart enhancer elements of the Drosophila melanogaster bHLH transcription factor Hand. *Gene Expr Patterns, 6*(4), 360-375.

Sherizen, D., Jang, J. K., Bhagat, R., Kato, N., & McKim, K. S. (2005). Meiotic recombination in Drosophila females depends on chromosome continuity between genetically defined boundaries. *Genetics*, *169*(2), 767-781.

Sinkins, S. P., & Gould, F. (2006). Gene drive systems for insect disease vectors. *Nat Rev Genet, 7*(6), 427-435.

Spradling, A. C., & Rubin, G. M. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. *Science*, *218*(4570), 341-347.

Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F., & Vignali, D. A. (2004). Correction of multi-gene deficiency in vivo using a single'self-cleaving'2A peptide–based retroviral vector. *Nature biotechnology*, *22*(5), 589-594.

Theilmann, D.A., &Stewart, S. (1992). Molecular analysis of the trans-activating IE-2 gene of Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. *Virology 187*(1), 84-96.

Thomas, D. D., Donnelly, C. A., Wood, R. J., & Alphey, L. S. (2000). Insect population control using a dominant, repressible, lethal genetic system. *Science*, *287*(5462), 2474-2476.

Thorpe, H. M., Wilson, S. E., & Smith, M. (2000). Control of directionality in the site-specific recombination system of the Streptomyces phage φC31. *Molecular microbiology*, *38*(2), 232-241.

Tolle, M. A. (2009). Mosquito-borne diseases. *Current problems in pediatric and adolescent health care*, *39*(4), 97-140.

Uemura, M., Niwa, Y., Kakazu, N., Adachi, N., & Kinoshita, K. (2010). Chromosomal manipulation by site-specific recombinases and fluorescent protein-based vectors. *PLoS one*, *5*(3), e9846.

Van Dyke, D. L., Weiss, L., Roberson, J. R., & Babu, V. R. (1983). The frequency and mutation rate of balanced autosomal rearrangements in man estimated from prenatal genetic studies for advanced maternal age. *American journal of human genetics*, *35*(2), 301.

Walker, T., Johnson, P. H., Moreira, L. A., et al. (2011). The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. *Nature*, *476*(7361), 450-453.

Wang, S., & Jacobs-Lorena, M. (2013). Genetic approaches to interfere with malaria transmission by vector mosquitoes. *Trends in biotechnology*, *31*(3), 185-193.

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS one*, *6*(2), e16765.

Whitten, M. J. (1971). Insect control by genetic manipulation of natural populations. *Science*, *171*(3972), 682-684.

Windbichler, N., Menichelli, M., Papathanos, P. A., et al. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*, *473*(7346), 212-215.

World Health Organization (2000). Global Burden of Disease Study. Retrieved April 30, 2014, from http://www.who.int/evidence/bod.

World Health Organization (2014a). Malaria factsheet. Retrieved April 30, 2014, from http://www.who.int/mediacentre/factsheets/fs094/en/index.html

World Health Organization (2014b). Dengue factsheet. Retrieved April 30, 2014, from http://www.who.int/mediacentre/factsheets/fs117/en/index.html

Yu, Y., & Bradley, A. (2001). Engineering chromosomal rearrangements in mice. *Nature Reviews Genetics*, *2*(10), 780-790.

Zettler, J., Schutz, V., & Mootz, H. D. (2009). The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett, 583*(5), 909-914.

Zhou, X., Vink, M., Klaver, B., Berkhout, B., & Das, A. T. (2006). Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene therapy*, *13*(19), 1382-1390.

Zhu, X. D., & Sadowski, P. D. (1995). Cleavage-dependent ligation by the FLP recombinase; characterization of a mutant flp protein with an alteration in a catalytic amino acid. *Journal of Biological Chemistry*, *270*(39), 23044-23054.

Appendix A

Cloning Methods

A.1 Enzymatic Assembly Cloning Method

Without tremendous advances in cloning technologies over the last decade, engineering underdominance would have been significantly more difficult, if not impossible, in the span of a graduate career. Prior to 2009, the traditional cloning techniques used in the Hay lab could only assemble one or two DNA fragments at a time, and depended on unique restriction sites for each step. This was quite limiting, as our underdominance constructs were often very complex and required many parts to be frequently moved around. In 2009, Gibson et al. described a cloning method that allowed for the cloning of numerous fragments at a time in a directionally specified way, required only one restriction site for each step, and could be used to generate novel restriction sites and short sequences between fragments. This enzymatic assembly (or one-step, as it is frequently called) technology was used for the vast majority of cloning steps described in this work.

