PROMOTERS, ENHANCERS AND INSULATORS FOR IMPROVED MOSQUITO TRANSGENESIS

A Dissertation

by

CHRISTINE ELIZABETH GRAY

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

DOCTOR OF PHILOSOPHY

August 2005

Major Subject: Genetics

PROMOTERS, ENHANCERS AND INSULATORS FOR IMPROVED

MOSQUITO TRANSGENESIS

A Dissertation

by

CHRISTINE ELIZABETH GRAY

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

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Craig J. Coates
Committee Members,	Ellen W. Collisson
	Timothy C. Hall
	Dorothy E. Shippen
Chair of Genetics Faculty,	Geoffrey Kapler

August 2005

Major Subject: Genetics

ABSTRACT

Promoters, Enhancers and Insulators for Improved

Mosquito Transgenesis. (August 2005)

Christine Elizabeth Gray, B.S., DePaul University

Chair of Advisory Committee: Dr. Craig J. Coates

Low level and variable transgene expression plague efforts to produce and characterize transgenic lines in many species. When transformation efficiency is high, productive transgenic lines can be generated with reasonable effort. However, most efforts to date in mosquitoes have resulted in suboptimal levels of transformation. This, coupled with the large space and intensive labor requirements of mosquito colony maintenance makes the optimization of transformation in mosquitoes a research priority. This study proposes two strategies for improving transgene expression and transformation efficiency. The first is to explore exogenous promoter/enhancer combinations to direct expression of either the transgene itself, or the transposase required for insertion of the transgene into the genome. An extension of this strategy is to investigate the use of a powerful viral *trans*activating protein and its cognate enhancer to further increase expression of these targets. The second strategy involves the identification of an endogenous boundary element for use in insulating transgenes and their associated regulatory elements. This would mitigate the inappropriate expression or silencing of many transgenes inserted into "unfavorable" genomic environments as a

consequence of an inability to specifically target the integration of transposons currently used in mosquito transgenesis.

The IE1 *trans*activating protein and its cognate enhancer from a baculovirus were shown to significantly increase expression of a reporter gene from three different promoters in cultured mosquito cells. Other heterologous enhancer/promoter combinations resulted in minimal increases or insignificant changes in expression.

Orthologues of the vertebrate insulator-binding factor, CTCF, were cloned and characterized in two mosquito species, *Aedes aegypti* and *Anopheles gambiae*. The expression profile of mosquito CTCF is consistent with its role as a putative insulator-binding protein. Preliminary binding site studies reveal a C/G-rich binding site consistent with that known in vertebrates and indicate that CTCF may bind widespread sites within mosquito genomes.

DEDICATION

The publication of this work brings my formal doctoral studies to a close, yet invites me to continue on a lifelong journey that will continue to reveal new insights, birth new possibilities and bring new dreams to fruition. I am in awe of the immense realm of possibility that my education affords me. As such, I find myself reflecting upon the many gifts and blessings that have brought me to this time in my life. I thank God each day for the precious gift of life, and I ask for His blessing upon my parents as I dedicate this work to them.

Mom and Dad

Thank you for encouraging me to reach beyond my self-imposed limitations and for believing in me throughout this journey. You fostered my hunger to learn early in my life and encouraged me to find my own unique way, even when you didn't understand all of my choices and decisions. As a result, I learned to trust myself and to believe that my desire could carry me where I dreamed to go. You gave me wings to fly, but always supported me in the difficult times. Know that I love you and that no words that I could ever write would be sufficient thanks.

ACKNOWLEDGEMENTS

No one lives in a vacuum. I count many faculty, friends, colleagues, students and family members among those who have significantly impacted my life and have helped to bring me to this place in my life. It is not my desire to leave out anyone, yet, at the risk of doing so, I would like to take the time to thank the following individuals.

Dr. Craig J. Coates—you encouraged me to think "outside the box" by leaving the door of possibilities open. Though I sometimes grew frustrated when I needed "an answer" or even a project, I was always able to step back, allow it to develop and figure it out. I never could have dreamt where this project would go. It's been a great experience, and I have truly learned much in the process. I will remember how freeing that "hands off" style was for me as I seek to carve out the space for future students to grow into their own in my own classroom and lab. Thank you for treating me as a colleague and for valuing what I think. That is truly a gift that you give to your students.

Dr. Dorothy E. Shippen—you listened when I was "really green" and were truly supportive in the messiness of a "project taking shape". I value your insight, and I appreciate the support you have gifted me with in my graduate career.

Dr. Michael P. Kladde—you served initially on my advisory committee, but still continued to freely offer great suggestions and a very helpful graduate student.

Dr. Ellen W. Collisson and Dr. Timothy C. Hall—you served on my advisory committee, lent me many good ideas and engaged in several fruitful discussions as this project unfolded.

To the members of the Coates Lab, past and present, for your support, camaraderie and generally for making the lab a good place to work during these past six years. I am particularly grateful to: Ahmed Mohammed—for your sense of humor and for providing me a window into a culture with which I did not have much experience; Andrea Julian—for your programming expertise and for writing the scripts which enabled me to complete material presented in section 5 of this dissertation; Darren Hagen—for your unique sense of humor, for being a sounding board on some of the crazier days and for some *Anopheles gambiae* genomic DNA and total RNA samples with which I completed some of the RT-PCR and RACE work presented in section 3 of this dissertation; David Pledger—for sharing your protocols for protein expression and purification; Jeremy Haag—for your invaluable help in cloning those many constructs early on in this project; Haiwen Li—for doing countless little things to make my life easier and for confirming Western blot data "under the gun".

Many friends have continued to support me throughout this time. I particularly thank my fellow long-time teaching colleague, Carol Larson—for your perspective and continued encouragement when times got difficult. I also recognize Heather and Paul Ketchum—for providing me, and certainly Mae and Mati, with a refuge from reality on many occasions. I cannot imagine better friends than you have been. I am grateful to Archana Dhasarathy—for your willingness to bring a ChIP novice up to speed, for early morning racquetball, and, most of all, for your warm friendship.

To the many students who have journeyed through my classroom, past and present—each of you brought your own unique perspective and vision. I often marvel at

the wealth of experience that you have collectively shared with me. I am heartened by your energy and your optimism, and I believe that the future is one of hope. Thank you for teaching me so much.

Finally, I acknowledge my family, particularly my parents, my sister, Julie and my brother, Raymond—for always loving me and being there no matter what life has "dished out". We do not choose our families, yet I cannot imagine a more special group of people. To each of the "little ones"…I challenge you to courageously dream. Dreams often become reality, though not always exactly as we initially imagine. If you reach high and you believe, good things will follow.

TABLE OF CONTENTS

Page

Page

Page

Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

1.1. Defining the Problem

1.1.1. Mosquito-vectored pathogens

Numbering greater than 3500 species and subspecies, mosquitoes are an enormously diverse and widespread group of mostly anantogenous dipterans, and, as such, are arguably the most important arthropod group of medical and veterinary significance (Eldridge, 2005). The females of some species preferentially blood-feed on particular host species and have co-evolved with various viral, protozoan and nematode pathogens, resulting in the ability to efficiently transmit a multitude of vertebrate pathogens. This study directly involves two of the most potent mosquito vectors of human pathogens, *Aedes aegypti*, commonly known as the Yellow Fever mosquito, and *Anopheles gambiae*, commonly known as the Malaria mosquito.

1.1.2. Ae. aegypti*: life cycle, distribution and human impact*

The mosquito, *Ae. aegypti*, inhabits virtually all tropical and sub-tropical areas of the world (Centers for Disease Control, 2005). Sprawling urbanization, increasing human population density, insecticide resistance, and an increasingly favorable climate continue to favor further expansion of the *Ae. aegypti* range (Gubler, 1998). *Ae. aegypti* females, like all anautogenous mosquitoes, require the large amounts of protein found in a vertebrate blood meal in order to complete egg development and maturation prior to oviposition (Eldridge, 2005). In addition, *Ae. aegypti* females are day-biting,

---

This dissertation follows the style and format of Gene.

urban-dwellers, and have a distinct preference for humans (Gubler and Clark, 1999). Once a viremic blood meal is ingested, viral particles must successfully infect the cells lining the midgut, replicate, escape the midgut and then migrate to and invade other tissues for dissemination, particularly the salivary glands. Though they usually mate only once, females retain enough sperm for the fertilization of several batches of eggs, each requiring a new blood meal. Once infected with a virus, the female will remain infected for the duration of her life (Blair et al., 2000).

Ae. aegypti is the principal vector of two important flaviviruses, the causative agents of Dengue Fever (DF) and Yellow Fever (YF). DF is the most widespread vector-borne human viral disease with as much as one third of the world's population at risk (Stephenson, 2005). 50 to 100 million cases of DF and several hundred thousand cases of the more severe Dengue Haemorrhagic Fever (DHF) occur annually across the globe (Centers for Disease Control, 2005). Approximately 5,800 cases of Yellow Fever (YF) are reported, but its incidence is likely higher and rising. It is predicted that urban epidemics of Dengue and Yellow fever may recur, because nearly all major urban centers have been reinfested with *Ae. aegypti* over the past 20 years (Gubler and Clark, 1999). Though an effective and cost-effective vaccine exists for YF, efficient delivery to disease-endemic areas is problematic. Because four distinct serotypes of Dengue virus exist, each with several genotypes, an effective vaccine against DF is unlikely to be produced. Immune responses to successive Dengue virus infections caused by different serotypes often exacerbate rather than mitigate the disease due to a phenomenon known as antibody-mediated disease enhancement (ADE) (Stephenson, 2005).

1.1.3. An. gambiae*: life cycle, distribution and human impact*

Of approximately 400 anopheline mosquito species, 60 serve as potential vectors of human malaria parasites, with ~30 of these posing a significant threat to human health in over 100 countries inhabited by roughly 40% of the world's population (World Health Organization, 2002). As many as 300 million cases of malaria occur across the globe each year, resulting in more than one million deaths. Upwards of 90% of the fatalities occur in sub-Saharan Africa, with most of these being children under the age of five (World Health Organization, 2003). The causative agent of human malaria is one of four protozoan species: *Plasmodium vivax, P. ovale, P. malariae and P. falciparum*, with *P. falciparum* causing the most severe disease. It is no coincidence that the most effective *P. falciparum* vector, *An. gambiae*, resides in sub-Saharan Africa where many countries lack the infrastructure and financial resources to mount effective and sustainable anti-malarial campaigns (World Health Organization, 2003). Exceedingly anthropophilic, *An. gambiae* essentially co-habitates with human populations, and thus feeds almost exclusively upon them. In addition, plasmodium-infected *An. gambiae* females tend to take more blood meals than non-infected individuals (Koella et al., 1998). Malaria parasites exhibit a complex lifecycle requiring both mosquito and vertebrate hosts. Upon ingestion of gametocytes in an infective blood meal, *Plasmodia* must complete their sporogonic cycle within the mosquito. To do so, the parasites must successively invade the midgut, escape the midgut and invade the salivary gland tissues.

1.1.4. Current control strategies are inadequate

Control of mosquito vectors must be multi-faceted and integrated. This requires mobilization of considerable financial and human resources, as well as constant public education and awareness. Currently, control of mosquito vectors includes chemical insecticides, environmental management and biological control, however serious drawbacks are associated with each of these strategies.

Economics drives the insecticide market to focus primarily on insects of agricultural importance with different formulations of these same insecticides being produced for use in human health applications (Hemingway and Ranson, 2005). As such, many insect vectors have already been exposed to all four major classes of insecticides (organochlorines, organophosphates, carbamides and pyrethroids) through breeding and resting in agricultural areas (Hemingway and Ranson, 2005). Repeated exposure of insect vectors to the same insecticides drives selection of those individuals with mutant alleles conferring insecticide resistance, eventually rendering an insecticide largely ineffective. Multiple examples of this have been described (Hemingway and Ranson, 2000; Brooke et al., 2002; Chandre et al., 1999). Arguably the cheapest, and possibly the most effective insecticide for use in mosquito control is the organochlorine, DDT. DDT poses significant environmental risk however, due to its excellent stability and bioaccumulation in the fat tissues of many non-target organisms further up the food chain (reviewed by Jaga and Dharmani, 2003). The greater expense associated with other classes of insecticides greatly impacts poor nations who must often choose between food production and efficient vector control. Certainly, the scope of

compounds exploited for chemical control could be expanded to include developmental inhibitors, attractants, repellents and chemosterilants (World Health Organization, 1982).

Knowledge of the specific environmental requirements of a particular vector species allows one to manipulate the environment in order to make it less hospitable. Mosquitoes all require water at some point in their life cycle—most often for both larval and pupal development. The connection between aqueous habitat and mosquito-borne disease incidence has been known for a long time (Ronald Ross' study of anophelines and malaria transmission in the 1890's and Finley and Reed's study connecting *Ae. aegypti* and Yellow Fever transmission in the early 1900's) and was suspected as far back as the 1st century, yet the advent of synthetic insecticides seems to have engendered apathy in environmental control efforts (Small, 2005). Each species prefers certain water conditions with optimal turbidity, organic content, oxygen content, salinity, temperature and surrounding vegetation of optimal density (reviewed by Small, 2005). Marshes and ponds can be drained or filled, while streams and rivers can be impounded or channeled into canals. Vegetation can be cleared or cut back to alter the ground cover and shade characteristics of the surrounding habitat. Care must always be taken to not trade one problem for another by creating conditions favorable to other endigenous vector species, as well as to avoid wholesale destruction of native ecosystems. Continued urbanization has introduced new incidental breeding sites such as flower pots and vases, discarded plastic containers, water cisterns, hoof- and footprints and used tires, while rice cultivation and wetlands preservation in residential areas provide ample breeding sites nearer to human populations (Small, 2005). Most importantly, enlisting

the help of the general population through community-wide education and awareness programs will significantly reinforce publicly-funded control measures.

Biological control efforts may include the introduction of a predator, pathogen, parasite, competitor or microbial toxin in order to reduce the target pest population (Hemingway, 2005). Greater species specificity makes this strategy desirable, though most efforts to date have been focused on agricultural pest species. Introduction of the mosquito fish, *Gambusia affinis*, in the early 1900's provided initial larval control, but was largely supplanted by use of insecticides such as DDT in the 1940's (reviewed by Hemingway, 2005). As concerns grew over the long-term effects of insecticides upon the environment and non-target species, the nematode *Romanomeris culicivorax* and the protozoan *Nosema algerae* were extensively studied as potential vehicles for larval control (reviewed by Kaya and Gaugler, 1993; Chapman, 1974; Legner, 1995). Field trials with these organisms met with mixed success, and ultimately, high production costs and difficulties in shipping and storage made these organisms unacceptable (Hemingway and Ranson, 2005). The single most effective biological agent for mosquito larval control to date is the toxin produced by *Bacillus thuringiensis* subsp. *israelensis* (*Bti*). The toxin is highly specific for mosquito larvae, and the development of slow-release briquettes and floating granules has increased its persistence in the environment, making it more economically feasible for widespread use (Hemingway, 2005; Becker and Ascher, 1998). An extension of biological control is genetic control via sterile-insect technique (SIT), similar to the campaign that eradicated the screw worm fly from all of North America, much of Central America, and, more recently, in

Libya (Wyss, 2000; Lindquist et al., 1992). Thus far, attempts to control vector mosquito populations have largely failed due to high population density and massive immigration/emigration patterns of target populations, failure to produce enough sterile males on a daily basis to achieve effective sterile to wild-type ratios and inefficient sexing mechanisms to ensure the release of males only (reviewed by Wood, 2005). Though potentially a powerful tool, biological control is not sufficient in itself to reduce the disease burden imposed by mosquito vectors.

1.2. A Proposed Ideal Solution

1.2.1. Creation of strains incapable of pathogen transmission

In the last 15 years, scientific focus has shifted from population elimination to population modification using an alternative control strategy to genetically engineer strains of mosquitoes that are refractory to pathogen transmission (Meredith and James, 1990; James, 2000). The advent of stable transformation technologies in mosquitoes (Jasinskiene et al., 1998; Coates et al., 1998) provided the key tool to make this possible. Researchers have succeeded in the creation of refractory strains of anopheline mosquitoes in which transmission of *Plasmodium* parasites is greatly reduced, but not completely eliminated (Ito et al., 2002; Moreira et al., 2002; Kim et al., 2004). In *Ae. aegypti*, RNA interference (RNAi), a natural anti-viral defense in a wide-range of organisms, is being investigated and has shown great promise in targeting viral-specific transcripts necessary for viral replication (Travanty et al., 2004; Sanchez-Vargas et al., 2004). Recently, stable and heritable gene silencing via RNAi triggered from stably-integrated transgenes with an inverted repeat configuration was demonstrated in *An.*

stephensi, an important malaria vector in Asia (Brown et al., 2003). None of the refractory strains created thus far completely abolishes the possibility of transmission, leaving a window for the development of counterstrategies of transmission by surviving pathogens. It is difficult to assess this risk however, because the experimental viral load/parasite numbers are typically much higher than occur in nature. This highlights the importance of continued investigation in an effort to identify and characterize multiple gene targets directed at multiple stages in the infection cycle that, when combined, will eliminate pathogen transmission. Single-chain antibodies, molecules involved in innate immune responses, antiparasite toxins and synthetic peptides could all potentially be exploited (Nirmala and James, 2003). Molecules associated with mosquito host-seeking behaviors (Zwiebel and Takken, 2004) could also be further characterized for use as potential transgene targets. Finally, the reality of parasite and viral co-evolution in response to host vector defense-mechanisms demands a much better understanding of the molecular interface between the pathogen and the mosquito vector. Introduction of additional genetic material will potentially trigger pathogen responses which may not be anticipated.

1.2.2. Movement of refractory genes into natural mosquito vector populations

Once suitable laboratory strains refractory to pathogen transmission are possible, the next step would be to introduce the refractory genes/alleles into natural populations in an attempt to replace competent vectors with incompetent ones. This task poses multiple problems and demands genetic drive mechanisms beyond simple Mendelian genetics. Mendelian transmission would take far too long to establish an introduced

gene. Assuming that introduced individuals are homozygous for the transgene, in one generation of mating with the wild population, all offspring of the introduced insects would be heterozygous. These heterozygous individuals that then mate with each other would produce 25% of offspring homozygous for the wild-type gene, 50% of offspring heterozygous like themselves and only 25% of offspring true-breeding for the transgene. The chances that the offspring mate with a wild-type insect rather than with a sibling would also be much higher due to sheer numbers. It would not be possible to introduce sufficient numbers of transgenic individuals to ensure that more individuals mate with the introduced individuals, thus further increasing the numbers of progeny homozygous for the transgene. If the goal is to drive the transgene to fixation in the natural vector population, Mendelian genetics will fail miserably on its own. Some mechanisms that could potentially be employed include active transposable elements (TEs), meiotic drive, endosymbiotic bacteria, viruses and male-specific insecticide resistance (reviewed by Braig and Yan, 2002).

The first, and perhaps most obvious choice is making use of the TEs that are used in the creation of transgenic individuals. TEs, first described by Barbara McClintock in maize (McClintock, 1950), are naturally-occurring, mobile elements capable of spreading rapidly within populations. Perhaps the most poignant example of this occurred with the spread of the *P*-element throughout the world's *Drosophila melanogaster* populations over a period of approximately 30 years (Kidwell et al., 1983; Anxolabehere et al., 1988). Prior to 1950, *D. melanogaster* wild-type populations contained no *P*-elements, while all natural populations sampled after 1978 carried

multiple copies of *P*-elements, introduced by horizontal transfer from *D. willistoni* (Anxolabehere et al., 1988; Ribeiro and Kidwell, 1993). Despite hybrid dysgenesis, which causes massive reductions of reproductive fitness upon initial introduction into some strains (M cytotype) due to increased genomic instability in the germ-line (Engels, 1989; Rio, 1990), *P*-elements replicate, presumably via DNA repair machinery, to high levels in the permissive strains (P cytotype). Active TEs, like *P*-elements, invade non-homologous chromosomes by targeting short, frequently occurring target-sites and ultimately stabilize with multiple copies per genome. Stabilization is hypothesized to be mediated by a balance between positive selection for greater replicative events and negative selection against unfavorable genome rearrangements and gene interruption caused by TE insertion (Kidwell and Lisch, 2000). Clearly TEs could provide a powerful tool for genetic drive, but these must be studied in the context of the target species in terms of both regulation and the unique genomic environment.