The mix for this method was made in-house, following the recipe from Gibson et al.:

5X Isothermal Reaction Buffer:

3 mL 1M Tris-HCl pH 7.5
150 μL 2M $MgCl_2$
60 μL 100 mM dGTP
60 μL 100 mM dATP
60 μL 100 mM dTTP
60 μL 100 mM dCTP
300 μL 1M DTT
1.5 g PEG-8000
300 μL 100 mM NAD
~ 4 mL total

→ Aliquot 330 μL into 0.6 mL tubes and store at -20°C

Enzymatic Assembly (EA) Mix:

320 μL 5x Iso Buffer (recipe above)
0.64 μL (10 U/uL) T5 Exonuclease (Epicentre)
20 μL (2,000 U/mL) Phusion DNA Polymerase (New England Biolabs)
160 μL (40,000 U/mL) Taq Ligase (New England Biolabs)
700 μL H_20
~ 1200 μL Total

→ Aliquot 15 μL into PCR tubes and store at -20°C (stable for 1-2 months)

To set up the reaction, thaw the 15 μL EA mix on ice, add 5 uL DNA total mix, incubate at 50°C for one hour, and transform 2-3 μL into competent cells of choice (typically JM109, Zymo Research, Irvine, CA). The backbone and each fragment are combined in a 1:1 molar ratio.

For the reaction to work, each fragment has to have at least 20 bps of overlap to the backbone and/or other fragments. To optimize reaction efficiencies, I typically used overlaps of at least 30 bps between each fragment and the backbone, and overlaps of 35 bps or more between neighboring fragments.

A.2 Golden Gate Cloning Method

Although the EA cloning method works well for most cloning applications, it is not particularly useful for putting together highly repetitive sequences or copies of the same fragment, as it depends on the annealing of complementary regions for proper directional assembly. A significant portion of the cloning described in this work required the assembly of shRNA sets composed of identical shRNA backbones with different guide and passenger strands (i.e., highly repetitive sequences). Since EA could not be used for this, it was typically done via traditional ligation, which only allowed for two shRNAs at most to be cloned at the same time. Thus, cloning different sets of shRNAs was a significant bottleneck, and limited the sets that could be tested.

However, another cloning technology that was adopted by the Hay lab around 2011, the Golden Gate method (Engler et al. 2008; Weber et al. 2011), greatly reduced the time involved in assembling shRNA sets. This method utilizes Type IIs

restriction enzymes that cleave outside of their recognition site, leaving 5' or 3' DNA overhangs of any desired sequence (Engler et al. 2008). Since this allows for the creation of many unique overhangs (e.g., 256 for 4 bp overhands), numerous DNA fragments can be cloned into a vector in a directionally controlled way in only one step, even if they are repeats of the same sequence. And since only one restriction enzyme is utilized, the restriction and ligation steps can be done together, and there are no buffer incompatibility issues.

Katie Kennedy, the lab coordinator at the time, adapted the Golden Gate method based on the BsaI type IIs enzyme described in Engler et al. 2008 and Weber et al. 2011 for use in the Hay lab, and her protocol is reproduced below. I used this method to build all sets of shRNAs detailed in this work, as well as any constructs involving highly repetitive sequences. Despite reports that claimed assembly of up to 10 DNA pieces with 95-100% efficiency in one step (Engler et al. 2009), in our hands, the Golden Gate method could only be reliably used to assemble six to eight fragments, and the efficiency was lower than that of the EA reaction. Thus, this technology was typically only used to clone shRNAs and other highly repetitive fragments.

Golden Gate Procedure:

1. Set up the following reaction in one tube:
   - 100 ng backbone
   - 100 ng each insert
   - 2 μL T4 DNA ligase buffer (10X; New England Biolabs)
   - 1 μL T4 DNA ligase (New England Biolabs)
   - 1 μL BsaI (New England Biolabs)
   - + H_20 up to 20 μL (total reaction volume)
2. Set up the following digestion/ligation reaction:
   a. (Digestion) 5:00 minutes at 37°C
   b. (Ligation) 10:00 minutes at 16°C
   c. Repeat steps "a" and "b" for 10 cycles total

d. (Final digestion) 5:00 minutes at 37°C
e. (Kill enzyme) 5:00 minutes at 80°C
f. 4°C hold or straight to ice

3. Transform 5 μL directly into competent cells (typically JM109, Zymo Research, Irvine, CA).

The backbone I used was OA669 (provided by O. Akbari), which is designed for blue/white screening via X-Gal. After the desired fragments were assembled into this backbone, they were moved to the final backbone via EA or tradition subcloning.

A.3 Small Hairpin RNA Design and Assembly

When I first started performing research in the Hay lab, shRNA assembly and multimerization involved consecutive traditional subcloning steps that could assemble shRNAs one or two at a time (see Matzen 2012 for a detailed protocol). The shRNAs described in this work that were provided by Kelly Matzen were constructed in this fashion.

However, with the lab's adoption of the Golden Gate method, the earlier technique was pretty much abandoned, and all of the shRNA assembly that I performed was done via a one-step Golden Gate reaction. The shRNA designs I used were based on either the endogenous Drosophila miR-6.1 (Chen et al. 2007) or miR-1 (Ni et al. 2011) backbone. The sequence of each backbone is provided below, with the guide (in red) and passenger (in blue) strands indicated in N's. Unless otherwise noted, the guide strands were always 100% complementary to the target mRNA; and in the case of miR-1, the guide and passenger strands were also completely complementary. However, for the miR-6.1 designs, the guide and passenger strands were mismatched at position three of the guide, as that had been speculated to render the shRNAs more efficient in the lab.