Meiotic drive results from the interaction between two genetic elements, a driver locus and a responder locus, causing the destruction or disabling of part, or all of the homologous chromosome and production of a disproportionate number of gametes carrying the linked alleles. The best-studied example of meiotic drive is the segregation distorter (*SD*) system of *D. melanogaster*, where males heterozygous for the *SD* chromosome and the wild-type SD^{+} chromosome normally produce in excess of 95-99% *SD*-carrying sperm (Hartl et al., 1967). Nearly all naturally-occurring populations of *D. melanogaster* examined possess *SD* chromosomes carrying *SD*, *E(SD)* (*Enhancer of Segregation Distortion*) and highly insensitive *Rsp* (*Responder*) alleles with much less

polymorphism within the immediate genomic region where these three alleles reside (reviewed by Akio et al., 2004). It appears that all closely-linked loci are simultaneously selected in a type of "genetic hitchhiking" effect (Palopoli and Wu, 1996). A meiotic drive system, with interesting parallels to the *SD* system in *D. melanogaster* (Wood and Newton, 1991), has long been known in *Ae. aegypti* (Hickey and Craig, 1966; Wood and Ouda, 1987). Sex is determined by a pair of alleles, M and m, on chromosome 1, where Mm results in males and mm results in females (Gilchrist and Haldane, 1947). The meiotic drive gene (M^D) produces a gene product that acts in *trans* upon an unknown gene, $m^{i/s}$, tightly-linked to the sex-determining locus. When the sensitive allele, m^s, is present, fragmentation of the m^s-carrying chromosome results (Newton et al., 1976). Renewed interest in the mechanisms of meiotic drive led to the establishment of a new laboratory strain, T37, from a natural population sample collected in Trinidad that will enable further characterization of this system in *Ae. aegypti* (Akio et al., 2004). Much remains unknown with regards to other alleles which may influence the interaction between M^D and m^s, as well as the natural frequencies of susceptible versus insensitive strains, yet meiotic drive represents a powerful strategy for driving refractory transgenes into native *Ae. aegypti* populations within a very short period of time.

An obligate intracellular bacterium, *Wolbachia*, spreads rapidly through populations of many different arthropod species, most commonly via cytoplasmic incompatibility (CI), producing patterns of crossing sterility within or between populations. Unidirectional CI results when infected males mate with uninfected females, resulting in complete or partial sterility, while the reciprocal cross is fully fertile.

This favors the production of infected offspring by infected females thus increasing the numbers of infected individuals with each additional generation (reviewed by Sinkins and O'Neill, 2000). CI likely results from the targeting of cell cycle regulatory proteins, as illustrated in the asynchronous nuclear envelope breakdown of the male and female pronuclei of the wasp, *Nasonia vitripennis*, with subsequent loss of the male chromosome complement (Curtis, 1976). Controlled, outdoor cage studies with the urban filariasis vector *Culex quinquefasciatus* established proof of principle that cytoplasmic replacement is possible, yet older males exhibited incomplete CI, making further studies necessary (Curtis, 1976). Improved and cheaper drug regimens for filarial infections have shifted the focus of disease control to treatment rather than elimination of transmission via vector control, so these studies remain incomplete (Curtis and Sinkins, 1998). Naturally-occurring *Wolbachia* infections, with wide tissue distribution, have been described in numerous *Culex* and *Aedes* species, but not in anophelines. It remains to be determined if *Anopheles* species can support *Wolbachia* and CI, as it is unclear if they have natural refractoriness or if the ecological conditions necessary for interspecific transfer have simply not been present naturally (Sinkins, 2004). Superinfections of two or more *Wolbachia* strains within individuals also commonly occur with different strains of *Wolbachia* showing no discernable host species preference (Sinkins, 2004). While much remains to be learned about specific *Wolbachia*-host interactions, two distinct strategies have been envisioned for using *Wolbachia* as an effective transgene drive mechanism: introduction of stably-transformed *Wolbachia*, or transformation of mosquitoes with both the desired transgene

and genes able to induce/rescue CI. Finally, superinfection and bidirectional CI make use of *Wolbachia* attractive, as a failed attempt could be rescued by introduction of an additional genetically-engineered *Wolbachia* strain (Sinkins, 2004).

Transduction involves the use of viruses to package and deliver genes of interest to cells via their natural route of infection. Of many possible viral transducing agents, densoviruses evoke the most interest as vehicles for the introduction and spread of transgenes within target mosquito populations. Densoviruses infect arthropods, primarily insects. Specific viruses within this group exhibit a narrow host range making them less likely to infect non-target species (Carlson et al., 2000). Infectious clones have been constructed for the *Ae. aegypti* densovirus (*Ae*DNV). These have proven to be easily manipulated to carry genes of interest, are capable of infecting multiple tissues in adult mosquitoes, and have shown some potential for vertical transovarial transduction (reviewed by Olson et al., 2005). *Ae*DNV is among the best-characterized of these viruses, yet much remains unknown regarding its molecular biology. This promising avenue continues to demand much attention from vector biologists.

Finally, a recent study describes the feasibility of linking a transgene to a male-specific insecticide resistance allele (Sinkins and Hastings, 2004). The construct would be propagated in both sexes, but the insecticide resistance mechanism would be linked to a male-specific promoter or splicing event, thus creating a fitness advantage in males that is greater than the genetic load imposed on females. The authors propose a number of attractive features: release of only non-biting males, release of small numbers of transgenic individuals, and positive selection pressure only in the presence of the

insecticide, thus allowing investigators to control the area of spread. Removal of the selective pressure by changing the class of insecticide would result in rapid loss of the transgene construct allowing reversal of a failed control strategy.

1.3. Mosquito Transgenesis: The Current State of Affairs

1.3.1. The P*-element paradigm*

The successful germ-line transformation of *D. melanogaster* using *P*-element derived vectors (Rubin and Spradling, 1982) provided a powerful tool for the development of the vinegar fly as a model genetic organism. *P*-element transformation has directly facilitated gene discovery and characterization, investigation of gene interactions at the DNA, RNA and protein levels and elucidation of gene regulatory mechanisms in this insect (reviewed by Coates, 2005). The lure of such a powerful tool perhaps hindered efforts to transform other insect species by restricting the focus of researchers to the *P*-element system. Two possible reasons have been proposed to explain the failure of *P*-elements to transform nondrosophilid insects: a.) the wide distribution of *P* and *P*-like elements in genomes has triggered the evolution of specific repression mechanisms to mitigate the mutagenic properties of transposon mobility; b.) host-specific factors that both mediate and regulate *P*-element activity are largely confined to *D. melanogaster* and closely-related sibling species, while the structure of the cleaved termini generated during the transposition process is not typical of other class II transposable elements (Atkinson and James, 2002). This drove the search for other class II TEs that could be harnessed for transgenesis and as drive mechanisms in medically-important mosquitoes and in other species of medical and agricultural interest.

Although *P*-elements proved unacceptable for transgenesis in nondrosophilid species, the basic binary transformation system utilized in *P*-element transformation remains the paradigm for transformation with subsequently identified TEs.

All transgenic insects produced thus far harness the components of a class II transposon (Finnegan, 1985) and are generated via microinjection of a mixture of two plasmid DNAs into preblastoderm embryos. Endogenous class II TEs catalyze their own movement within the host genome via a "cut and paste" mechanism with canonical transposition marked by duplication of the target site upon integration. In insect transgenesis, the system is partitioned into two plasmids (Fig. 1). The first plasmid, termed the donor plasmid, consists of the transposons inverted terminal repeats (ITRs), any internal sequences necessary for transposition, a marker gene and its promoter and a gene of interest along with its regulatory sequences. The second plasmid, termed the helper plasmid, carries the coding sequence for the transposase along with a strong, constitutive or inducible promoter to direct its expression at high levels within the cells of the early embryo. The transposase, *in-trans*, catalyzes the movement of the donor cassette from the donor plasmid into a target site within the genome. If the integration occurs in a germ-line cell, the donor cassette will be inherited by the progeny of the individual, as a dominant allele, in a Mendelian fashion. This only occurs for a small fraction of integration events, with the vast majority being somatic integrations that affect only the current generation. The helper plasmid cannot integrate, thus the ability to produce the transposase is lost with subsequent cell division and dilution/degradation of the helper plasmid. Assuming that the host species does not possess active,

endogenous copies of a similar TE, the donor element will remain stably integrated at its new site.

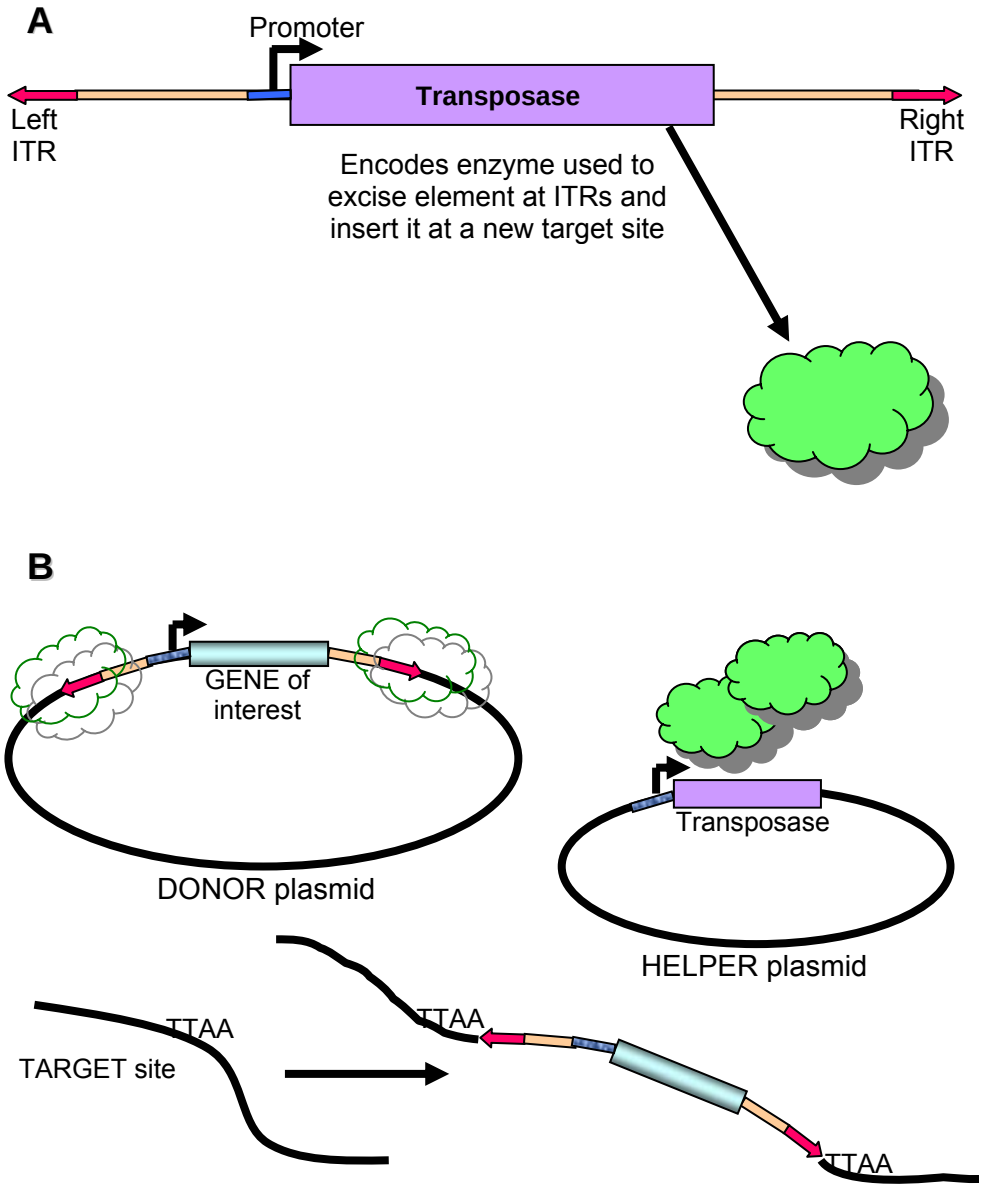


Fig. 1. Use of class II transposable elements (TEs) for stable germ-line transformation. (A) Class II Transposons move autonomously via the enzymatic gene product (transposase) encoded by the open reading frame. This product recognizes the Inverted Terminal Repeats (ITRs) and moves the entire element into a new site within the genome. (B) Binary Transposable Element System involves the production of transposase by a non-integrating helper plasmid. This transposase recognizes the ITRs flanking the transgene cassette in the donor plasmid and catalyzes the movement of the transgene into the target genome at a cognate target site. This movement is characterized by a target site duplication.

1.3.2. Mosquito transgenesis

The failure of *P*-element transformation in mosquito species led researchers to test newly-identified and characterized class II TEs, *Hermes*, *MosI*, *Minos* and *piggyBac* (Atkinson et al., 1993; Medhora et al., 1991; Franz and Savakis, 1991; Fraser et al., 1996), using *in vivo* interplasmid transposition assays (Sarker et al., 1997). These assays are used to assess the mobility of the TE within the target species, prior to investing the significant time and effort required to generate stable transgenic lines. In these assays, the donor and helper plasmids are injected into preblastoderm embryos along with a target plasmid. Twenty-four hours later, the target plasmid is recovered and analyzed by restriction analysis and/or PCR amplification for integrations of the donor cassette. Following the development of universal, fluorescent marker genes with no host cell requirements (Pinkerton et al., 2000; Horn and Wimmer, 2000; Horn and Peterson, 2002), several different class II TEs have subsequently been used to generate germ-line transformants in a number of medically important mosquito species (Table 1). Nonetheless, mosquito germ-line transformation is largely inefficient, labor intensive and costly compared to *Drosophila* transformation. Typical mosquito transgenesis experiments begin with 10-20 fold more injected embryos, but result in far fewer stable lines. Maintenance of mosquito colonies also requires much more space due to the different environmental conditions needed for rearing each life stage, and the actual physical segregation of life stages requires much time. Single-pair matings are all but impractical, and the blood meal required for egg production necessitates the care and maintenance of numerous vertebrate hosts. With the exception of *Ae. aegypti* and a few

related species, for which egg papers can be stored for a period of time, mosquito lines must be continuously maintained, making the prolonged study of more than a few transgenic lines impractical, even in large insectaries. While there will likely never be genetic transformation studies on the scale possible for *Drosophila*, further optimization of mosquito transformation would advance both the understanding of mosquito vector biology and the development of genetic control strategies for important vector species.

The reasons for such low transformation efficiencies are likely complex. It has become apparent that TEs do not always behave as expected in the context of different genomic environments. The *Hermes*, *MosI* and *piggyBac* TEs all occasionally show non-canonical integration events where the entire donor cassette along with some or all of the flanking plasmid sequence inserts into the germ-line in *Ae. aegypti* (reviewed by Tu and Coates, 2004). A similar phenomenon occurs with *Hermes* insertions into the *Cx. quinquefasciatus* genome (Allen et al., 2001). Large, unstable tandem arrays of the *piggyBac* TE are observed in approximately 50% of *Ae. aegypti* transgenic lines (Adelman et al., 2004). Attempts to remobilize integrated transgenes in the germ-line by supplying an exogenous source of transposase have largely been ineffective (O'Brochta et al., 2003). Recent whole genome sequencing projects have revealed a diverse range of TEs in mosquitoes, so these unexpected observations could reflect interactions between endogenous and exogenous TEs with similar ITR sequences (Tu and Coates, 2004). Clearly, the "black box" of endogenous TE regulation must be better understood in mosquito species.

Table 1

Medically-important mosquito species transformed[a]

Species	Transposable Element (host species)	Marker Gene/Promoter	Transformation Efficiency[b] (published)	Reference
Aedes aegypti	*Hermes* (*Musca domestica*)	*cn*$^+$	~8%	(Jasinskiene et al., 1998)
	Mos1 (*Drosophila mauritiana*)	*cn*$^+$	~4%	(Coates et al., 1998)
	piggyBac (*Trichoplusia ni*)	*EGFP*/3 X P3	~5-10%	(Kokoza et al., 2001)
	Tn5 (*Escherichia coli*)	*DsRed*/3 X P3	0.22%	(Rowan et al., 2004)
Culex quinquefasciatus	*Hermes*	*EGFP*/*Actin5C*	~12%	(Allen et al., 2001)
Anopheles stephensi	*Minos* (*Drosophila hydei*)	*EGFP*/*Actin5C*	~7%	(Catteruccia et al., 2000)
	piggyBac	*DsRed*/*Actin5C*	~4%	(Nolan et al., 2002)
Anopheles gambiae	*piggyBac*	*EGFP*/*Hr5-IE1*	~0.5%	(Grossman et al., 2001)
Anopheles albimanus	*piggyBac*	*EGFP*/*PUBnls*	20-43%	(Perera et al., 2002)

[a]Only the first published account of each species transformed with a particular element is represented.

[b]Represented as the number of transgenic lines per fertile G_0 offspring.

A number of efforts have been made to increase the frequency of germ-line integration events. Alternative sources of transposase have been used, such as purified recombinant protein (Coates et al., 2000) and *in vitro* transcribed RNA (Kapetanaki et al., 2002), in an effort to circumvent inefficient heterologous promoter utilization, avoid codon bias and to reduce the time required from injection to integration. Both efforts resulted in greater observed transformation frequencies, however both strategies highlighted the need to investigate endogenous regulation of TEs and how those unknown regulatory mechanisms might impact TE-mediated transgene integration. Recently, creation of transgenic *Ae. aegypti* using the *Tn5* TE from *E. coli* harnessed pre-formed synaptic complexes—intermediates formed in the "cut and paste" integration process (Rowan et al., 2004). Though not effective in this case, the strategy certainly warrants further investigation as a means to facilitate the required interaction between the transposase and donor DNA. Current research focuses on identification and utilization of endogenous, germ-line-specific promoters and developmental mRNA targeting sequences to maximize transposase expression at developmentally appropriate times and in specific regions of the early embryo. In this project, heterologous promoter/enhancer combinations were tested in cultured cells for potential use in producing higher levels of transposase, as well as for their potential to drive high-level expression of transgenes.

1.3.3. Addressing variable expression of transgenes due to position effect and position effect variegation

Transgene expression often varies both between and within transgenic lines due primarily to two well-known phenomena, position effect (PE) and position effect variegation (PEV). Position effect arises as a direct consequence of different integration sites (Wilson et al., 1990), due to the transcriptional status of the site and/or the influence of nearby enhancers or repressors of gene expression (Fig. 2). An example of this is particularly evident in the early transformation experiments in *Ae. aegypti*, where the kynurenine hydroxylase gene (*cinnabar*) from *D. melanogaster* was used as the transgenic marker in the white-eyed kw^h (Bhalla, 1968) strain, and different transgenic families exhibited marked variation in eye color, from light orange to very dark purple/black that was nearly indistinguishable from wild-type (Jasinskiene et al., 1998; Coates et al., 1998). The only difference from one family to the next was the integration site of the transgene. Position effect variegation results when expression is varied between siblings of a family with a single, common insertion site that likely borders a euchromatic/heterochromatic boundary (Fig. 3). It can also be observed as "mosaic"-type expression within a particular tissue of a single individual as heterochromatin boundaries are not rigidly fixed (reviewed by Grewal and Elgin, 2002). Both types of variable expression have been observed in transgenic mosquito lines (Jasinskiene et al., 1998; Coates et al., 1998). This phenomenon complicates efforts to identify and characterize transgenic lines in organisms like *Ae. aegypti* and other culicine mosquitoes, where there are large amounts of interspersed heterochromatin throughout the genome

(Knudson et al., 1996). Perhaps the most promising strategy lies in the identification of endogenous insulators or boundary elements which could be used to insulate transgenes from the effects of their genomic environment.