The shRNAs were PCRed in three steps, using three distinct sets of primers. The first set was self-annealing, with a 26 bp region of complementarity, and unique to

each target mRNA site. The first round of PCR created the passenger and guide fragment of the stem loop. The second primer set was universal to all shRNAs, and added the 5' and 3' flank arms to each basic stem loop. The third set of primers was used to put in the BsaI restriction sites and 4 bp unique overhangs that would make Golden Gate assembly of the shRNAs possible. Additionally, if the particular shRNA was to be on the 5' or 3' terminus of the set, the third round of PCR also added a splice site on the end of that stem loop. In this way, each set of shRNAs (whether a three-mer, four-mer, or six-mer) would be contained within an intron that could splice out of the 3'UTR of the open reading frame (ORF) it was contained in.

The general protocol was as follows:

1. Perform first round of PCR with self-annealing primers
2. Perform second round of PCR on the product from the first round (diluted at least 1:10)
3. Perform third round of PCR on the product from the second round (diluted at least 1:10)
4. Gel purify fragments
5. Set up Golden Gate reaction with fragments and appropriate backbone

The Golden gate reaction was most efficient for putting in four shRNAs at once, and could be used to put in six shRNAs with decent efficiency. However, attempting to insert more fragments than that did not work very well in my hands.

The particular region (5'UTR, 3'UTR, CDs) of a given mRNA targeted by the shRNAs differed depending on a number of factors, which are discussed on a case by case basis in the main text. However, no matter what general area was targeted, specific sites were selected based on either rules described by Chun-Hong Chen, use of the Thermo Scientific siRNA Design Center (http://www.thermoscientificbio.com /design-center/), or some combination of both.

miR-6.1 Backbone

GATCTTTTAAAGTCCACAACTCATCAAGGAAAATGAAAGTCAAAGTTGGCAGCTTACTT
AAACTTAATCACAGCCTTTAATGTNNNNNNNNNNNNNNNNNNNNNTAAGTTAATA
TACCATATCTANNNNNNNNNNNNNNNNNNNNNGTACCTAAAGTGCCTAACATCATT
ATTTAATTTTTTTTTTTTTTGGCACACGAATAACCATGCCGTTTTG

miR-1 Backbone

AAGTGAGTAGTGCCACCAAAAGTTAGCCGCGTTGTGGAAAATCCAAAACAAAGACCATC
CCATATTCAGCCTTTGAGAGTTNNNNNNNNNNNNNNNNNNNNNGTTATATTCAAGC
ATANNNNNNNNNNNNNNNNNNNNNCGAAATCTGGCGAGACATCGGAGTTGAAACT
AAAACTGAAATTTGATTGAAACAGAAGTAGAACCGTAATGAAATGAAT

Consensus Splice Sites

The splice donor and acceptor are capitalized.

5’ site cagGTaagt

3’ site tactaattcttcttttccctttttttAGg

Appendix B

UD with shRNA toxins

B.1 CG5266 Sequence Data

Intein Sequences

Intein N Fragment (attached to the 3' end of the N-terminal portion of CG5266)
TGCTTGTCCTATGAAACGGAAATCTTGACGGTTGAATATGGCCTGCTCCCAATCGGCAAG
ATCGTGGAAAAACGTATCGAATGCACGGTCTATTCGGTGGATAACAATGGTAACATCTA
CACCCAACCGGTGGCCCAGTGGCACGACCGCGGAGAACAGGAGGTTTTTGAATATTGCCT
GGAAGATGGTTCCCTCATTCGCGCCACTAAGGATCACAAGTTCATGACCGTGGACGGACA
AATGCTGCCGATCGACGAGATATTTGAACGGGAGCTGGATCTGATGCGCGTGGACAACCT
GCCGAATTAG

Intein C Fragment (attached to the 5' end of the C-terminal portion of CG5266)
ATGATCAAGATTGCCACCAGGAAGTATTTGGGAAAGCAGAATGTCTACGATATTGGCGT
GGAACGAGATCATAATTTCGCGCTGAAAAACGGATTCATAGCCTCGAACTGTTTTAAC

shRNA Target Sites

Table B.1 contains the 22 bp mRNA sequences of CG5266 targeted with the various sets of shRNAs utilized in this work.

shRNA Set	mRNA Target Sequence (5' to 3')
Original UTR-targeting set[1]	TCATTTTACACATCAAATCACT CCAGGAAATATGGCTACCGAAC CCACAAGCTAAGCTTTCTTAAT
CD-targeting set A[1]	AAACACAAGTCACCGCTGTATG CCATCCTTACGCTGAAAGAAGG GAATCTGCGATCAGAACGGATT
CD-targeting set B[2]	CATGGGCAAGAACGCAGTGAAC CACCGCCATCCTTACGCTGAAA TGAAAGAAGGTTTTGAGGGAAA

5'UTR-targeting set	GCATTCAGCATTCGGTCACACT CAATTGCTGTGTCATTTTTGTT GTGCGAAATAAGTCGGCTTTTT TTACACATCAAATCACTGCATT
3'UTR-targeting set	TAAGATTACCACAAGCTAAGCT TTACCACAAGCTAAGCTTTCTT ATAAGTACGGGACATTAAAAAA AAAATTTTACAACTGGTGTTAT

Table B.1: Messenger RNA (mRNA) sites of CG5266 targeted by various sets on shRNAs.
[1] Target sites selected by K. Matzen.
[2] Target sites selected by G. Pittman.

CG5266 Recoding

The CG5266 rescue was recoded to be resistant to each set of shRNAs that targeted the CDs (no recoding was necessary for the UTR-targeting shRNAs, since the transcript's endogenous UTRs were not utilized as part of the rescue). In each sequence, exons are in uppercase, while introns are in lowercase. Sites that have been recoded are bolded.

CG5266 recoded for original UTR-targeting shRNAs

ATGGCAACTGAGCGATACAGCTTTTCGTTGACCACGTTCAGgtgagagcatcgctatagccggag
cggagcattcgcttcgccccgcgcaacacaaactgccgccatgactgggcaataagtacaattttctaccgtttgcagTC
CTTCCGGAAAACTCGTCCAACTGGAGTATGCATTGGCGGCCGTATCTGGCGGAGCTCCCT
CCGTGGGCATTATAGgttagcatgccgcacagccagtcagcctgtcagccagtactgtgcattcaactaacacact
attatttatttttcttggctgtagCTTCCAACGGCGTCGTCATTGCCACAGAGAACAAACACAAGTCA
CCGCTGTATGAGCAGCACAGTGTACATCGCGTGGAGATGATCTACAACCACATCGGCATG
GTGTACTCGGGAATGGGTCCGGACTACCGCCTGCTGGTCAAGCAGGCCCGCAAGATCGCC
CAGACGTACTACCTGACCTACAAGGAGCCGATTCCAGTGTCACAGCTGGTGCAGCGCGTG
GCCACGCTCATGCAGGAGTACACTCAGTCCGGgtaagtgtgcgcgcttttctcgactttgggttcccatag
ataatcgtatgcaatcccaccaacagTGGCGTTCGTCCCTTTGGCGTTTCCCTACTGATCTGCGGCT
GGGACAATGATCGTCCGTATCTCTACCAATCCGATCCTTCGGGCGCCTACTTCGCCTGGA
AGGCCACTGCCATGGGCAAGAACGCAGTGAACGGCAAAACTTTCCTGGAGAAGCGgtaagt
accaataatagtacaccaattaaagtcgcaagtcactcatattttcatcaacctgcagCTACAGCGAAGATCTGG

AGCTGGACGACGCTGTTCACACCGCCATCCTTACGCTGAAAGAAGGTTTTGAGGGAAAAA
TGACTGCCGACAACATTGAGATCGGAATCTGCGATCAGAACGGATTCCAGCGTCTGGACC
CCGCCTCAATCAAGGACTACTTGGCCAGCATCCCCTAA

CG5266 recoded for both CD-targeting shRNA sets

ATGGCTACCGAACGATACAGCTTTTCGTTGACCACGTTCAGgtgagagcatcgctatagccggagc
ggagcattcgcttcgccccgcgcaacacaaactgccgccatgactgggcaataagtacaattttctaccgtttgcagTCC
TTCCGGAAAACTCGTCCAACTGGAGTATGCATTGGCGGCCGTATCTGGCGGAGCTCCCTC
CGTGGGCATTATAGgttagcatgccgcacagccagtcagcctgtcagccagtactgtgcattcaactaacacacta
ttatttatttttcttggctgtagCTTCCAACGGCGTCGTCATTGCCACAGAGAAC**AAGCATAAATCC**
CCATTATACGAGCAGCACAGTGTACATCGCGTGGAGATGATCTACAACCACATCGGCAT
GGTGTACTCGGGAATGGGTCCGGACTACCGCCTGCTGGTCAAGCAGGCCCGCAAGATCGC
CCAGACGTACTACCTGACCTACAAGGAGCCGATTCCAGTGTCACAGCTGGTGCAGCGCGT
GGCCACGCTCATGCAGGAGTACACTCAGTCCGGgtaagtgtgcgcgcttttctcgactttgggttcccata
gataatcgtatgcaatcccaccaacagTGGCGTTCGTCCCTTTGGCGTTTCCCTACTGATCTGCGGC
TGGGACAATGATCGTCCGTATCTCTACCAATCCGATCCTTCGGGCGCCTACTTCGCCTGG
AAGGCCACTGC**TATGGGAAAAAATGCCGTCAAT**GGCAAAACTTTCCTGGAGAAGCGgta
agtaccaataatagtacaccaattaaagtcgcaagtcactcatattttcatcaacctgcagCTACAGCGAAGATCT
GGAGCTGGACGACGCTGTTCA**TACTGCTATTCTCACCCTCAAGGAGGGATTCGAAGGT**
AAAATGACTGCCGACAACATTGAGATC**GGTATTTGTGACCACAATGGTT**TCCAGCGTC
TGGACCCCGCCTCAATCAAGGACTACTTGGCCAGCATCCCCTAA

FLEX Switch F3/FRT Site Arrangement

F3 and FRT recombinase sites were arranged around the first exon of CG5266 as described in Schnutgen and Ghyselinck 2007. In the sequence of this arrangement shown below, F3 sites are underlined, FRT sites are italicized, and CG5266 is in red (with exons in uppercase and introns in lowercase).