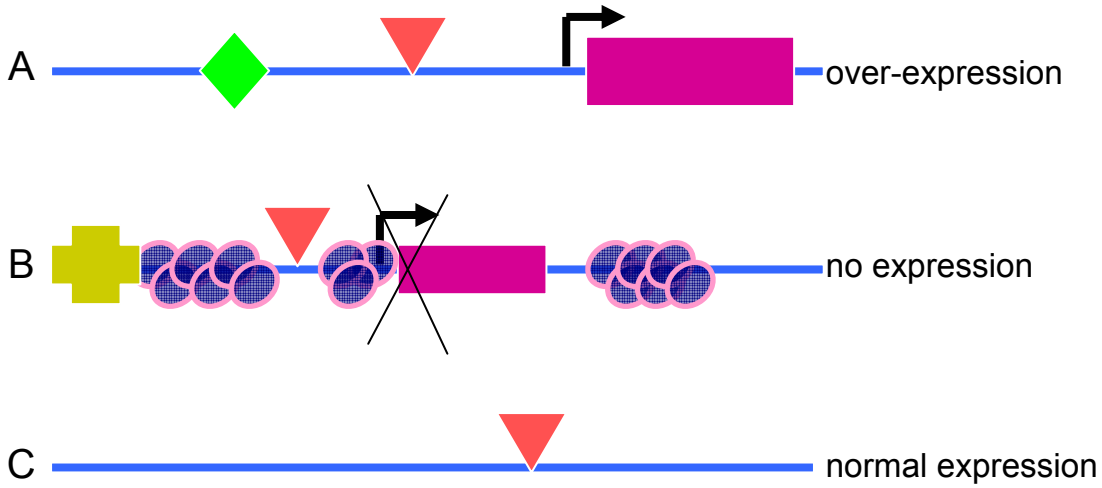


Fig. 2. Position effect. Due to different insertion points of a transgene (orange triangle) into the genome, one of three different expression patterns may result: (A) over expression of the transgene due to the influence of an upstream enhancer (green diamond), (B) no expression of the transgene due to the influence of an upstream silencer (black/yellow cross), (C) normal expression due to the absence of endogenous gene regulatory elements in the adjacent chromatin.

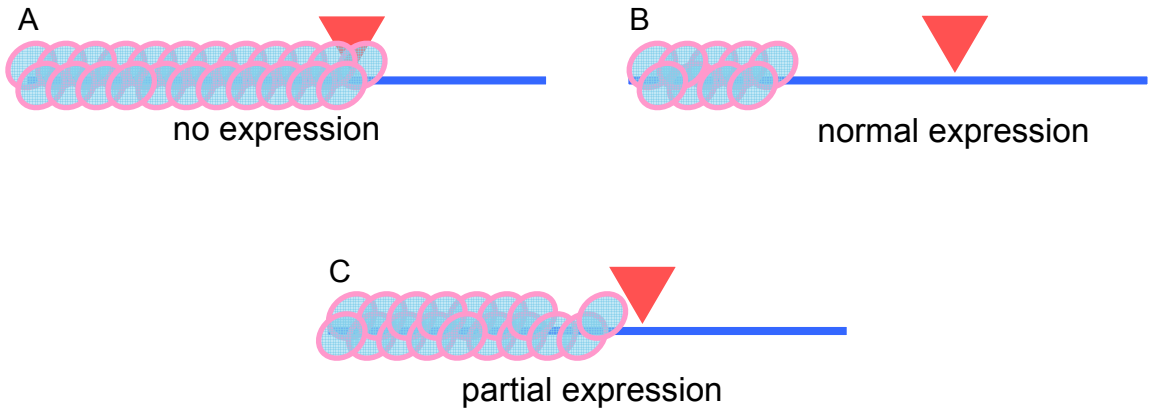


Fig. 3. Position effect variegation. Within a single transgenic family or within individual cells of a single tissue, heterochromatin (purple/pink ovals) boundaries may be variable with reference to the transgene. No expression (A) results when heterochromatin encroaches upon the transgene, normal expression (B) results when the transgene is clearly within open chromatin, partial expression (C) results when the heterochromatin boundary coincides with the position of the transgene.

Insulators, also referred to as boundary elements, vary widely in their DNA sequence and in the proteins that bind to them. Nonetheless, they have at least one of two properties (Fig. 4) related to barrier formation: ability to block the action of neighboring enhancers/repressors and/or ability to inhibit position effects (Bell et al., 2001). A number of different DNA sequences with insulating activity have been identified in both invertebrate and vertebrate species: scs/scs' (Udvardy and Maine, 1985) and a portion of the *gypsy* retrotransposon from *D. melanogaster* (Geyer and Corces, 1992), sites in the sea urchin histone H3 genes (*sns*) (Palla et al., 1997), human Matrix Attachment Regions (MARs) (Namciu et al., 1998), the chicken β-globin genes (cHS4) (Chung et al., 1993), the ribosomal RNA genes of *Xenopus* (Robinett et al., 1997), and the human T cell receptor (TCR)-α/δ locus (Zhong and Krangel, 1997). All of these insulating elements occur between genes with independent expression profiles, consistent with their ability to protect genes from neighboring regulatory elements (Bell et al., 2001). Several of these insulators occur in conjunction with DNase I-hypersensitive sites at the outermost boundaries of genes, consistent with their ability to protect genes from the encroachment of silencing heterochromatin. cHS4 has been used to generate consistent expression levels in transgenic rabbits, mice and cell lines (reviewed by Bell et al., 2001). In addition, a study of the specific binding site of the boundary element-associated factor (BEAF) protein in *Drosophila* has revealed an entire class of insulators, that includes scs', with the simple sequence motif of CGATA. Up to this point, insulators did not appear to have notable sequence homologies or common binding properties (Cuvier et al., 1998). Increasing evidence also points to regulation of

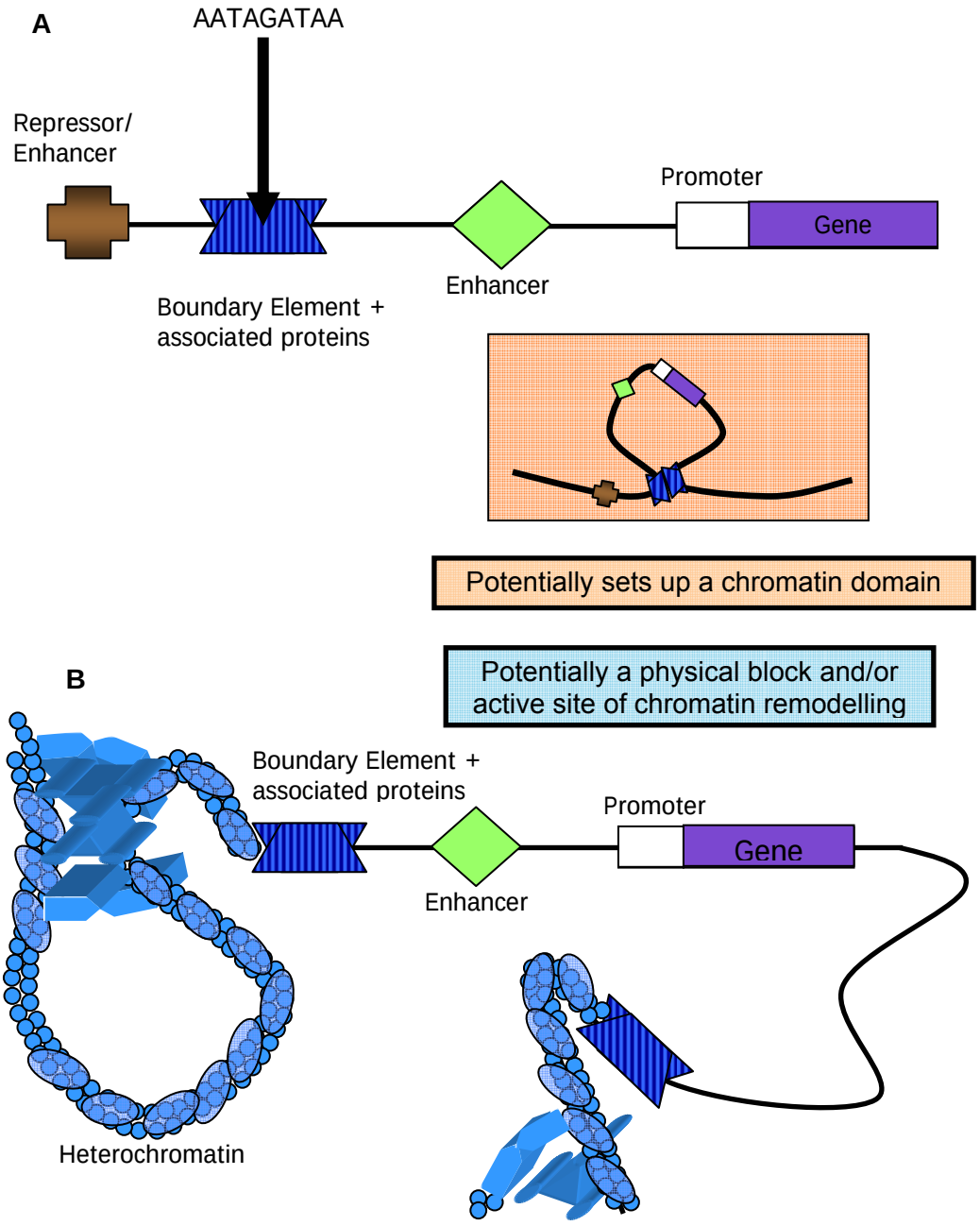


Fig. 4. Boundary elements. Functional boundary elements possess one or both of the functions depicted. (A) Enhancer-blocking and (B) Barrier Activity.

insulators (reviewed by Bell et al., 2001). Several mechanisms have been proposed to explain insulator activity, but a dominant model is one of binding proteins anchoring specific segments of DNA to the nuclear matrix and other proteins perhaps further cordoning off a region of DNA, thereby causing the formation of looped domains and specific regions within these domains. No enhancer or silencer could then act outside its own domain, nor could encroaching heterochromatin advance beyond this barrier. The anchoring proteins and the proteins further subdividing the domains into smaller regions could then be regulated differently, perhaps accounting for differential expression during development or in response to environmental stimuli (Bell et al., 2001). This study focuses on the identification and characterization of a mosquito protein orthologous to a well-known vertebrate insulator-binding protein, CCCTC-binding factor (CTCF). It is hoped that insulator sequences bound by this protein can be utilized to protect transgene constructs from PE and PEV in mosquitoes. Future study of this protein will also likely reveal important information regarding genome organization and the regulation of gene expression in these critical disease vector organisms.

2. HIGH-LEVEL GENE EXPRESSION IN *Aedes albopictus* CELLS USING A BACULOVIRUS HR3 ENHANCER AND IE1 *TRANS*ACTIVATOR

2.1. Introduction

Through the efforts of many individuals in the past few years, it has become possible to genetically transform a wide variety of non-drosophilid insects of medical and agricultural importance (Robinson et al., 2004). The ability to genetically transform mosquito species allows researchers to better understand mechanisms of vector competence, design novel methods to disrupt vector-pathogen relationships and develop new insect control strategies ((Collins and James, 1996); (Beerntsen et al., 2000); Blair, 2000; (Alphey and Andreasen, 2002). New molecular methods could potentially augment continued traditional efforts to control malaria and other re-emerging arthropod-borne diseases. Similar approaches may also be used to stem the devastating infestation of economically important crops by insecticide-resistant pest strains.

Mosquitoes transmit to humans some of the most debilitating and deadly diseases known. According to the World Health Organization, malaria alone is responsible for one million deaths annually (World Health Organization, 2003). Additionally, the transmission of yellow fever, dengue fever, West Nile virus and a variety of other encephalitis viruses permanently disrupt or end untold numbers of lives. Both anopheline (Catteruccia et al., 2000; Grossman et al., 2001; Nolan et al., 2002; Perera et al., 2002) and culicine (Jasinskiene et al., 1998; Coates et al., 1998; Pinkerton et al., 2000; Kokoza et al., 2001; Allen et al., 2001; Lobo et al., 2002; Ito et al., 2002) mosquito species have been successfully transformed. In all cases, the process is labor-

intensive with a few successful experiments yielding transformation efficiencies ranging from 0.5% to 13%. These transformation efficiencies are low compared to the nearly 50% previously reported in *Drosophila* with vectors up to 8 kb in size (Spradling, 1986). Additionally, transgene expression in the yellow fever mosquito varies considerably both between and within families (Coates et al., 1998; Jasinskiene et al., 1998; Coates et al., 1999), likely due to differences in the transcriptional environments of specific insertion sites within the genome, such as the proximity of the transgene to enhancers or heterochromatic stretches of DNA. This phenomenon is of particular concern in the *Ae. aegypti* genome given its large size (~780 Mb) and its apparent pattern of short-period interspersion where single copy genes (1 to 2 kb) alternate with short (200-600 bp) or medium (1-4 kb) length repetitive sequences (Knudson et al., 1996). The problem is complex; however, transposition has been shown to be dependent upon the amount of transiently available transposase to catalyze vector integration (Kapetanaki et al., 2002; Davidson et al., 2003). Also, the effective use of genetically-altered mosquitoes to augment current disease vector control requires the ability to create and maintain transgenic lines with consistent, predictable and high-level expression patterns of effector transgenes.

With a view to maximizing the transcription of transgenes that insert into favorable genomic environments and to potentially increase the levels of transiently available transposase, we tested the ability of three different enhancer elements; SV40 (Moreau et al., 1981), *copia* ULR (*Drosophila*) (Matyunina et al., 1996) and Hr3 (*Bombyx mori* nuclear polyhedrosis virus or NPV) (Lu et al., 1997), to increase the

levels of transcription from each of three heterologous promoters from the following genes: actin5C (Thummel et al., 1988) and polyubiquitin (Ubi-p63E—hereafter referred to as pUb) (Lee et al., 1988) from *Drosophila* and the intermediate early gene (IE1) (Pullen and Friesen, 1995) from the *Autographa californica* multicapsid nuclear polyhedrosis virus (MNPV). Additionally, we tested the ability of the *B. mori* baculovirus IE1 gene product (Lu et al., 1996), which binds to repetitive sequences within the baculovirus homologous regions (Hrs) (Guarino and Summers, 1986; Kremer and Knebel-Mordsdorf, 1998) and has previously been shown to function as a powerful *trans*activator in transfected lepidopteran cells (Lu et al., 1996), to yet further increase gene expression in mosquito cells.

2.2. Materials and Methods

2.2.1. Construction of the luciferase expression plasmids

Maps for all plasmid constructs built in this study can be found in Appendix A, in the order in which they are mentioned in the text. A 2.7-kb *Hind*III-*Sal*I fragment from pGL2-Basic (Promega, Madison, WI 53711), containing the firefly luciferase coding region and the SV40 poly-Adenylation signal, was inserted into the corresponding sites of pBCKS+ (Stratagene, LaJolla, CA 92037) to create pBCLuc (A-1). A 2.7-kb *Sma*I-*Sal*I fragment from pBCLuc was inserted into the *Sma*I-*Sal*I sites of pSL*fa*1180*fa* (Horn and Wimmer, 2000) to create pSLLuc (A-2). The *Drosophila* Actin5C promoter was excised from pHermesA5CEGFP (Pinkerton et al., 2000) by *Pst*I and *Bam*HI digestion and inserted into the corresponding sites of PSLLuc to create pSLAct5CLuc (A-3). The *Sac*II site was removed from pIE1-3 (Novagen, VWR

International, Bristol, CT 06011) and then the 657-bp *Eco*RI-*Bam*HI fragment containing the AcMNPV IE1 promoter was inserted into the corresponding sites of pSLLuc to create pSLIE1Luc (A-4). A 2-kb *Kpn*I-*Bam*HI fragment from pB[pUB-nls-EGFP] (Handler and Harrell, 2001) containing the *Drosophila* polyubiquitin promoter was inserted into the corresponding sites of pSLLuc to create pSLpUbLuc (A-5). The *copia* ULR was amplified by polymerase chain reaction (PCR) from *copia* LTR-ULR-CAT (Wilson et al., 1998) using the primers *copia*ULRForw and *copia*ULRRev (Table 2), cleaved by digestion with *Hind*III and *Sac*II and inserted into the corresponding sites of pBCKS+ to create pBCcULR (A-6). The *Hind*III-*Sac*II fragment from this plasmid was then inserted into the corresponding sites of pSLAct5CLuc and pSLIE1Luc to create pSLcULRAct5CLuc (A-7) and pSLcULRIE1Luc (A-8). pBCcULR was digested with *Hind*III and the site filled with the Klenow fragment of DNA polymerase I (Promega), then digested with *Sac*II and ligated into pSLpUbLuc which had been cut with *Not*I and the site filled with Klenow fragment, then cut with *Sac*II to create pSLcULRpUbLuc (A-9). The SV40 enhancer region from pRL-SV40 (Promega) was PCR-amplified using the primers SV40Forw and SV40Rev (Table 2), digested with *Hind*III and *Sac*II and inserted into the corresponding sites of pBCKS+ to make pBCeSV40 (A-10). The *Hind*III-*Sac*II fragment of pBCeSV40 was then inserted into the corresponding sites of pSLAct5CLuc and pSLIE1Luc to create pSLeSV40Act5CLuc (A-11) and pSLeSV40IE1Luc (A-12). pBCeSV40 was digested with *Hind*III, the site filled with Klenow fragment, then digested with *Sac*II and ligated to pSLpUbLuc digested with *Not*I, the site filled with Klenow fragment and subsequently digested with *Sac*II to

produce pSLeSV40pUbLuc (A-13). A 1.2-kb *Pst*I-*Bam*HI fragment containing the *B. mori* NPV Hr3 enhancer from p153 (Lu et al., 1997) was inserted into the corresponding sites of pBCKS+ to create pBCHr3 (A-14). The *Pst*I-*Bam*HI fragment of pBCHr3 was inserted into the corresponding sites of pSLAct5CLuc to create pSLHr3Act5CLuc (A-15). The *Hind*III-*Sac*II fragment of pBCHr3 was inserted into the corresponding sites of pSLIE1Luc to create pSLHr3IE1Luc (A-16). The *Eco*RV-*Sac*II fragment of pBCHr3 was ligated to pSLpUbLuc digested with *Not*I, the site filled with Klenow fragment, and then digested with *Sac*II to create pSLHr3pUbLuc (A-17). phsp82*Renilla*Luc (Mohammed, unpublished) was created by inserting a 1-kb *Kpn*I-*Bam*HI fragment from pKhsp82 (Coates et al., 1996) into the corresponding sites of pBCKS+ and then inserting the *Kpn*I-*Pst*I fragment from this plasmid into the corresponding sites of pRL-SV40. ppUb*Renilla*Luc (A-18) was created by first digesting pSLpUbLuc with *Not*I, filling in the site with Klenow fragment, then digesting with *Pst*I to produce a 2-kb fragment which was ligated to pRL-SV40 prepared by digestion with *Bgl*II, the site filled with Klenow fragment and then digested with *Pst*I.

Table 2
Primers for section 2

Identifier	Sequence
*copia*ULRForw	5’ aagcttgggcccagtccatgccta 3’
*copia*ULRRev	5’ ccgcggattacgtttagccttgtc 3’
SV40Forw	5’ aagcttctgaggcggaaagaacca 3’
SV40Rev	5’ ccgcggaaaattagccagccatgg 3’

2.2.2. Cell cultures and transfections

Aedes albopictus C7-10 cells were maintained at 25ºC with 5% CO_2 in Eagle's media plus 5% fetal calf serum with the following additions per liter: 10 mL 10% (wt/vol) D(+)glucose, 10 mL 200mM L-glutamine, 10 mL MEM vitamin solution, 20 mL MEM non-essential amino acids, 10 mL Penicillin/Streptomycin (10,000 U/mL), 29.3 mL sodium bicarbonate (7.5% w/v) (Shih et al., 1998). 400 µL of cells at a density of 2 x 10^6 cells/mL were seeded into 24-well microtiter plates and incubated at 25ºC for 24 hrs. Cells were transfected with 0.4 µg total DNA and 0.8 µL LipofectAMINE 2000 (Invitrogen, Carlsbad, CA 92008) in 10 µL serum-free, antibiotic-free media. phsp82*Renilla*Luc and the firefly constructs were transfected at a 1:2 ratio. The IE1 transactivator plasmid (Lu et al., 1997) was present as 1/10 of the total DNA amount.

2.2.3. Luciferase assays

Transfected cells were assayed 24 hrs. post-transfection using a Turner Designs 20/20 luminometer and a Dual Luciferase Assay (Promega). The manufacturer's passive lysis protocol was followed. In addition, cell lysates were snap-frozen in liquid nitrogen immediately after lysis to minimize luciferase protein degradation. All samples were diluted 20-fold in 1X PLB (passive lysis buffer) in order to obtain a reading within the range of the luminometer.

2.3. Results

2.3.1. The Hr3 enhancer and the IE1 trans*activator increase reporter gene activity in transiently transfected C7-10* Ae. albopictus *cells*

In transiently transfected C7-10 cells, the Act5C promoter resulted in the highest luciferase reporter activity in comparison with the remaining promoters alone (Fig. 5A). The level of measured activity directly corresponds to the amount of luciferase protein expressed by the transfected cells and thus presumably the level of transcription. Among the enhancers, Hr3 improved luciferase expression by 4-fold, 47-fold and 22-fold over the basal level expression from the Act5C, IE1 and pUb promoters respectively; cULR improved luciferase expression over basal level from the Act5C, IE1 and pUb promoters 2-fold, 11-fold and 10-fold respectively; and eSV40 resulted in 2-fold, 8-fold and 7-fold increases respectively in luciferase expression from the Act5C, IE1 and pUb promoters (Fig. 5A). Addition of the IE1 *trans*activator unexpectedly resulted in large increases in expression from the *Renilla* control plasmid, as well as from the Hr3-containing reporter plasmids. This is seen in the apparent drop of expression indicated by the red-shaded bars in Fig. 5A. This was confirmed in several independent experiments and was seen even with decreased concentrations of the *Renilla* plasmid (Fig. 6). In order to determine the relative effect of the IE1 transactivator on expression from the Hr3/promoter constructs, the raw firefly luciferase values were converted to a percentage of average pSLIE1Luc expression and plotted on a log scale (Fig. 5B). Firefly luciferase expression increased 50 to 200-fold over the basal level expression of all of the promoters with the addition of the IE1 *trans*activator.

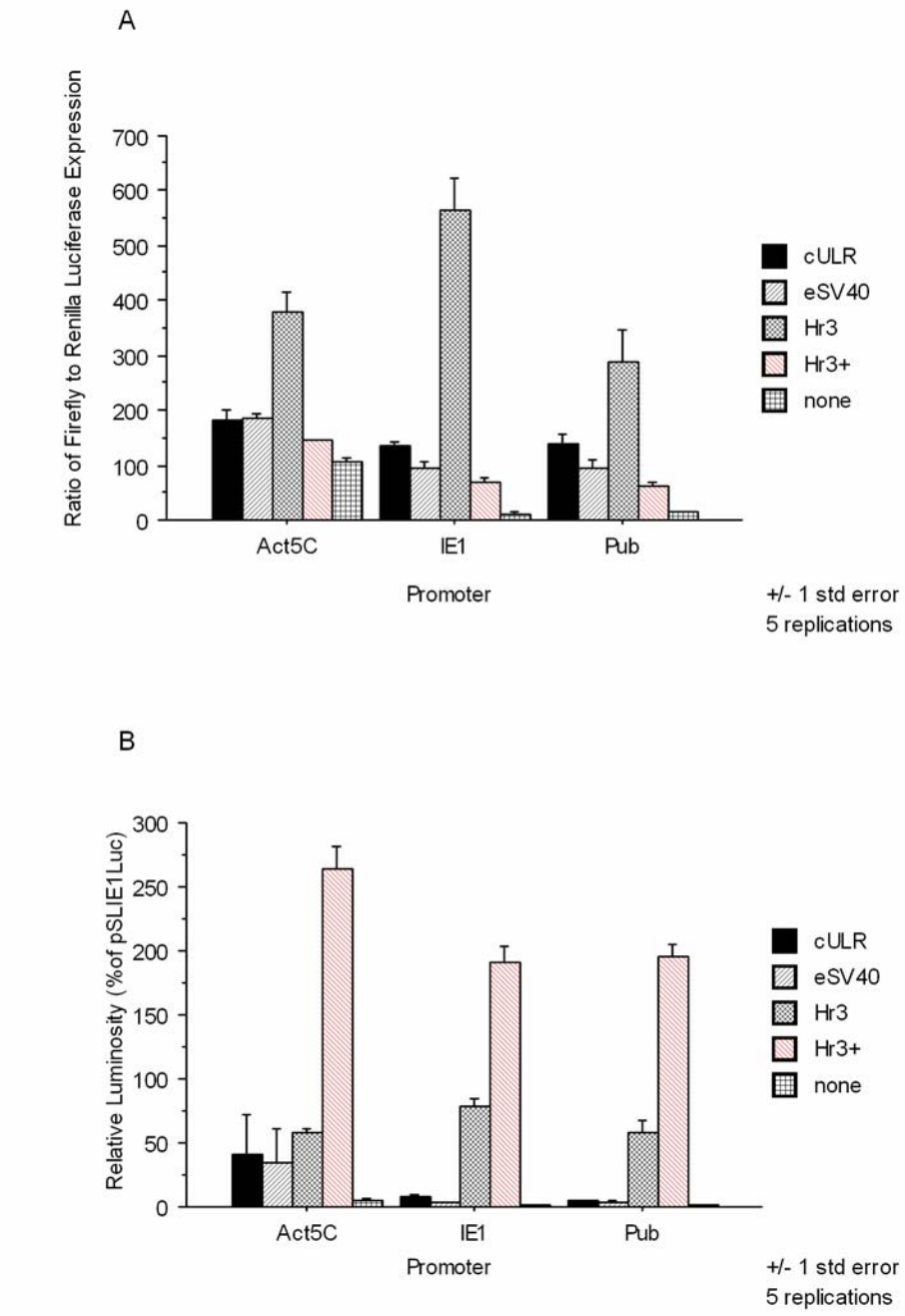


Fig. 5. Firefly luciferase expression from various promoter/enhancer plasmids in *Aedes albopictus* C7-10 cells. Cells were assayed for luciferase expression 24 hrs. post-transfection. The averages of five replications are reported and error is reported as +/- 1 standard error. Bars in red indicate the presence of the IE1 *trans*activator. (A) To normalize for differences in transfection efficiency and cell cycle state within the experiment, the firefly luciferase luminescence values for each construct were divided by the corresponding *Renilla* luciferase luminescence values measured in a dual luciferase assay. The Hr3 enhancer clearly outperforms both the cULR and the eSV40 enhancers in combination all promoters. (B) Raw firefly luciferase values are reported as a % of sSLIE1Luc expression on a log scale. Addition of the IE1 *trans*activating protein (Hr3+) increased firefly luciferase expression 2.5-4-fold over all Hr3-promoter combinations alone and 50 to 200-fold over basal promoter expression.

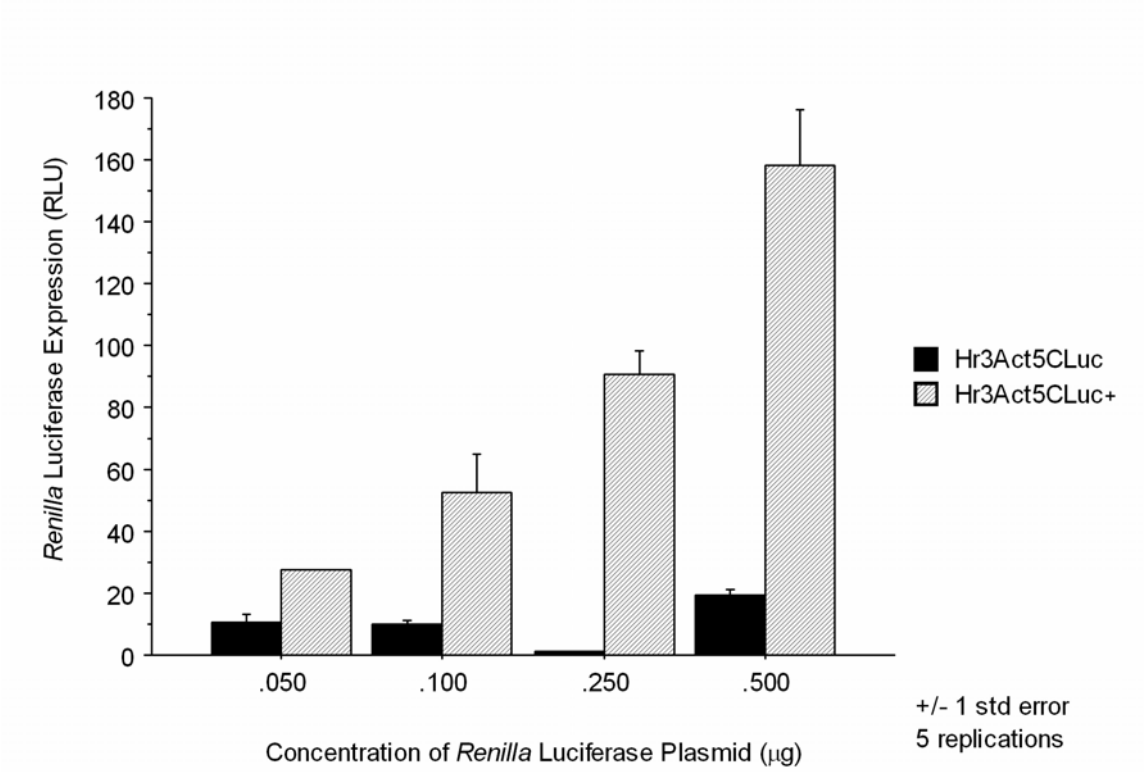


Fig. 6. The IE1 *trans*activator significantly affects expression from the *Renilla* luciferase control plasmid. Cells were transfected as detailed in the methods with 0.27 µg Hr3Act5CLuc (firefly) plasmid, 0.04µg IE1 *trans*activator plasmid and the indicated amount of *Renilla* luciferase control plasmid. Because this is a control plasmid with no Hr3 enhancer element present, one would expect the expression levels to parallel those shown in the absence of the IE1 *trans*activator (solid bars).

2.3.2. IE1 trans*activator interacts with promoter sequences in addition to the Hr3 enhancer*

Analysis of multiple experiments (Fig. 7, Table 3, and data not shown) revealed an interesting trend regarding the effect of the IE1 *trans*activating protein upon the promoters themselves. This effect was different for each promoter when co-transfected with an identical plasmid expressing the *Renilla* luciferase control. Addition of the IE1 *trans*activator resulted in a 17-fold increase in expression of firefly luciferase from the

Act5C promoter over its basal level expression and a concomitant 30-fold increase from the *Renilla* luciferase reporter under the control of the hsp82 promoter from *Drosophila pseudobscura* (Blackman and Meselson, 1986). When the IE1 promoter was used to drive expression of firefly luciferase, expression increased 169-fold over the basal level expression, while *Renilla* luciferase expression from the hsp82 promoter increased 138-fold. Finally, firefly luciferase expression increased 11-fold relative to basal level expression from the pUb promoter with a corresponding 202-fold increase in *Renilla* luciferase expression from the hsp82 promoter.

Table 3
Change in basal luciferase expression from promoters with the addition of the IE1 *trans*activator

Promoter	Firefly Luciferase	*Renilla* Luciferase	Ratio
Act5C	↑ 17 x	↑ 30 x	↓ 1.9 x
IE1	↑ 169 x	↑ 138 x	↑ 1.2 x
pUb	↑ 11 x	↑ 202 x	↓ 18.3 x

This summary of the data presented in Fig. 5 shows the fold change of firefly luciferase expression from each basal promoter following addition of the IE1 *trans*activator, the fold-increase in expression from the control *Renilla* luciferase plasmid under control of the hsp82 promoter and the overall change in ratio.

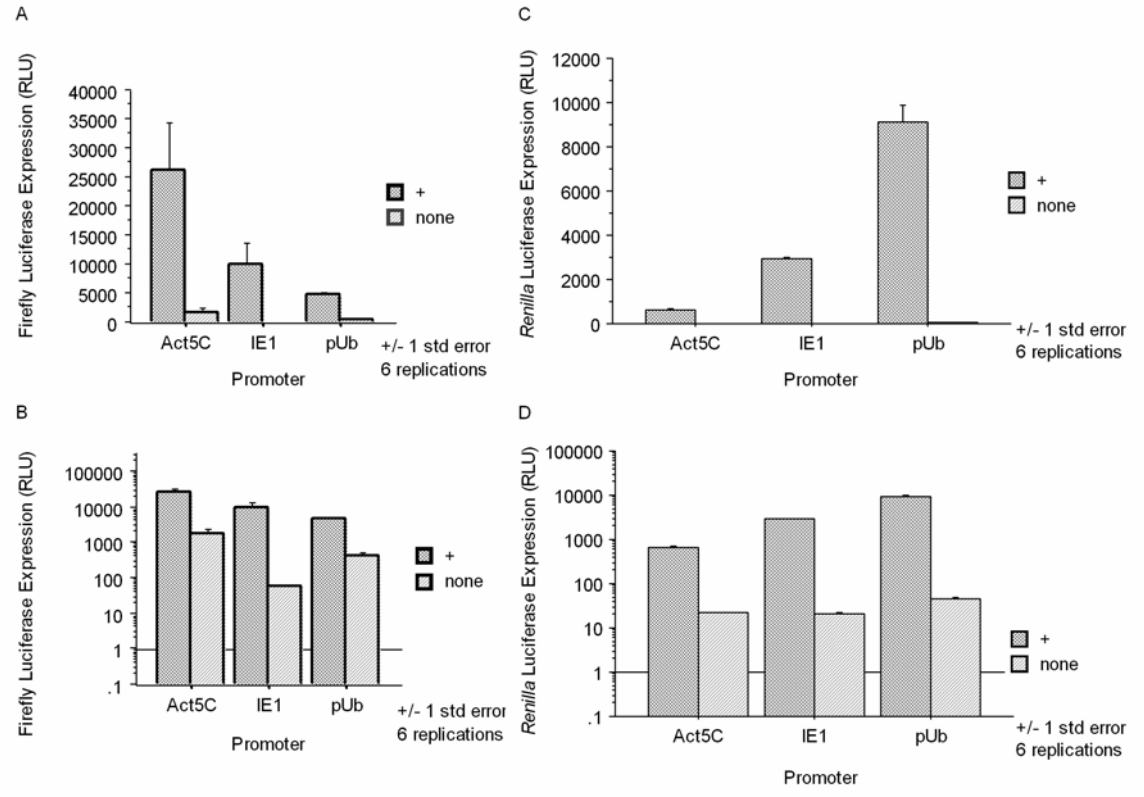


Fig. 7. Differential effect of the IE1 *trans*activator on the transcription levels from various promoters. Each promoter-Luc construct was co-transfected with phsp82*Renilla*Luc, both in the absence and presence of the IE1 *trans*activator, and assayed for both firefly and *Renilla* luciferase expression, 24 hrs. post-transfection. One experiment with six replicates was performed with the same batch of cells, DNA/liposome complexes and luciferase reagents. Error is reported as +/- 1 standard error. Each set of data is plotted both on a linear and a log scale. (A) Addition of the *trans*activator (+) caused a 16-fold increase in firefly luciferase expression from the Act5C promoter, a 169-fold increase in expression from the IE1 promoter and an 11-fold increase in expression from the pUb promoter. (B) The same data as shown in (A) but plotted on a log scale. (C) Addition of the *trans*activator resulted in different levels of expression from the phsp82*Renilla*Luc construct depending upon which promoter was used to drive expression of the firefly luciferase construct. (D) The same data as shown in (C) but plotted on a log scale.

2.4. Discussion

2.4.1. The IE1 trans*activator interacts with the* Renilla *luciferase promoter*

Unexpectedly, the internal control for transfection and protein recovery, *Renilla* luciferase, could not reliably be used as such in the presence of the *trans*activator. The data presented reveal a differential effect of the IE1 *trans*activator (Fig. 7 and Table 3)

that profoundly affects expression levels from the two luciferase plasmids in an enhancer/promoter-dependent manner. This compromises the ability to compare expression values both within an experiment where IE1 is present in some samples but not in others and between experiments where different batches of cells and assay reagents are employed. The results of a single experiment involving the *trans*activator are reported here, however additional experiments show similar results. The ratio of firefly to *Renilla* luciferase is reported for all promoter/enhancer combinations to allow accurate comparison of the three promoters alone and in combination with each of the three enhancers. It should be noted that the addition of the transactivator does significantly increase firefly luciferase expression from all three promoters with the Hr3 enhancer sequence, though this is masked by the simultaneous increase in expression from the *Renilla* luciferase control plasmid.

2.4.2. The IE1 trans*activator differentially interacts with exogenous promoters independently of the Hr3 enhancer*

The IE1 *trans*activator is clearly interacting with the promoters in *trans*, even in the absence of the Hr3 enhancer element (Fig. 7, Table 3 and data not shown). This observation agrees with previously published data that the cytoplasmic A3 actin gene promoter of *B. mori* was upregulated as much as one hundred-fold by the co-transfection of a plasmid encoding the *B. mori* IE1 gene product (BmIE1) (Lu et al., 1996). When the Hr3 enhancer is present, there is a cooperative effect, and luciferase expression increases as much as 200-fold (Fig. 5) over that of the promoter alone. This cooperativity is consistent with results obtained with Hr3-enhanced CAT expression

cassettes driven by the *B. mori* cytoplasmic actin gene promoter co-expressed with the BmIE1 protein in lepidopteran cell lines *Bm5* and *Sf*21 (an increase of up to three orders of magnitude) (Lu et al., 1996).

The significant differences seen in expression from each of the promoters tested (Fig. 7 and Table 3) reveal that not all promoters are affected in the same manner, nor is the co-transfected plasmid. The presence of the Hr3 enhancer region upstream of the promoter driving expression of the IE1 *trans*activator protein, results in high levels of IE1 protein from a relatively low amount of plasmid DNA. Despite this abundance of IE1 protein, it appears that transcription from the pUb promoter, in the absence of the Hr3 enhancer, increases only 11-fold (Fig. 7A and Table 3), while transcription from the hsp82 promoter driving *Renilla* luciferase expression is exceptionally high (Fig. 7B and Table 3). The simplest explanation is that the IE1 protein has different affinities for binding sites on the various promoters and/or the IE1 protein is sequestering necessary basal transcription factors. It has also been observed that some viral promoters, IE-0, IE-2 and PE-38, are inhibited by IE1 expression (Carson et al., 1991; Kovacs et al., 1991; Leisy et al., 1997). Clearly, the actions of the IE1 *trans*activator in this study are consistent with its ability to bind Hrs (Guarino and Dong, 1991, 1994; Rodems and Friesen, 1995). In addition, the protein has two independent functional acidic activation domains and two potential positively-charged inhibitory domains (Choi and Guarino, 1995; Slack and Blissard, 1997), consistent with its observed ability to both enhance and inhibit expression from different promoters. Also, lower concentrations of the plasmid bearing the IE1 gene sequence in these transient assays result in greater increases in

luciferase expression (Fig. 8). This observation is consistent with the mechanism of negative regulation by the IE1 protein previously proposed (Leisy et al., 1997) where the cooperative binding of the Hrs occurs at a lower concentration than that required for binding to the half sequence regions (Hs) present in negatively regulated promoters. It is also consistent with the presence of the Hr3 enhancer sequence on the plasmid producing the IE1 *trans*activating protein, which results in up-regulation of IE1 transcription, consequently reducing the number of plasmid copies needed to produce optimal protein levels. When this experiment was repeated using pUb to drive *Renilla* luciferase expression (data not shown and Fig. 9), significant differences between promoters were also observed, though not the same differences described above with the hsp82-*Renilla* expression plasmid. Finally, it should be noted that each of the enhancers alone also differentially affected the expression from each promoter. These data collectively highlight the value of evaluating the effects of new promoter/enhancer/*trans*activator combinations on the expression of a reporter gene within a related cell line, prior to investing significant time and effort in the creation of transgenic lines. Though cell lines do not completely mimic the cellular and nuclear environment of an entire organism, they can yield significant insight into both the potential interaction between regulatory elements driving transgene expression and the potential impact of unknown endogenous *trans*acting factors.