GAAGTTCCTATTCCGAAGTTCCTATTCTTCAAATAGTATAGGAACTTCAGCTGAACAAGC
TAAACAATCTGCA*GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC*g
gctgacaggctgactggctgtgcggcatgctaacCTATAATGCCCACGGAGGGAGCTCCGCCAGATAC

GGCCGCCAATGCATACTCCAGTTGGACGAGTTTTCCGGAAGGActgcaaacggtagaaaattgta
cttattgcccagtcatggcggcagtttgtgttgcgcggggcgaagcgaatgctccgctccggctatagcgatgctctcacC
TGAACGTGGTCAACGAAAAGCTGTATCGTTCGGTAGCCATtgttgGAAGTTCCTATACTAT
TTGAAGAATAGGAACTTCGGAATAGGAACTTCTTCCTGGTGGCAATCTTGATCATtttgttg
*GAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGAACTTC*agtactgtgcattcaacta
acacactattatttatttttcttggctgtag

B.2 Rpl35a Sequence Data

shRNA Target Sites

Table B.2 contains the 22 bp mRNA sequences of Rpl35a targeted with the various sets of shRNAs utilized in this work.

shRNA Set	mRNA Target Sequence (5' to 3')
Original UTR-targeting set[1]	ATACATTCATCGAACAACAAGC AACAACAAGCTAAGCCATGGCC TAAGGCGAGATACCGATTGAAA
CD-targeting set	ACACAAGCCAAGTCCACTACTG CCAAGAAGTACAAGCGTCACGG TGAGGAACCAGCACGAGAACCA ACAAGGCCGAGACCAAGAAGTG AGAGGCTAAAGTTTCCGCCAAG GGAGCACGGATCCTTCTACGTT
Mixed set A[2]	GTACAAGCGTCACGGACGCCTC CGGCTACAAGCGTGGTCTGAGG CAAGATTGAGGGCGCCCGCCGC TAAGGCGAGATACCGATTGAAA
Mixed set B[2]	CGAGCGCAAGACCCGCGTCCGC TGCCCGTTTCAACAGGAACCTG CATCATGCTGTACCCATCAAGG TAAGGCGAGATACCGATTGAAA
3'UTR-targeting set	ATACCGATTGAAAGTTGACGGT TTTAATTAAAACGCAACTTGGC TTTTTTATTAAGGCGAGATACC AATTACTGACCTGCAGGAGTAA

Table B.2: Messenger RNA (mRNA) sites of Rpl35a targeted by various sets of shRNAs.

[1] Target sites selected by K. Matzen.
[2] These sets each contain three novel sites in the CDs (chosen because they would allow for more extensive recoding of the rescue transcript) plus the original 3'UTR-targeting site selected by K. Matzen.

Rpl35a Recoding

The Rpl35a rescue was recoded to be resistant to each set of shRNAs that targeted the CDs and the UTRs, since both were included in the rescue fragment. In each sequence, exons are in uppercase, while introns are in lowercase. Sites that have been recoded are bolded. The 5'UTR and 3'UTR regions are underlined. For the 5'UTR, intronic UTR regions are replaced with "…" for sake of space. Note that not the entire genomic region used in the rescue fragment is shown; details on the additional utilized sequences upstream of the 5'UTR and downstream of the 3'UTR can be found in Chapter 3.

Rpl35a recoded for original UTR-targeting shRNAs

TCCTTCTTTTCGCTTTCGTTTCCGGCGAA…CGGTGTTA**ACACTTTTATTGAGCAAAAAC**
…GTGAAAC**ATGGCT**GACACACAAGCCAAGTCCACTACTGCGCCCAAGGCCGCCAAGGCC
CAGAAGGCTCCCAAGGCCGTCAAGGCGCCTAAGGCCGAGAAGCCCGCCGCCTCAGAGGCT
AAAGTTTCCGCCAAGAAGTACAAGCGTCACGGACGCCTCTTCGCCAAGGCCGTCTTCACC
GGCTACAAGCGTGGTCTGAGGAACCAGCACGAGAACCAGGCCATCCTCAAGgtatgtctatact
tcaatgtgatgggtccggacttcacagagtttttcaaataataattaatacattccagttctgtaaaaacattttagtaatgt
aattgattaaccaacatattacattgcagATTGAGGGCGCCCGCCGCAAGGAGCACGGATCCTTCTAC
GTTGGGAAGCGTTGCGTCTATGTCTACAAGGCCGAGACCAAGAAGTGCGTGCCGCAGCAT
CCCGAGCGCAAGACCCGCGTCCGCGCTGTCTGGGGCAAGGTCACCCGCATCCACGGCAAC
ACCGGCGCTGTGCGTGCCCGTTTCAACAGGAACCTGCCCGGTCATGCCATGGGCCACCGC
ATCCGCATCgttagttcaaacacacctttctgtccgacaattgggttgtattaacatgccttctttttttcttgcagATG
CTGTACCCATCAAGGATTTAAGTTAATATCCGACTTGAATTACTGACCTGCAGGAGTAA
AAAATCCGTTTTACATTAAATGAAACACTTTAAATTTAATTAAAACGCAACTTGGCTTT
TTTATT**GAGACGTGACACGGACTAG**AAGTTGACGGTAATCTGTA