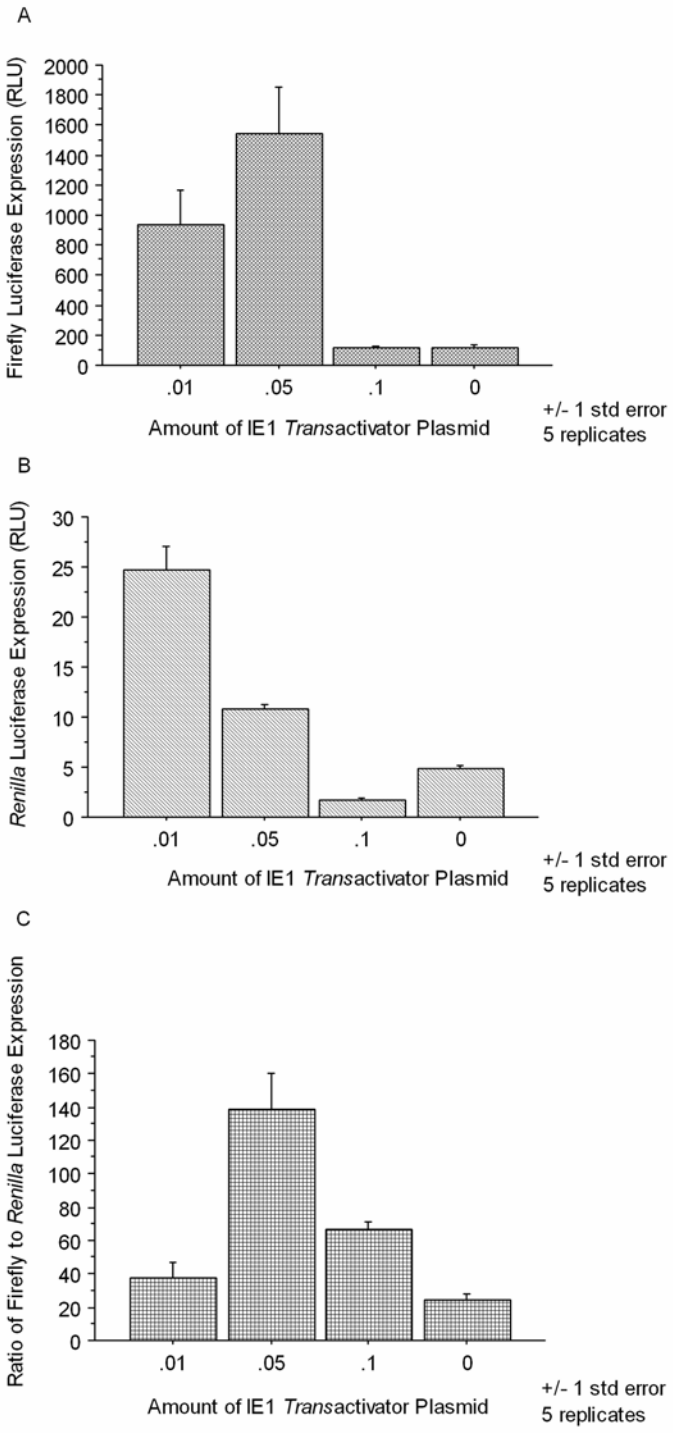


Fig. 8. Lower concentrations of the IE1 *trans*activator result in greater expression from both target and non-target promoters. Expression is indicated in relative light units (RLU) from (A) the Hr3Actin5C (firefly luciferase) and (B) the hsp82 (*Renilla* luciferase) promoters. (C) indicates the ratio of firefly to *Renilla* luciferase expression. Cells were transfected as detailed in the methods with 0.27 µg Hr3Act5CLuc (firefly) plasmid, 0.14 µg hsp82*Renilla*Luc plasmid and the indicated amount (µg) of IE1 *trans*activator plasmid.

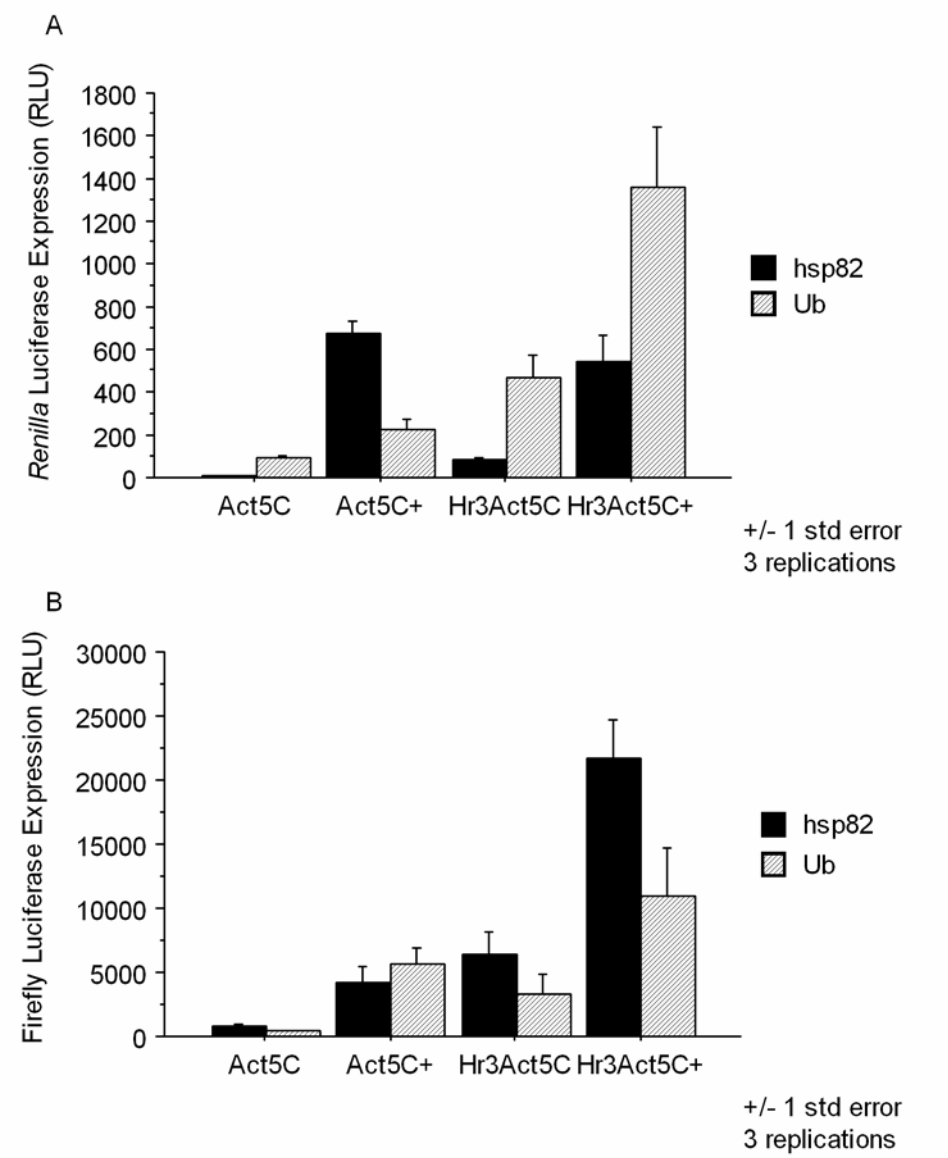


Fig. 9. The hsp82 promoter versus the pUb promoter for *Renilla* luciferase expression. Cells were transfected as detailed in the methods with the only difference being the promoter used to drive expression of the *Renilla* luciferase control plasmid. (A) The pUb promoter seems to be upregulated more than the hsp82 promoter in the presence of the *trans*activator. Interestingly, the corresponding expression from (B) the Hr3Act5C firefly luciferase plasmid is less when the pUb *Renilla* plasmid is co-transfected. Other differences were seen when this experiment was repeated with Hr3IE1 and Hr3pUb firefly luciferase plasmids. Clearly, the IE1 *trans*activator binds sequences other than the Hr3 enhancer sequence in eukaryotic promoters and the effect is dependent upon the combination of promoters present.

2.5. Conclusions

Clearly, we have shown that the baculovirus homologous region, Hr3, along with the IE1 transactivating protein, significantly increases transgene expression from each of the three heterologous, constitutive promoters tested in mosquito cells. Some concern does exist that endogenous promoters might be down-regulated by the presence of the IE1 protein, and that available host cell transcription factors might be sequestered by complexes stabilized by the IE1 transactivator, however a lower concentration of IE1 transactivator would likely mitigate these effects. Preliminary transposition assays confirm the ability of the Hr3 enhancer/IE1 transactivator combination to function in syncytial preblastoderm mosquito embryos and to significantly increase observed transposition frequencies when used to drive transposase expression (Coates, *et al*., unpublished data). Use of tissue-specific promoters/enhancers and/or inducible expression may effectively reduce any potential fitness load imposed by interactions of the IE1 protein with endogenous regulatory elements. Perhaps the most promising application is the use of HR/IE1 in helper plasmids transiently expressing transposase in an attempt to increase the number of stable transgene integration events by increasing the amount of available transposase, particularly if germ-line-specific promoters were used to express the transactivator protein. The HR/IE1 strategy is a promising tool for high-level transgene expression and/or increased transposition frequency in culicine mosquitoes and possibly other insect species as well.

3. CLONING AND CHARACTERIZATION OF cDNAS ENCODING PUTATIVE CTCFS IN THE MOSQUITOES, *Aedes aegypti* AND *Anopheles gambiae*

3.1. Introduction

CTCF (CCCTC-binding factor) was originally identified as a transcriptional repressor in studies of the chicken lysozyme silencer (Baniahmad et al., 1990) and the regulation of the chicken c-myc gene (Lobanenkov et al., 1990). Since that time, CTCF has been extensively characterized in vertebrates as a ubiquitously-expressed, highly-conserved, multivalent transcription factor that utilizes different zinc finger (ZF) combinations to specifically bind to diverse nucleotide sequences, resulting in the repression or activation of target genes, creation of hormone-responsive silencers and in the formation of enhancer-blocking boundary elements (reviewed by Ohlsson et al., 2001). Multiple, independent studies have established vertebrate CTCF as a central player in the regulation of gene expression via its association with every known vertebrate insulator (Ohlsson et al., 2001; West et al., 2002; Filippova et al., 2001). Further characterization of these proteins revealed their function to be: (a) constitutive to the insulation of the chicken β-globin 5'HS4 site (Bell et al., 1999; Recillas-Targa et al., 2002) and the 5' boundary of the human apolipoprotein B gene (Antes et al., 2001), (b) a role in imprinting via methylation-sensitive binding to the *Igfr2-H19* control locus (Bell et al., 1999; Hark et al., 2000; Ishihara and Sasaki, 2002; Fedoriw et al., 2003; Du et al., 2003; Schoenherr et al., 2003; Pant et al., 2003), the *DM1* locus (Filippova et al., 2001) and the *DLK1/GTL2* locus (Wylie et al., 2000), and (c) as part of a more complex,

multipartite insulator regulated by ligand binding (Lutz et al., 2003). Most recently, CTCF-dependent insulators have been identified in transitional chromatin, with high levels of H3 acetylation and essentially no CpG methylation, between escape genes and inactivated genes on both mouse and human inactive X chromosomes (Filippova et al., 2005). Finally, *Tsix* and CTCF have been proposed to comprise a regulated epigenetic switch for X-inactivation in mammals (Chao et al., 2002). Clearly, CTCF plays a pivotal role at multiple levels of gene regulation and genome organization in vertebrate organisms.

Long thought to be exclusive to vertebrates, a CTCF orthologue was recently characterized in *Drosophila melanogaster* with domain structure, binding site specificity and transcriptional repressor activity similar to that of vertebrate CTCF (Moon et al., 2005). Significantly, these researchers also demonstrated that a known *Drosophila* insulator, *Fab8*, mediates enhancer-blocking via CTCF in both *Drosophila* and vertebrate cell lines. We have cloned and characterized two mosquito CTCF-like cDNAs encoding polypeptides with significant similarity and insulator binding properties to both the vertebrate and *Drosophila* CTCFs. Analysis of available genome sequence from numerous invertebrate species yields promising candidates for additional CTCF orthologues. Clearly, this versatile protein has more ancient roots than once thought.

3.2. Materials and Methods

3.2.1. Isolation of RNA and preparation of cDNA by reverse-transcriptase PCR

Total RNA was isolated from ~30 mg each of *Ae. aegypti*, *An. gambiae* and *D. melanogaster* larvae using the RNeasy® Mini Kit (Qiagen, Valencia, CA 91355), followed by DNase I-treatment with DNA-free™ (Ambion, Austin, TX 78744) and was used to synthesize first strand cDNA using the SuperScript II™ reverse transcriptase (Invitrogen, Carlsbad, CA 92008) following the manufacturer's instructions. Briefly, 0.2 μM c-anchor dT-primer (Table 3), 1.03 μg total RNA, 1 μl of 10 mM dNTPs were brought to a total volume of 12 μl with ddH_2O, heated to 65 ºC for 5 minutes and quick-chilled on ice. After a brief centrifugation, 5X 1st strand buffer, 10 mM DTT and 40 units of RNase OUT™ Recombinant Ribonuclease Inhibitor (Invitrogen) were added and the mixture incubated at 42 ºC for 2 minutes. After addition of 200 units of SuperScript II™ (Invitrogen), the reaction was incubated for 50 minutes at 42 ºC followed by heat inactivation (70 ºC for 15 minutes). The product was digested with RNase H (Promega, Madison, WI 53711) to produce single-stranded cDNA and stored at -20 ºC. In order to increase the efficiency of the reverse-transcription reaction, 150 ng/μL of T4 Gene 32 Protein (Villalva et al., 2001) was added to the 1st strand buffer.

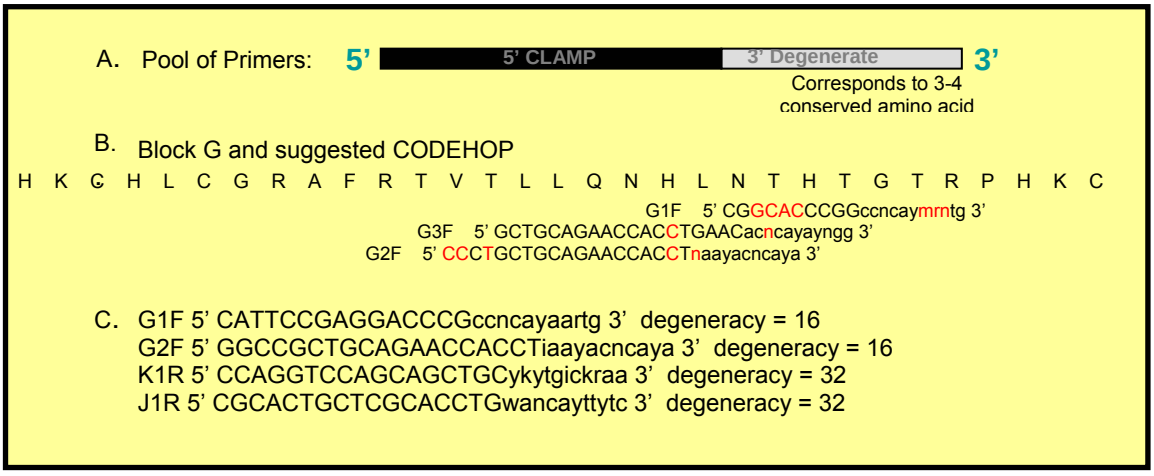


Fig. 10. Degenerate PCR primer design. (A) Schematic of CODEHOP primers. (B) Block G and suggested CODEHOP primers before being edited (bases in red). (C) Four CODEHOP primers used in this study along with their degeneracies. The CLAMP is in upper case while the degenerate core is in lower case. Inosine was utilized where necessary to reduce the degeneracy of the primer. In some cases, additional bases were added to the 5' CLAMP in order to increase the melting temperature to match a partner primer.

3.2.2. Isolation of Ae. aegypti *CTCF by degenerate PCR amplification*

The amino acid sequences of all known and predicted CTCFs (EAA11339.1, AAL78208, AAG40852, NP_031820, NP_114012, P49711 and Q08705) were identified using the BLAST search algorithm at the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov) and aligned using the ClustalW algorithm in the Vector NTI™ Suite (InforMax, Inc., 1999). Two completely nested and degenerate PCR primer pairs (Fig. 10, Table 4) were designed to a highly-conserved 168 amino acid region using CODEHOP (Rose et al., 1998; Rose et al., 2003). A 504 base pair nested PCR product was obtained from *Ae. aegypti* larval cDNA using G-1F and K-1R primers in the first PCR reaction, followed by a nested reaction with primers G-2F and J-1R. Each reaction was performed with 2 mM $MgCl_2$, 0.2 µM each primer, 10 mM dNTPs, 0.5 µL cDNA or 1st reaction product and 2.5 units of *Taq* polymerase

(Continental Lab Products, San Diego, CA 92111). The following touchdown PCR conditions were used: 96 ºC for 4'; 2 cycles of 96 ºC for 20", 72 ºC for 1'; 11 cycles of 96 ºC for 20", 71 ºC –1.0 ºC/cycle for 15", 72 ºC for 45"; 25 cycles of 96 ºC for 15", 59 ºC for 15", 72 ºC for 45"; final extension at 72 ºC for 2'. The product was visualized on a 1% agarose gel, extracted from the gel, cloned into a pGEM-T (Promega) vector and its sequence determined on an ABI 3100 capillary sequencer.