Rpl35a recoded for the CD-targeting shRNA set and the original UTR-targeting set

TCCTTCTTTTCGCTTTCGTTTCCGGCGAA…CGGTGTTA**ACACTTTTATTGAGCAAAAAC**
…GTGAAAC**ATGGCT**GAC**ACCCAGGCGAAATCGACCACCG**CGCCCAAGGCCGCCAAGGC
CCAGAAGGCTCCCAAGGCCGTCAAGGCGCCTAAGGCCGAGAAGCCCGCCGCCTC**CGAAGC**
CAAGGTGTCGGCTAAAAAATATAAACGCCATGGCCGCCTCTTCGCCAAGGCCGTCTTC
ACCGGCTACAAGCGTGGTC**TCAGAAATCAACATGAAAATCA**AGCCATCCTCAAGgtatgtc
tatacttcaatgtgatgggtccggacttcacagagtttttcaaataataattaatacattccagttctgtaaaaacattttagt
aatgtaattgattaaccaacatattacattgcagATTGAGGGCGCCCGCCGCAA**AGAACATGGCTCGT**
TTTATGTGGGGAAGCGTTGCGTCTATGTCT**ATAAAGCTGAAACGAAAAAATG**TGTGCC
GCAGCATCCCGAGCGCAAGACCCGCGTCCGCGCTGTCTGGGGCAAGGTCACCCGCATCCA
CGGCAACACCGGCGCTGTGCGTGCCCGTTTCAACAGGAACCTGCCCGGTCATGCCATGGG
CCACCGCATCCGCATCgttagttcaaacacacctttctgtccgacaattgggttgtattaacatgccttctttttcttt
gcagATGCTGTACCCATCAAGGATTTAAGTTAATATCCGACTTGAATTACTGACCTGCAG
GAGTAAAAAATCCGTTTTACATTAAATGAAACACTTTAAATTTAATTAAAACGCAACTT
GGCTTTTTTATT**GAGACGTGACACGGACTAG**AAGTTGACGGTAATCTGTA

Rpl35a recoded for the mixed-target shRNA sets

TCCTTCTTTTCGCTTTCGTTTCCGGCGAA…CGGTGTTAATACATTCATCGAACAACAAG
…CTAAGCCATGGCCGACACACAAGCCAAGTCCACTACTGCGCCCAAGGCCGCCAAGGCCC
AGAAGGCTCCCAAGGCCGTCAAGGCGCCTAAGGCCGAGAAGCCCGCCGCCTCAGAGGCTA
AAGTTTCCGCCAAGAAA**ATATAAAAGGCATGGCAGGTTG**TTCGCCAAGGCCGTCTTCAC
GGGATATAAAAGGGGCTTACGCAACCAGCACGAGAACCAGGCCATCCTGAAgtatgtctata
cttcaatgtgatgggtccggacttcacagagtttttcaaataataattaatacattccagttctgtaaaaacattttagtaatg
taattgattaaccaacatattacattgcag**AATCGAAGGAGCTAGAAGA**AAGGAGCACGGATCCTTC
TACGTTGGGAAGCGTTGCGTCTATGTCTACAAGGCCGAGACCAAGAAGTGCGTGCCaCAG
CATCC**gGAAAGAAAAACGAGGGTGAGG**GCTGTCTGGGGCAAGGTCACCCGCATCCACGG
CAACACCGGCGCTGTGCG**CGCTAGATTTAATCGCAATTTA**CCCGGTCATGCCATGGGCC
ACCGCATCCGTATTgttagttcaaacacacctttctgtccgacaattgggttgtattaacatgccttctttttcttgc
ag**ATGTTATATCCGAGCCGA**ATTTAAGTTAATATCCGACTTGAATTACTGACCTGCAGG

AGTAAAAAATCCGTTTTACATTAAATGAAACACTTTAAATTTAATTAAAACGCAACTTG

GCTTTTTTATGATAATAGAACTTGTTGGTCTAGTTGACGGTAATCTGTA

Rpl35a 3'UTR recoded for the regular 3'UTR-targeting shRNA set

GTTAATATCCGACTTG**AGTAAATGATCTATGACCAGTG**AAAATCCGTTTTACATTAAA
TGAAACACTTTAAA**TTGAAGATCACATACTAATCTTGATTATGTGATAATAGAACTG
TTTCTTCAAAGTTCTGTC**AATCTGTA

Rpl35a 3'UTR recoded for the 3'UTR-targeting shRNA set bearing mismatches to the targeted mRNA

GTTAATATCCGACTTG**AATTAATGATCTATGACCAGTG**AAAATCCGTTTTACATTAAA
TGAAACACTTTAAA**TTTAAGATCACATACTAATCTTTTTTATGTGATAATAGAACTG
TTTCTTCAAAGTTCTGTC**AATCTGTA

Appendix C

Fly Methods

C.1 FLP Excision

Removal of FRT- or B3-flanked blocker sequences utilized in the engineered underdominance (UD) constructs depended on crossing in germline (GL) expressing FLP or B3 recombinase into unflipped UD individuals. The generation of germline FLP and B3 recombinase lines and Bloomington Stock Center lines used for this are described in Chapter 2. A schematic of how the crosses were set up (using the FRT/FLP system as an example) is shown in Figure C.1.

For kill crosses (Figure C.1A), individuals bearing unflipped UD alleles were first crossed to homozygous GL FLP flies of the opposite sex, generating individuals with one copy of the underdominant allele and one copy of male or female specific GL FLP. These individuals were then crossed out to white minus to score killing. As described in Chapter 2, blocker sequences contained GFP, so the presence of unflipped UD alleles was verified by observable tissue-specific GFP expression. UD flies also contained a different marker not dependent on blocker presence—usually white plus, and often eye-specific dsRed expression present in the attP insertion lines. Thus, upon outcross to white minus, the above heterozygous UD/FLP individuals were expected to produce only white minus progeny if killing was complete, and 50% white minus/50% white plus individuals if there was no killing or if the blocker was not removed (in the former case, the white plus progeny would have no GFP expression; in the latter, they would have tissue-specific GFP). Alternatively, males with unflipped UD alleles were crossed to virgins from the BicC-FLP line (see Chapter 2). These females expressed enough FLP to result in 100% efficiency of blocker excision in all of their progeny, even if their mates contributed the UD allele. Thus, this was a faster way of scoring killing, since only one cross was required for blocker removal.

To test the ability of complementary UD alleles to bring about rescue, UD/FLP individuals bearing opposite UD alleles were crossed to each other (Figure C.1B). If killing was complete and the rescue worked, one would expect 25% of progeny from such a cross to be UD transheterozygotes (UD1/UD2). If killing was incomplete, transheterozygotes were identified by intensity of white plus expression (i.e., redness of eyes) and verified by outcrossing.