Table 4
Primers for section 3

Identifier	Sequence
c-anchor	5' gctaatacgtacgatcggtcgacaagttttttttttttttttv 3'
G-1F	5' cattccgaggacccgccncayaartg 3'
G-2F	5' ggccgctgcagaaccacctiaayacncaya 3'
J-1R	5' cgcactgctcgcacctgwancayttytc 3'
K-1R	5' ccaggtccagcagctgcykytgickraa 3'
AgaCTCFfor	5' caaacgccatatggaggacgtggagctgatat 3'
AgaCTCFrev	5' attacctcttgcggccgcttccgtggagaggataaact 3'
AgaCTCFseq-a	5' gccacctcaaaacgcactcc 3'
AgaCTCFseq-b	5' tgtccgcactgtacgtacgc 3'
AgaCTCFseq-c	5' tgcacgcgaaaacgcacgag 3'
AgaCTCFseq-d	5' agtcgacggtgagcaaggag 3'
AgaCTCFseq-e	5' gagtgcgaaaaaagaaccgg 3'
DmelCTCFfor	5' gaaggcatgattaatgccaaggaggacaaaaaag 3'
DmelCTCFrev	5' attacctcttgcggccgcggtcactagttgagcaag 3'
DmelCTCFseq-a	5' ccgaggatctgcagaccttc 3'
DmelCTCFseq-b	5' ccaccaataaatccatcaat 3'
DmelCTCFseq-c	5' ctgaggtctacgaatttgag 3'
DmelCTCFseq-d	5' aaagcgccttcaccaccagc 3'
DmelCTCFseq-e	5' acctgcgcgttcacattaaa 3'
DmelCTCFseq-f	5' gttgtccgcgggagtttacc 3'
DmelCTCFseq-g	5' gaaaagcagccgatgatcat 3'
DmelCTCFseq-h	5' ctcctgaaaacaaaagctaatggat 3'
AaeGSP1	5' gtctgtcttgcgcccacatgttg 3'
AaeGSP2	5' cgaaagcacgtttacaacttctgg 3'
AgaGSP1	5' ccacaggtcgtcgggcagagtttgca 3'
AgaGSP2	5' caatcggagtaagattgtccgaagaaaggtct 3'

Table 4 continued

Identifier	Sequence
AaeCTCFseq-a	5’ agtgtgaatgtgcgagcaat 3’
AaeCTCFseq-b	5’ gtgtggaagatgaattttgg 3’
AaeCTCFseq-c	5’ atgcttcgatttcggcgcgc 3’
AaeCTCFseq-d	5’ aagtgtcggcgatgcaacaa 3’
AaeCTCFseq-e	5’ tctatccgacgaaatgctta 3’
AaeCTCFSouthernForw	5’ actcactatagggcaagcagtggtatcaa 3’
AaeCTCFSouthernRev	5’ tgtgatttatcccgccagtggatt 3’
AaeRT-Forw	5’ gtgtttcattgcgagctttgcc 3’
AaeRT-Rev	5’ tgtctcgatcctccggaatg 3’
*S17*RT-Forw	5’ cgaagcccctgcgcaacaagat 3’
*S17*RT-Rev	5’ cagctgcttcaacatctccttg 3’
DmelRT-Forw	5’ atggagactcacgatgattcgg 3’
DmelRT-Rev	5’ ctcgtcgccattaaccagct 3’
*Rp49*RT-Forw	5’ gcgcaccaaggacttcatc 3’
*Rp49*RT-Rev	5’ gaccgactctgttgtcgatacc 3’
5’HS4Forw	5’ gagctcacggggacagcccccc 3’
5’HS4Rev	5’ aagctttttccccgtatccccc 3’
*Fab8*Forw	5’ ggcacaatcaagttaatgttgg 3’
*Fab8*Rev	5’ gcaagcgaagagttccattc 3’

3.2.3. PCR-amplification and cloning of An. gambiae *and* D. melanogaster *CTCFs*

The predicted ORF of the *An. gambiae* CTCF was PCR amplified from ~100 ng of cDNA with 0.2 μM each of the primers AgaCTCFforw and AgaCTCFrev (Table 4), 10 mM dNTPs and 2.5 units of Herculase® Hotstart DNA Polymerase (Stratagene, La Jolla, CA 92037) in a 50 μL reaction using the following conditions: 95 ºC for 2’; 5 cycles of 95 ºC for 30”, 55 ºC for 30”, 72 ºC for 2’45”; 25 cycles of 95 ºC for 30”, 65 ºC for 30”, 72 ºC for 2’45”; final extension at 72 ºC for 5’.

The *D. melanogaster* CTCF orthologue was PCR-amplified from ~100 ng of cDNA with 0.2 μM each of the primers DmelCTCFfor and DmelCTCFrev (Table 4), 10 mM dNTPs and 1 μL of Advantage2 Polymerase Mix (BD Biosciences Clontech, Palo

Alto, CA 94303) in a 50 µL reaction using the following conditions: 95 ºC for 4'; 5 cycles of 95 ºC for 30", 55 ºC for 30", 72 ºC for 3'; 25 cycles of 95 ºC for 30", 65 ºC for 30", 72 ºC for 3"; final extension at 72 ºC for 5'. An aliquot of each of the PCR products was visualized on a 1% agarose gel to confirm the correct size, and the remainder of the product was purified using the QIAquick® PCR Purification Kit (Qiagen). The purified PCR products were double-digested with *Nde*I and *Not*I restriction endonucleases (Promega) and cloned into the altered pET-30 vector described below. The sequence was determined on an ABI 3100 capillary sequencer with the primers AgaCTCFseq-a, b, c, d and e and the primers DmelCTCFseq-a, b, c, d, e, f, g and h (Table 4) for the resulting clones AgaCTCFpET (A-19) and DmelCTCFpET (A-20) respectively.

The pET-30 LIC/Xa Vector (Novagen, VWR International, Bristol, CT 06011) was altered by filling in the LIC sites with 2.5 units of Klenow Polymerase (Promega) and 10 mM dNTPs, followed by self-ligation with T4 DNA ligase (Promega). The resulting plasmid was propagated in electrocompetent DH10B cells, purified and the sequence of its multiple cloning site (MCS) confirmed. This vector was then double-digested with *Nde*I and *Not*I restriction endonucleases (Promega) and the 5' terminal phosphates removed with calf-intestinal phosphatase (CIAP) (Promega). This digested vector was used to clone both the *An. gambiae* and *D. melanogaster* CTCF orthologues.

3.2.4. Rapid amplification of cDNA ends (RACE) in Ae. aegypti *and* An. gambiae

Total RNA was prepared from freshly collected and snap-frozen larvae using the RNeasy® Mini Kit (Qiagen) and immediately DNase I-treated with DNA-free™

(Ambion) according to the manufacturers' instructions. The BD SMART™ RACE cDNA Amplification Kit (BD Biosciences Clontech) was then used to prepare first-strand cDNA and to amplify 5' and 3' RACE products according to the manufacturer's instructions. See Table 4 for gene-specific primers (GSPs) used for each species. GSP1 indicates the primer used for 5' RACE reactions while GSP2 indicates the primer used for 3' RACE reactions. Reaction conditions were as follows: 94 ºC for 5'; 5 cycles of 94 ºC for 10", 72 ºC for 3'; 5 cycles of 94 ºC for 10", 70 ºC for 10", 72 ºC for 3'; 25 cycles of 94 ºC for 10", 68 ºC for 10", 72 ºC for 3'; final extension at 72ºC for 8'. RACE products were visualized on a 1% agarose gel, gel purified, cloned into pGEM-T (Promega) and their DNA sequence determined using an ABI 3100 capillary sequencer with M13 (-20) and M13 Reverse primers followed by primer walking. All species-specific sequencing primers are listed in Table 4. The sequences were first edited to remove all vector sequences, assembled into complete sequence contigs and then translated using the Vector NTI™ Suite 8 (InforMax, Inc., 1999) software package. At least 3 different clones were analyzed for each RACE product.

3.2.5. Multiple sequence alignment and analysis

One vertebrate CTCF amino acid sequence (*Homo sapiens*, P49711) was aligned with the amino acid sequences of the three putative dipteran CTCF sequences from this study using the ClustalW algorithm in the Vector NTI™ Suite 8 (InforMax, Inc., 1999) package. Each of the sequences was independently submitted to each of the following databases in order to identify any potential conserved protein motifs: Conserved Domain Database (CDD) (Marchler-Bauer et al., 2003), Pfam (Protein families database

of alignments and HMMs) (Bateman et al., 2004), SMART (Simple Modular Architecture Research Tool) (Schultz et al., 1998; Letunic et al., 2004) and Interpro (Mulder et al., 2003).

3.2.6. Phylogenetic analysis

BLAST searches of the non-redundant databases at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST/), Ensembl (http://www.ensembl.org/), Flybase (http://bugbane.bio.indiana.edu:7151/blast/) and Baylor College of Medicine Human Genome Sequencing Center (http://www.hgsc.bcm.tmc.edu/) yielded additional putative orthologues. All putative orthologues were determined to possess the eleven highly-conserved C2H2 zinc finger domains in the central region of the protein as well as greatest similarity to known CTCFs when analyzed individually by BLAST at the NCBI website.

All putative orthologues were aligned using several algorithms: MultAlin (Corpet, 1988), DCA (Stoye et al., 1997), PRRN (Berger and Munson, 1991), T-Coffee (Notredame et al., 2000) and Poa (Lee et al., 2002). The resulting alignments were inspected and the best one chosen based upon alignment of the conserved residues within the eleven zinc finger domains. Both MultAlin and T-Coffee produced very good alignments though MultAlin introduced fewer gaps. Consequently, the MultAlin alignment, using the Blosum62 model with a gap opening penalty of 35, a gap extension penalty of 0.5 and no end gap penalty, was chosen for use in the construction of a phylogenetic tree for this gene. This alignment was refined by submitting it to RASCAL (Thompson et al., 2003) in order to remove any poorly aligned or divergent regions. The

refined alignment was submitted to Gblocks (Castresana, 2000) in order to identify and assemble blocks appropriate for phylogenetic analysis of this gene. The Gblocks output was converted to algorithm-specific formats using ReadSeq (http://bimas.dcrt.hih.gov/molbio/readseq/). Ultimately, the alignment was limited to the eleven zinc fingers plus five flanking amino acid residues.

The resulting alignment was analyzed using the Phylip software package (Felsenstein, 1989): bootstrapped (5000 replicates) with Seqboot, a distance matrix computed using Protdist (5000 datasets), the matrix submitted to Neighbor or Fitch (5000 trees), a consensus tree determined using Consense and the tree drawn using Drawgram. The MultAlin alignment was also submitted to Tree-Puzzle (Strimmer and von Haeseler, 1996) with 200,000 replicates and to BAMBE (Larget and Simon, 1999) with 200,000 cycles and 20,000 burn-in.

3.2.7. Preparation of genomic DNA

If mosquito DNA is to be used for PCR, the heads must be removed prior to homogenization due to an eye pigment that inhibits *Taq* DNA polymerase activity. 3 female and 3 male mosquitoes, or 5 females, or 7 males were ground in 80 µL of Bender buffer (0.1 M NaCl; 0.2 M Sucrose; 0.1 M Tris-Cl, pH 9.1; 0.05 M EDTA; 0.5% SDS) in a 1.7 mL sterile microtube with a Kontes pestle and Pellet Pestle® Motor (Kontes). After addition of 20 µL of 20 mg/mL Proteinase K (Promega), the tube was incubated at 50ºC overnight. The sample was then gently (so as not to shear the DNA) extracted twice with 200 µL Phenol/Chloroform/Isoamyl alcohol (25:24:1) and once with Chloroform/Isoamyl alcohol (24:1). The DNA was precipitated by addition of 4 µL of

3M NaOAc (pH 4.8) and 200 µL isopropanol, mixing gently for 5 minutes and centrifugation at 14,000 x g for 10 minutes. The pellet was washed with 70% ethanol, spun 5 minutes and air dried. The pellet was resuspended in 230 µL of nuclease-free water for digests or 100µL for PCR and stored at -20ºC. Alternatively, the protocol just described can be scaled up to extract genomic DNA from 175 to 200 mosquitoes by grinding in ~15 mL Bender buffer with a mortar and pestle, using 1 mL of proteinase K, 100 µL of 3 M NaOAc, pH 4.8 and 10 mL of isopropanol. The final pellet is resuspended in ~5 mL of nuclease-free water.

3.2.8. Southern blotting

Approximately 10µg of genomic DNA was completely digested overnight at 37 °C with 10 units of a restriction endonuclease (*Eco*RV, *Hind*III, *Sal*I or *Xho*I) (Promega). After phenol/chloroform extraction and ethanol precipitation, the digested DNA was size-fractionated on a 0.8% TAE agarose gel at 105 V for 1.25 hours, stained with ethidium bromide, destained with ddH_2O and imaged. The DNA in the gel was prepared for Southern transfer by: a.) depurination in 0.25 N HCl for 15-30 minutes, b.) denaturation in 0.5 N NaOH + 1.5 M NaCl for 30-60 minutes, c.) neutralization in 1 M Tris-HCl (pH 7.5) + 1.5 M NaCl for 30-60 minutes. A Stratagene PosiBlot® 30-30 Pressure Blotter (Stratagene) was used, according to the manufacturer's instructions, to transfer the DNA to a nylon membrane (Nytran +) overnight in 10X SSC buffer. The DNA was crosslinked to the membrane using a UV Stratalinker 2400 (Stratagene) on the auto-crosslink setting.

3.2.9. Probe preparation and hybridization

A 736 base pair fragment was PCR-amplified and labeled by incorporation of 800 Ci/mmol α-32-dCTP (Amersham Pharmacia Biotech, Piscataway, NJ 08855) from ~5 ng of 5'RACE clone #15 with the 0.2 μM each of the primers AaeCTCFSouthernForw and AaeCTCFSouthernRev (Table 4), 0.5 mM dNTPs (minus dCTP) and 0.5 μL of *Taq* DNA polymerase (Continental Lab Products) according to the following conditions: 94 °C for 5'; 30 cycles of 94 °C for 30", 60 °C for 30", 72 °C for 1'; 72 °C for 5'. The probe was cleaned up using a QIAquick® Nucleotide Removal Kit (Qiagen) according to the manufacturer's instructions. The blot was prehybridized for 2 hours at 68 °C in 12.5 mL of Church's buffer (1% Fraction V BSA, 1 mM EDTA, 0.5 M Na2HPO4—pH 7.2, 7% SDS), 625 μL 100X Denhardt's solution (2% w/v Ficoll 400, 2% w/v polyvinylpyrrolidone, 20 mg/mL Fraction V BSA) and 0.5% Non-Fat Dry Milk (NFDM). The prehydridization solution was replaced with fresh solution of the same composition, the probe was added and the probe was hybridized to the blot overnight at 68 °C. The blot was washed at 68 ºC for 30 minutes each: 2X SSC/0.1% SDS, 1X SSC/0.1% SDS, 0.3X SSC/0.1% SDS and exposed overnight on a phosphor screen. The screen was scanned on a STORM phosphorimager.

3.2.10. Reverse-transcriptase polymerase chain reaction (RT-PCR) analysis/ developmental profile in D. melanogaster *and* Ae. aegypti

Total RNA was prepared from freshly collected and snap-frozen samples using the RNeasy® Mini Kit (Qiagen) and immediately DNase I-treated with DNA-free™ (Ambion) according to the manufacturers' instructions. Sterile, nuclease-free,

disposable pestles were used along with a hand-homogenizer (Kontes) to grind the tissue finely. The concentration and 260/280 ratio of each RNA sample was determined using a spectrophotometer (Beckman Coulter, Fullerton, CA 92834). Reverse-transcriptase reactions were performed as described in 3.2.1.

Polymerase chain reactions (PCR) were assembled with 100 ng cDNA, 10X buffer, 1.5 µL 10 mM dNTPs, 0.2 µM each primer (Table 4) and 1 µL Advantage2 *Taq* Polymerase (BD Biosciences Clontech) in a total volume of 50 µL. Reaction conditions were as follows: 95 ºC for 5'; 20, 25 or 30 cycles of 95 ºC for 15", 55 ºC for 15", 72 ºC for 30"; final extension at 72 ºC for 2'. Products were resolved on a 2% agarose gel, stained with ethidium bromide, destained with ddH_2O and imaged. The constitutively expressed *D. melanogaster Rp49* gene (153 bp product) and *Ae. aegypti S17* gene were used as controls.

3.2.11. Generation of polyclonal antibody against An. gambiae *CTCF*

The coding sequence of a C-terminal region (amino acid residues 444-680) was PCR amplified and cloned into the pET-30 plasmid (Novagen) to create AgaC-termCTCFpET (A-21), expressed in *E. coli* (BL21-DE3) and His-tag purified on a Ni-NTA column (Novagen). The purified protein was used to immunize two New Zealand white rabbits following standard procedures at the Texas A&M University Laboratory Animal Research Resources (LARR) facility. Sera was separated from whole blood by centrifugation, aliquoted and stored at -70 ºC.

3.2.12. Immunoblotting

Bacterial pellets were resuspended and lysed in 1X disruption buffer (50mM Tris-HCl, pH 8.0; 4 % SDS; 4% β-mercapto-ethanol; 10% glycerol; 1 mg/mL bromophenol blue) by forcing the cells through a 25-gauge needle. The whole cell lysate was then sonicated at 30% output power for 3 cycles of 10 seconds each, with samples cooled on ice between sonication cycles.

Ten microliters of each sample were separated on a denaturing gel (8% separating gel and 4% stacking gel) in 1X Running Buffer (25 mM Tris-HCl, pH 8.0; 0.2 M glycine; 0.1% SDS) at 20 mA for ~2 hours. Kaleidoscope Prestained Standards (Bio-Rad, Hercules, CA 94547) were run in an adjacent lane and used to estimate the sizes of the resulting immunostained bands as well as to assess the transfer efficiency of the protein from the gel to the membrane. The proteins were transferred to Westran® Clear Signal membrane (Schleicher & Schuell BioScience, Inc., Keene, NH 03431) by electroblotting in 1X Towbin transfer buffer (25 mM Tris-HCl, pH 8.0; 0.2 M glycine; 0.1% SDS; 10% methanol) for one hour in a Hoefer™ tank transfer unit (Amersham Pharmacia Biotech) according to the manufacturer's instructions.

Upon completion of the protein transfer, the gel was washed twice for 10 minutes in 1X TBS buffer (10 mM Tris-HCl, pH 7.5; 150 mM NaCl). It was then washed in blocking buffer (1% non-fat dry milk (NFDM), 1% fraction V Bovine Serum Albumin (BSA), 1X TBS, 0.05% Tween-20) with 20% 5X casein (Novagen), in a sealed bag overnight at 4 ºC. The blot was then washed twice for 10 minutes in 1X TBSTT and once for 10 minutes in 1X TBS and was incubated for 1 hour at room temperature on an

orbital shaker with His-Tag Monoclonal Antibody (Novagen), resuspended in 600 µL of 1X TBS and then diluted 1:50 in blocking buffer for a final dilution of 1:10,000 or rabbit polyclonal antisera diluted 1:250 in blocking buffer. After antibody binding, the blot was washed twice in 1X TBSTT (1X TBS, 0.05% Tween-20, 0.2% Triton X-100) for 10 minutes and once in 1X TBS for 10 minutes. Goat Anti-Mouse IgG-AP conjugate (Novagen) was diluted 1:3,500 in blocking buffer and incubated with the blot for 1 hour at room temperature on an orbital shaker. The blot was then washed for 10 minutes five times in 1X TBSTT. Finally, it was developed for 1-10 minutes in Sigma-FAST™ (Sigma Aldrich Chemical Company, St. Louis, MO 63178) according to the manufacturer's instructions.

3.2.13. Electromobility Shift Assay (EMSA)

Sua4 cells were lysed in ice-cold lysis buffer (50mM Tris, pH 7.8; 150 mM NaCl; 1% IGEPAL CA360 (Sigma) with Complete Protease Inhibitor Cocktail (Roche, Indianapolis, IN 46250) and 1 mM PMSF. Total protein of the cell lysate was quantitated using the BCA Protein Assay (Pierce, Rockford, IL 61105), aliquoted and frozen at -20 ºC. The 5'HS4 and *Fab8* probes for EMSA were amplified and simultaneously labelled with $^{\alpha\text{-}32}$P (Amersham) in a 50 µL PCR reaction with ~10 ng DNA template, 10X reaction buffer, 2 mM $MgCl_2$, 8 µL of 0.5 mM 3dNTP mix (minus dCTP), 5 µL of 40 µM dCTP, 0.2 µM each primer (Table 4) and 2.5 µL of 10 mCi/mL (800 Ci/mmol) α^{32}dCTP. Amplifications were performed according to the following conditions: 95 ºC for 4'; 30 cycles of 95 ºC for 30", 55 ºC for 15", 72 ºC for 30"; final extension at 72 ºC for 5'. A scintillation counter was used to measure the specific

activity of the probe. This number was used to calculate the amount of probe produced in the reaction given that the reaction conditions produce products with ~10% of the "C" nucleotides labelled (Pollock, 1996). The binding reaction protocol was adapted from (Filippova et al., 1996). Approximately 10 fmol of labelled probe was incubated for 15 minutes on ice with 1, 5 or 10 μL of total cell lysate in binding buffer (1X PBS with 5 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 1mM DTT, 0.1% IGEPAL CA360 (Sigma), 10% glycerol) in the presence of a mixture of non-specific, cold, double-stranded competitor DNAs (500 ng polydI·polydC, 500 ng polydG·polydC, 500 ng SpI oligos, 500 ng Egr1 oligos). The SpI and Egr1 ds oligos contain strong, C/G-rich binding sites for the zinc-finger proteins SpI and Egr1 respectively. Sample 5 contained 150-fold excess unlabeled specific competitor. For the supershift, anti-sera against the *An. gambiae* CTCF was then added and the reactions incubated an additional 15 minutes on ice. Complexes were separated from the free probe on a 5% native PAGE gel in 0.5X TBE. The gel was run for 3.5 hours at 4 ºC at 10 V/cm.