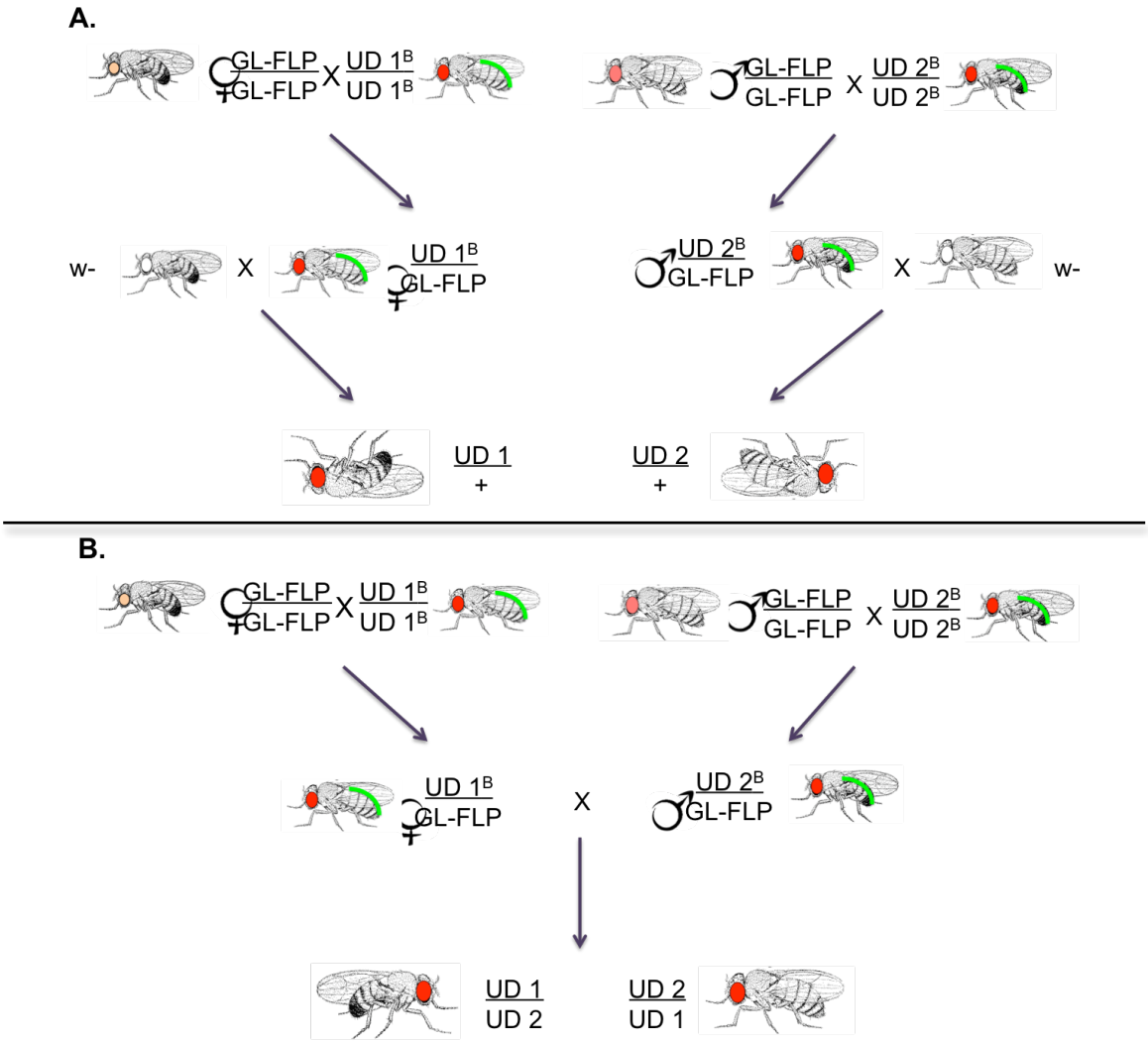

Figure C.1: A. The crossing scheme used to remove blocker sequences from UD allele-bearing flies, and to test for killing by UD alleles. The green stripe on the dorsal fly side represents GFP expressed when the blocking sequence is present. The ^{B} symbol by UD alleles signifies that the blocker is present B. The crossing scheme used to test whether complementary UD alleles rescued.

C.2 Translocation Crosses

To generate translocation individuals, both of the translocation constructs and a source of I-SceI had to be present in the same fly. A stock of such individuals was created as diagrammed in Figure C.2. Individuals from a line expressing I-SceI from heat shock (hs) promoter (marked by scutoid) on the second chromosome, obtained from Bloomington Stock Center line #6956, were crossed to flies homozygous for the opie2-dsRed translocation construct on the third chromosome to generate animals with second-chromosome hs-I-SceI and a third-chromosome translocation construct. These individuals were then crossed to flies homozygous for the svp-GFP translocation construct on the second chromosome to generate a line with both of the translocation alleles and the hs-I-SceI. (Although, as discussed in Chapter 4, svp-driven GFP was not visible in flies, individuals bearing the svp-GFP translocation construct were distinguishable by their distinct orange eye color).

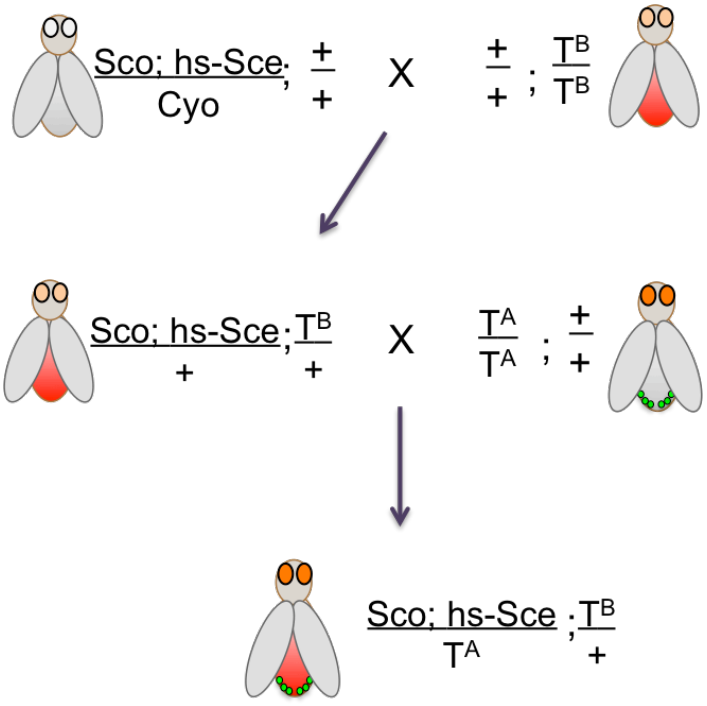

Figure C.2: The crossing scheme used to generate individuals with both of the translocation constructs and a source of heat shock (hs) induced I-SceI. The I-SceI was marked with scutoid, and was crossed in from BSC stock #6935.

C.3 Drive Experiments

Unless otherwise noted, all drive experiments were set up against the empty attP insertion site lines used to generate the specific transformant lines. Fly husbandry was performed under standard conditions at 25°C. For each drive experiment, fly bottles were set up so that all flies to be used would be collected at the same time under the same conditions. Newly emerged males and virgin females were collected and aged for approximately four days. Then an equal number of UD or translocation males and females were placed in a bottle with an equal number of males and females of the corresponding attP line to make up drive experiments with various thresholds (i.e., for an 80% release experiment, 40 transgenic males and 40 transgenic females were mixed with 10 attP males and 10 attP females, so 80% of the starting drive mixture was transgenic). Each bottle was set with 80-100 flies, and every different threshold was set at least in duplicate, but usually in triplicate or greater.

After around five days of mating, the parental flies from each bottle were discarded, leaving behind the developing larvae. Several days after most of the progeny in each bottle eclosed (~ 14 days after the experiment was originally set), half of the offspring were transferred to a new bottle to seed the next drive generation, and half were counted (this was done to avoid overexposure of flies to CO_2 while counting, and because there were usually hundreds of flies that emerged from each bottle, making scoring all of the progeny laborious). Counts were continued for two generations after phenotypic fixation was observed for the reproductively isolated UD lines, and for at least four generations for the translocation lines and the UD lines that had some heterozygote survival.

The initial experiments were carried out in a fly incubator that had a 12-hour day/night light cycle. However, to avoid any potential fitness advantages resulting from the red-colored eyes of the transgenics (versus the white eyes of the empty attP line flies), later experiments were set in a dark fly incubator.