3.3. Results

3.3.1. Cloning of Ae. aegypti *and* An. gambiae CTCF-*like cDNAs*

A BLAST search using the human CTCF protein sequence (Filippova et al., 1996) as a query uncovered a cDNA from *D. melanogaster* (AAL78208), subsequently characterized as an orthologous CTCF factor (Moon et al., 2005). This sequence was then used to query the *An. gambiae* genome assembly at the Ensembl database (http://www.ensembl.org/) resulting in a highly significant hit of the predicted novel gene ENSANGG00000015222 (e-139). These two dipteran sequences were aligned

with known vertebrate CTCF sequences, *Gallus gallus* (Klenova et al., 1993), *Mus musculus* and *Homo sapiens* (Filippova et al., 1996), *Rattus norvegicus* (NP_114012.1) and *Xenopus laevis* (Burke et al., 2002) using the ClustalW algorithm (Vector NTI™ Suite 8, InforMax, Inc., 1999). This alignment was used for degenerate PCR primer design. Degenerate PCR, using *Ae. aegypti* larval cDNA as a template, yielded a single PCR product of 504 base pairs, corresponding to a 168 amino acid polypeptide containing six of the eleven predicted zinc-finger domains (Fig. 11A). PCR amplification was initially performed with an *An. gambiae* larval cDNA template and primers corresponding to the 5' and 3' ends of the predicted novel coding sequence. This yielded a single product of 2040 base pairs, corresponding to a translated polypeptide of 680 amino acid residues (Fig. 12A). Subsequent 5' and 3' RACE (rapid amplification of cDNA ends) in both species yielded putative full-length cDNAs of 2616 and 4544 base pairs for *Ae. aegypti* (Fig. 11B and 11C) and *An. gambiae* (Fig. 12B and 12C) respectively. Alignment of the corresponding polypeptide sequences with both the *D. melanogaster* and *H. sapiens* CTCFs revealed significant differences in the N-terminal and C-terminal regions of the protein (Fig. 13), however there was 38% identity and 56% similarity across all eleven zinc finger domains (Fig. 14). Furthermore, 68% of the critical binding residues were conserved, despite at least 500 million years of divergence between invertebrate and vertebrate species (Peterson et al., 2004).

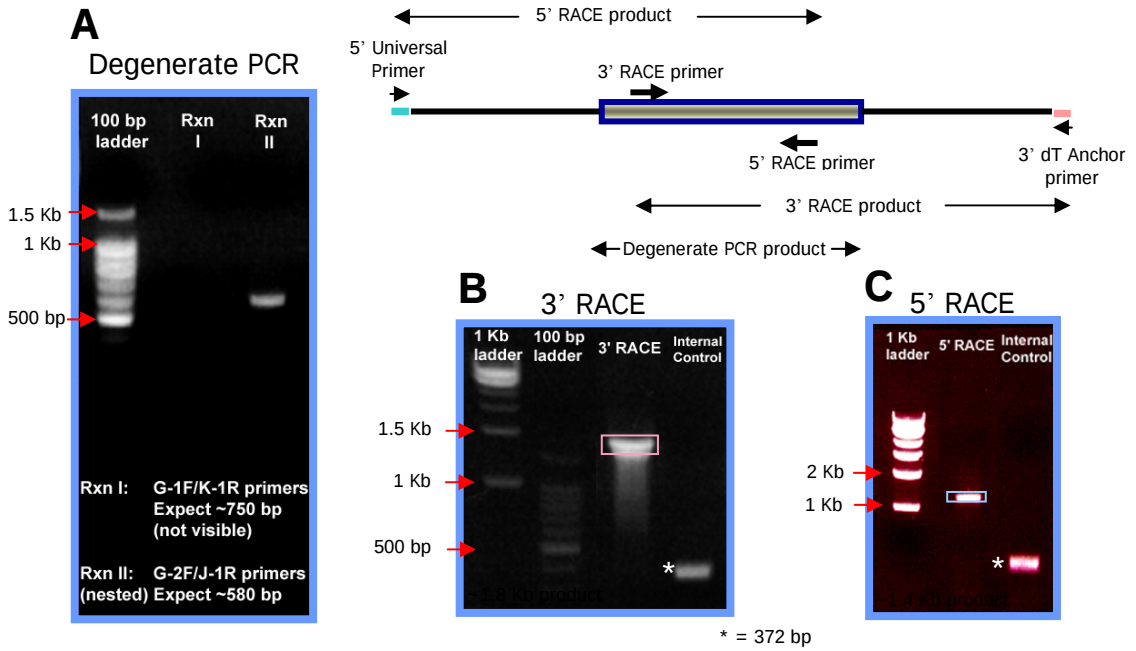


Fig. 11. Degenerate PCR and RACE products for *Ae. aegypti CTCF*. (A) Nested reaction yielded ~580 bp product that was cloned and sequenced. (B) 3' RACE product. (C) 5' RACE product. The * indicates the RACE internal control fragment.

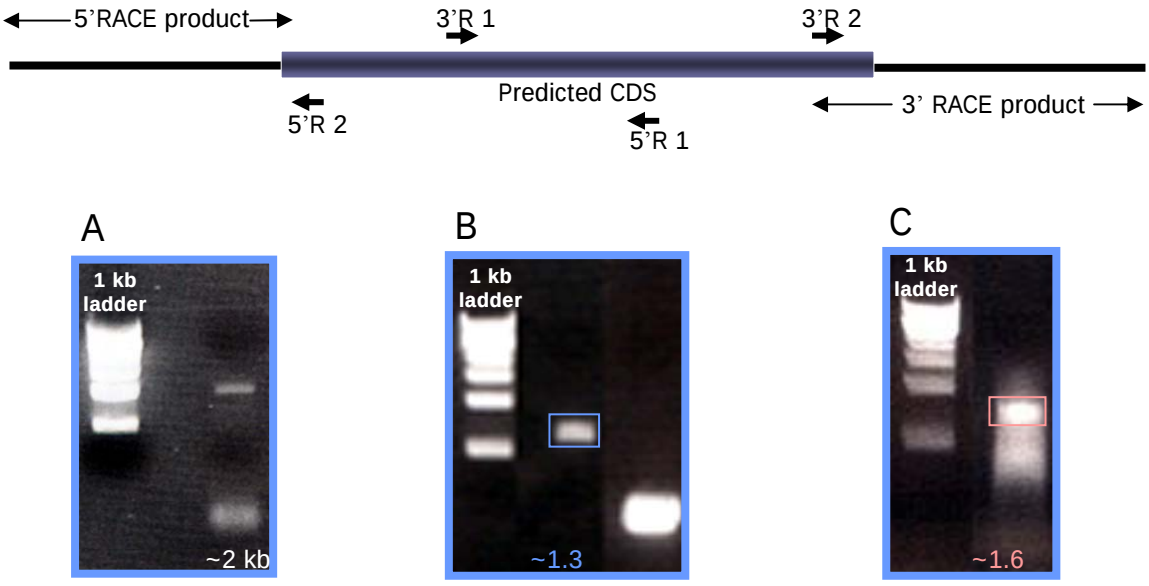


Fig. 12. Coding sequence and RACE products for *An. gambiae CTCF*. (A) ~2 Kb product using primers designed from predicted CTCF sequence. (B) 5' RACE product. The unboxed band is an internal PCR control amplified using 3'R1 and 5'R1 primers. (C) 3' RACE product.

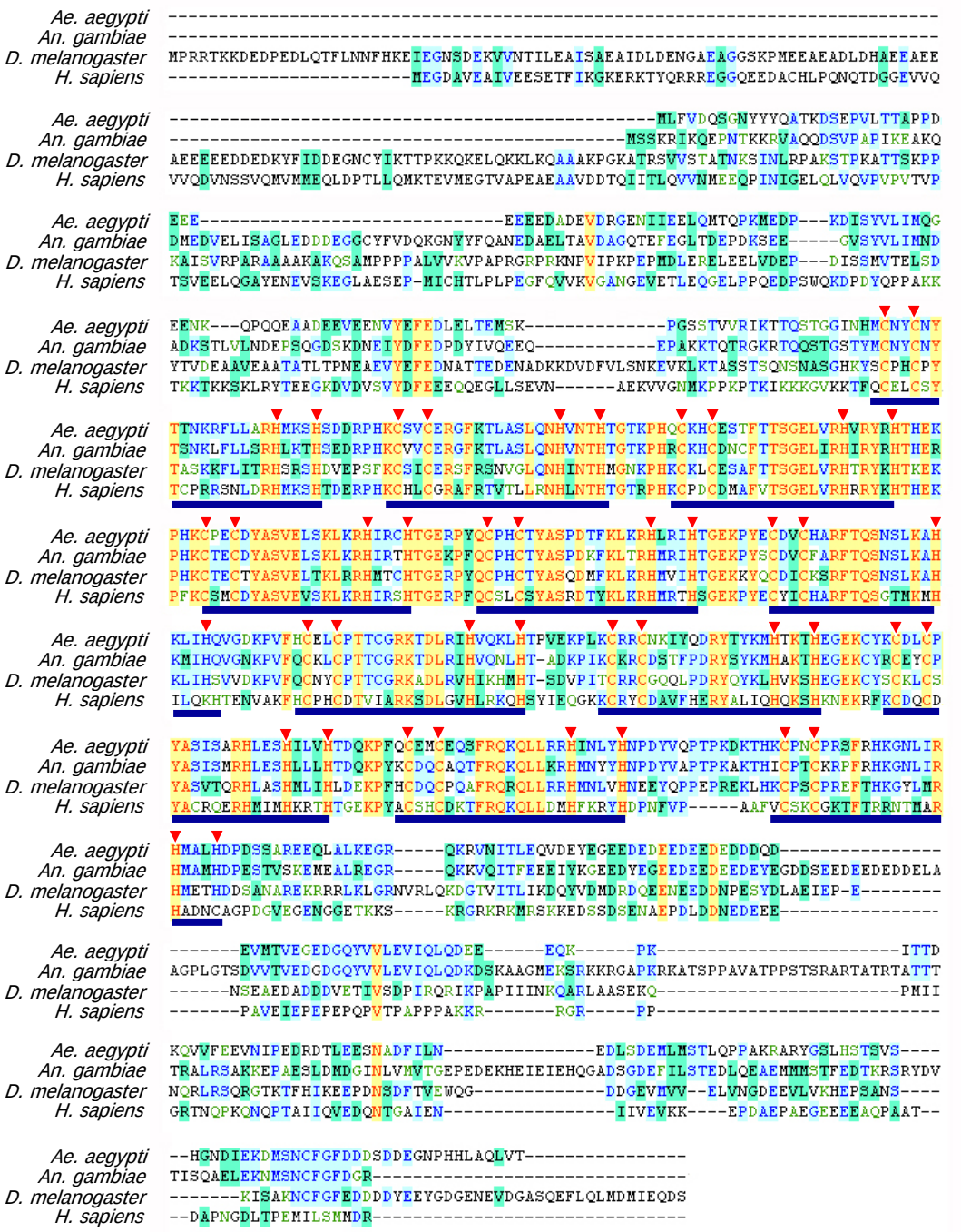


Fig. 13. ClustalW alignment of full-length, translated CTCF sequences for *Ae. aegypti*, *An. gambiae*, *D. melanogaster* and *H. sapiens*. Each of the eleven C2H2 zinc-finger domains is underlined, and the highly-conserved zinc-coordinating residues are indicated by the red arrowheads.

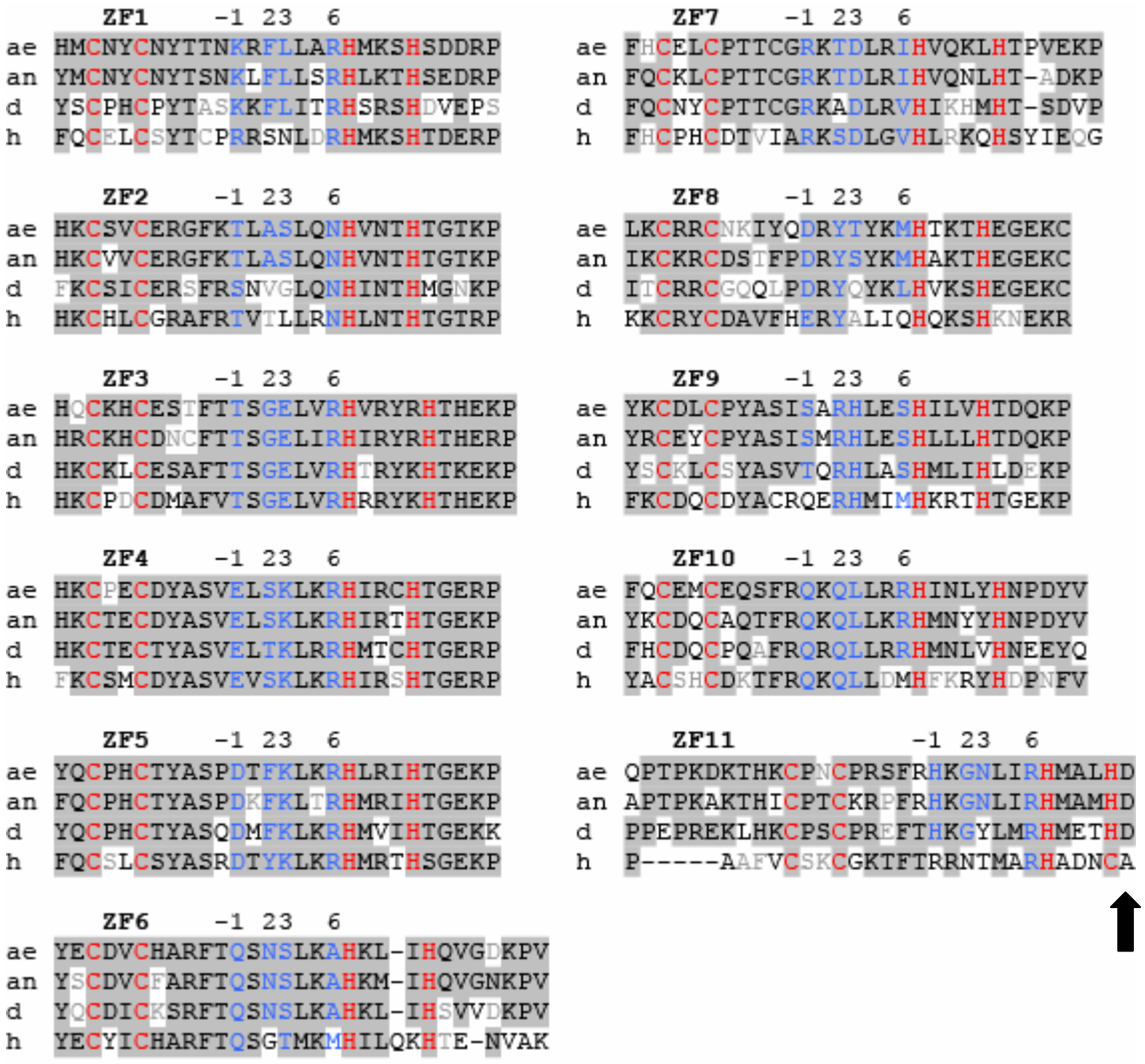


Fig. 14. The zinc-finger (ZF) domain is highly conserved between the dipteran insects, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*, and humans. Each of the eleven ZFs is aligned using the ClustalW algorithm. Identical and strongly conserved residues are highlighted in gray. Weakly conserved residues, the zinc-coordinating residues, and the amino acids with identical binding site recognition are indicated in gray, red and blue font respectively. Numbers indicate the position of each of the critical contact residues for DNA binding. The arrow denotes the C2H2 to C2HC change seen in vertebrate sequences.

3.3.2. CTCF *appears widespread in dipteran species*

Available genome sequence for multiple drosophilid species was queried at Flybase (http://bugbane.bio.indiana.edu:7151/blast/) using the *An. gambiae* amino acid

sequence and the tBLASTx algorithm. All species searched produced single hits of very high significance, ≤ e-126. Each of these was submitted as a BLASTp query of the non-redundant database at NCBI and confirmed to be a significant match to known CTCFs. Sequences with complete zinc finger regions were trimmed to the zinc-finger region plus five flanking amino acid residues and aligned with the corresponding region of CTCFs from *H. sapiens*, *G. gallus*, *X. laevis*, *Danio rerio* (NP_001001844), *Tetraodon nigroviridis* (CAF99566), and *Fugu rubripes* (Ensembl novel gene SINFRUG00000147322). The corresponding region of zinc finger protein 2 *Caenorhabditis elegans* (NP_500033) with 11 C2H2 zinc finger domains, a coil-coil region and predicted nuclear localization sequence was also included in the alignment and used as an outgroup in the subsequent phylogenetic analysis. Two consensus distance-based trees, Neighbor-Joining (Saitou and Nei, 1987) (Fig. 15A) and Fitch-Margoliash (Fitch and Margoliash, 1967) (Fig. 15B), were generated with 5000 bootstrap replicates using the Phylip software package (Felsenstein, 1989; Felsenstein, 1996). Additionally, a maximum-likelihood tree generated by 200,000 iterations of Tree-Puzzle (Strimmer and von Haeseler, 1996) (Fig. 15C) and a Bayesian analysis tree generated by 200,000 cycles of BAMBE (Larget and Simon, 1999) with 20,000 cycles of burn-in (Fig. 15D) yielded identical branch topologies.

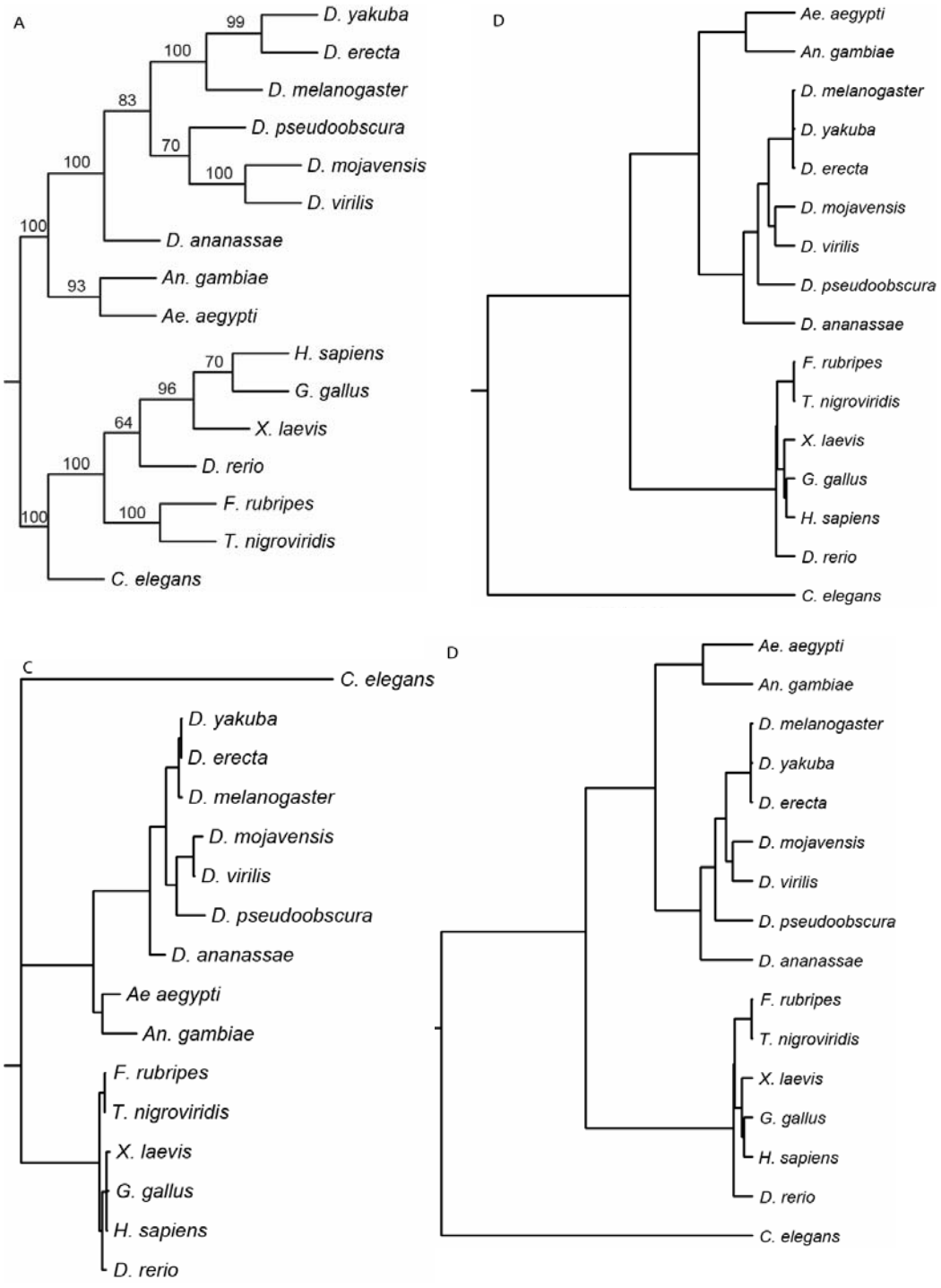


Fig. 15. Phylogenetic analysis of CTCF-like candidates in multiple species. Dendrograms consensus trees for an alignment of the 11 ZF region of known and predicted CTCFs. Four different trees were generated: (A) Neighbor-Joining (5000 bootstrap replicates), (B) Fitch-Margoliash (5000 bootstrap replicates), (C) Tree-Puzzle (200,000 iterations), (D) BAMBE (200,000 cycles). The tree topology is consistent regardless of the method used and agrees with the taxonomic classification of the *Drosophila* species.

3.3.3. CTCF *is a single copy gene in* Ae. aegypti

Available genome sequence data supports *CTCF* as a single copy gene within both *D. melanogaster* and *An. gambiae.* A Southern blot with *Ae. aegypti* genomic DNA (Fig. 16) was then performed to confirm a single copy locus in this mosquito. Ten micrograms of genomic DNA was digested with the restriction enzymes *Eco*RV, *Hind*III and *Xho*I, size-fractionated on a gel and probed with a radiolabeled 736 base pair fragment located upstream of the conserved zinc finger region. As expected, the *Eco*RV lane shows two distinct bands consistent with two fragments hybridizing to the probe. Partial sequence of the first intron reveals an *Eco*RV restriction enzyme site consistent with this pattern. The *Hind*III enzyme does not cut the region covered by the probe and produces the expected single band on the Southern blot, while the *Xho*I enzyme cuts the probed region once, and results in two bands. Subsequent release of a genome sequence with 5-fold coverage (http://tigrblast.tigr.org/er-blast/index.cgi?project=aabe) supports the conclusion that *CTCF* is a single-copy gene in *Ae. aegypti*.

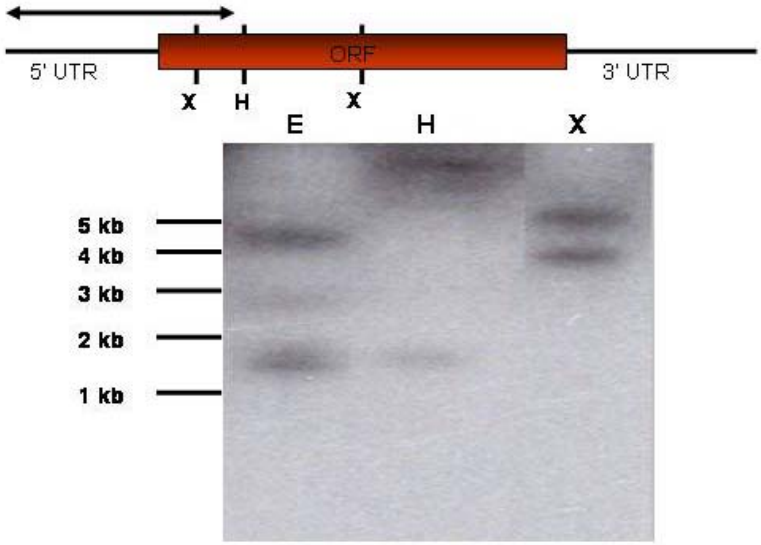


Fig. 16. Southern blot analysis confirms that *CTCF* is likely a single-copy gene in *Ae. aegypti*. Genomic DNA was digested with each of the following restriction enzymes, E = *Eco*RV, H = *Hind*III, X = *Xho*I, separated on a 1% agarose gel, transferred to a nylon membrane and probed with a 750 bp fragment (indicated by the double-headed arrow) corresponding to the 5' UTR and the coding sequence corresponding to the N-terminal region of the protein.

3.3.4. Mosquito CTCF is constitutively expressed in all developmental stages and is upregulated in early embryos and in the ovaries of blood-fed females

Reverse-transcriptase (RT)-PCR amplifications of RNA isolated from embryos, ovaries, larvae, pupae and adults shows CTCF expression across all stages of development and in ovarian tissues of both *Ae. aegypti* and *D. melanogaster* (Fig. 17). Early *Ae. aegypti* embryos and ovarian tissues from both species clearly show increased CTCF expression.

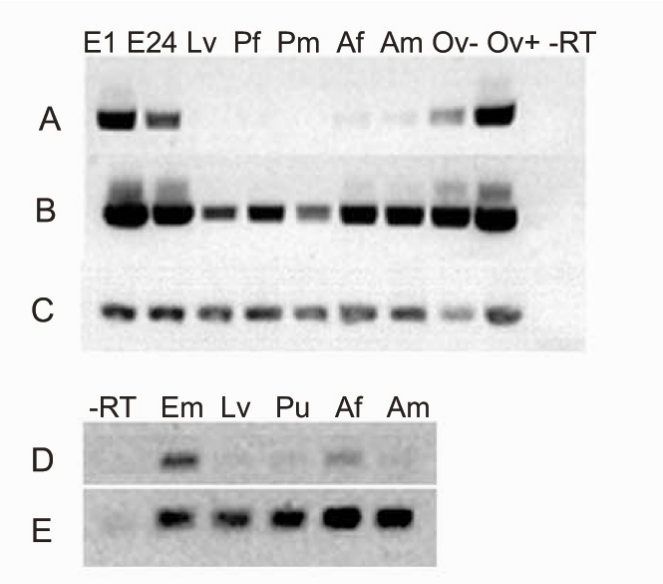


Fig. 17. Developmental expression profile of *CTCF* transcripts in *Ae. aegypti* and *D. melanogaster.* The expression of *CTCF* was analyzed using RNA isolated from multiple individuals at each of the indicated stages: E1 and E24 (embryos ≤ 1 hr and 24 hrs post-oviposition respectively), Lv (larvae), Pf (female pupae), Pm (male pupae), Pu (pupae), Af (adult females), Am (adult males), Ov- and Ov+ (ovaries from non-blood-fed and blood-fed females respectively). –RT, no reverse-transcriptase. (A) and (B) *Ae. aegypti CTCF*, 20 cycles and 30 cycles of PCR respectively. (C) *Ae. aegypti S17*, 20 cycles. (D) *D. melanogaster CTCF*, 25 cycles. (E) *D. melanogaster Rp49*, 20 cycles.

3.3.5. Specificity of the polyclonal rabbit antisera raised against a C-terminal fragment of An. gambiae *CTCF*

Immunoblot analysis of *E. coli* (BL21-DE3) cells expressing a His-tagged, C-terminal fragment of *An. gambiae* CTCF with both a His-tag monoclonal antibody (data not shown) and the rabbit polyclonal antisera raised against a C-terminal fragment (amino acid residues 443-638) of *An. gambiae* CTCF (Fig. 18) reveals a single band migrating at ~47 kD. The same band was identified in both total cell lysates (Fig. 18, lane 7) and in the His-purified fraction (Fig. 18, lane 9). Furthermore, total cell lysates from untransformed cells (Fig. 18, lane 7) show no specific binding to the CTCF antisera.

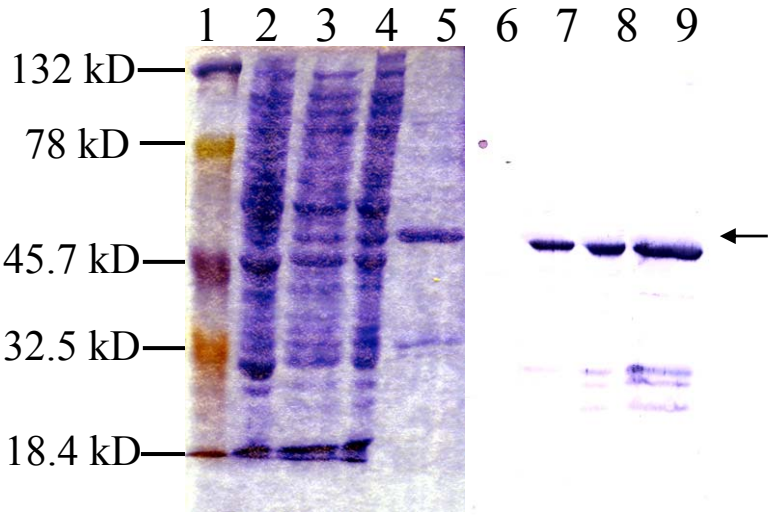


Fig. 18. Bacterial expression of the C-terminal fragment of *An. gambiae* CTCF. 1) size markers 2) untransformed cells 3) total cell protein 4) soluble fraction 5) His-tag purified 6) Western—untransformed cells 7) Western—total cell protein 8) Western—soluble fraction 9) Western—His-tag purified. Western anti body = polyclonal rabbit antisera diluted 1:250.

3.3.6. Mosquito CTCF binds in vitro *to both the chicken 5'HS4 and the* Drosophila Fab8 *insulators*

As we were unable to express the full-length protein in bacteria, whole cell lysates were prepared from the Sua4 (Muller et al., 1999) *An. gambiae* cell line and used in an electrophoretic mobility shift assay (EMSA) to assess whether mosquito CTCF can bind to known CTCF-associated insulator sequences (Fig. 19). The intensity of the shifted bands increased with application of greater amounts of protein lysate. The detectable complex was competed by cold, unlabeled probe, indicating that the binding was indeed specific. In addition, all reactions contained a 1200-fold excess of cold, non-specific C/G-rich sequences, further illustrating specific binding. Finally, the complex

could be partially shifted by polyclonal antibody sera generated against the C-terminal region of *An. gambiae* CTCF.

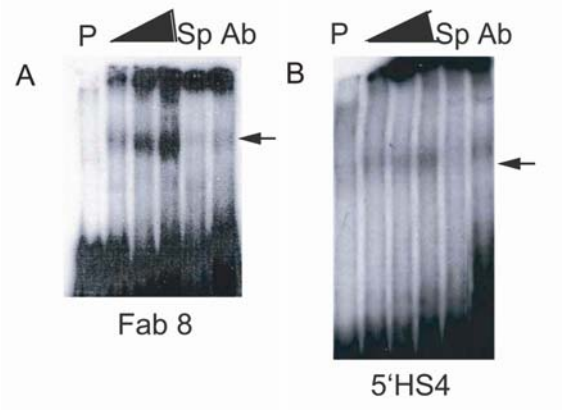


Fig. 19. *An. gambiae* CTCF specifically binds the chicken 5'HS4 and *Drosophila Fab8* insulator sequences. Sua4 cells were lysed and increasing amounts of total cell lysate (1, 5, 10 μL, represented as the solid triangle) were incubated with radiolabeled insulator sequences as follows: (A) *Drosophila Fab8* insulator sequence (Moon et al., 2005); (B) chicken *β-globin* FII insulator sequence (Bell et al., 1999). The complex was competed, indicated by C, with ~150-fold excess of cold, unlabeled probe DNA and supershifted, indicated by Ab, with polyclonal antibody sera raised against the C-terminal fragment of *An. gambiae* CTCF. The probe only lane is indicated by P.

3.4. Discussion

3.4.1. Differing rates of molecular substitution rate heterogeneity, smaller genomes and different genome organization may account for the decreased identity observed among dipteran CTCFs

Vertebrate CTCFs, from fish to human, are ≥ 98% identical across the entire zinc finger core of the protein. Comparison of the three dipteran CTCFs reveals 54% identity and 68% similarity within this same region. In addition, amino acid residues considered critical for DNA binding (Suzuki et al., 1994) are 89% conserved among these three insect species. This apparent discrepancy can be partially addressed by investigating the

molecular substitution rate heterogeneity among vertebrates and invertebrates. Recent maximum likelihood analysis of a set of 50 nuclear genes for vertebrates and dipterans, with *Arabidopsis* as an outgroup, suggests that the rate of vertebrate molecular evolution slowed considerably with respect to that of dipterans, before the origin of the crown-group Osteichthyes (Peterson et al., 2004). The much shorter generation times of dipterans have undoubtedly facilitated significant differences in their genome sizes (ranging from 179 Mb in *D. melanogaster* (Holt et al., 2002) to 813 Mb in *Ae. aegypti* (Warren and Crampton, 1991)) and gene organization patterns, attributable primarily to the amount and pattern of repetitive sequences (Severson et al., 2004). This would perhaps result in predictions of even greater sequence divergence than is observed in the CTCF genes. It seems likely that at least some of the many attributed vertebrate functions of CTCF are ancestral.

3.4.2. Evidence for conservation of CTCF in multiple insect orders

Each of the species examined yielded a single, extremely significant match followed by numerous matches of lesser significance, suggesting a single copy locus. Significant divergence in available N-terminal or C-terminal sequence supports the earlier observation that dipteran genomes have evolved very quickly, and thus these regions may not be critical to the conserved ancestral function(s) of this gene. Additionally, these regions may be more directly involved in protein-protein interactions with the interacting proteins having likewise undergone evolutionary adaptation. High bootstrap support and essentially identical trees generated by four independent methods establishes the trees presented in Fig. 15 as representative of the evolution of this gene

sequence. Less bootstrap support in the vertebrate clade is more indicative of the homogeneity of the sequence rather than uncertainty as to where these species should be located in the tree. Clearly, CTCF is present in vertebrates from fish through mammals and is highly conserved. Of interest is its consistent presence in all dipteran species queried. The relatedness of the protein sequences mirror the accepted taxonomic relationships among these species as presented at FlyBase (http://bugbane.bio.indiana.edu:7151/blast/), likely indicative of a conserved critical function. Significant EST evidence in the flour beetle, *Tribolium castaneum*, the honey bee, *Apis mellifera*, and in the silkworm moth, *Bombyx mori*, suggests the presence of CTCF-like genes in multiple insect orders.

3.4.3. Drosophila in situ *hybridization and microarray data support the mosquito RT-PCR expression profile*

The RT-PCR data from both mosquito and fly are consistent with one another, repeatable, and in agreement with both *in situ* hybridization data (Tomancak et al., 2002) posted for the fly at the Berkeley Drosophila Genome Project website (http://www.fruitfly.org) and the fly microarray data summarized at Yale University's Drosophila Developmental Gene Expression Timecourse website (http://genome.med.yale.edu/Lifecycle/query_gen.php?input1=FBgn0035769). *In situ* hybridization shows high-levels of *CTCF* transcript ubiquitously distributed throughout stage 1-3 embryos. mRNA levels then decrease until approximately stage 9 where they increase primarily in the developing nervous and sensory tissues. The neural-specific expression pattern also corresponds to findings in *X. laevis* where *in situ* hybridization

with staged embryos revealed weak homogeneous staining prior to stage 14, with subsequent upregulation in neural tissues and the sensory organs of the head (Burke et al., 2002). Also, over-expression of CTCF in mice during early embryogenesis resulted in decreased expression of the highly conserved homeobox gene *Pax6*, causing ocular defects (Li et al., 2004). Microarray data analysis clusters fly CTCF (CG8591) with genes exhibiting a single peak in expression during development, those showing significant expression increases in early embryogenesis, genes with expression changes of at least four-fold across development, and those expressed in the female germ-line (Arbeitman et al., 2002). Taken together, these expression data and the corresponding functional data from vertebrates suggest that CTCF may indeed be multi-functional in insects as well. Some possibilities include regulation of homeobox genes like *Pax6*, a role in the facilitation of chromatin organization during early development and establishment and/or maintenance of heterochromatic and euchromatic regions.

3.4.4. Conserved insulator function of mosquito CTCF is a promising tool for mosquito transgenesis

The EMSA data support a role for CTCF in endogenous mosquito insulator function and confirm recent findings that the insulator function of CTCF is conserved from invertebrate to vertebrate species (Moon et al., 2005). Currently, position effect and postion-effect variegation complicate efforts to establish effective transgenic lines in *Ae. aegypti* and other mosquitoes. Particularly problematic is the highly repetitive nature of much of the intergenic sequence, as well as the compact nature of the coding portion of the genome, which places regulatory elements from neighboring genes in

close proximity to one another, where they may inappropriately impact the transgene of interest. The ability to flank transgenes with short, conserved endogenous insulator sequences could significantly improve observed expression levels, and possibly increase the frequency of recovery of transgenic individuals.

3.5. Conclusions

We have cloned the cDNAs for two putative mosquito CTCF proteins. We have presented bioinformatics evidence that CTCF is likely present in many dipteran species and that the ancestral portion of the protein is clearly the zinc-finger region. Constitutively expressed in all life stages, mosquito CTCFs are distinctly upregulated in early embryos and in the ovarian tissues of blood-fed female mosquitoes. Finally, mosquito CTCF specifically binds both the chicken 5'HS4 *β-globin* and the fly *Fab8* insulator sequences. Further characterization of these CTCFs and their binding sites will provide a promising avenue for insulating transgenes in these medically-important mosquito species.

4. CTCF EXPRESSION IN THE MOSQUITOES *Aedes aegypti* AND *Anopheles gambiae* IS CONSISTENT WITH ITS POTENTIAL ROLE AS AN INSULATOR-ASSOCIATED PROTEIN

4.1. Introduction

Given the large relative size of their genomes, eukaryotes must organize and package their chromatin such that expressed genes remain accessible to the cellular transcription machinery and repressed genes remain quiescent. In addition, the effects of enhancers and repressors need to be restricted to their cognate genes, prohibiting inappropriate activation or repression of neighboring expression domains. Finally, developmental and tissue-specific expression demands dynamic reorganization of chromatin domains at specific times and in specific subsets of cells. This critical task of genome organization is attributed, in large part, to boundary elements or insulators, a diverse array of DNA sequences bound by one or more proteins (reviewed by Bell et al., 2001). Of several insulator-binding proteins characterized to date, only CTCF is common to both invertebrate and vertebrate organisms (Moon et al., 2005); Gray and Coates, in review).

Widely characterized in vertebrates, CTCF binds a variety of DNA sequences via different subsets of its eleven zinc finger (ZF) domains, making it a key player in a range of regulatory events: repression or activation of promoters, creation of hormone-responsive silencers and insulation of transcription domains via enhancer-blocking (reviewed by Ohlsson et al., 2001). In addition to the constitutive insulation of several genes, CTCF has recently been shown to be key to both the establishment and

maintenance of differentially methylated imprinting control regions (ICRs) that are critical to the proper expression of imprinted mammalian genes (reviewed by Lewis and Murrell, 2004). Notably, CTCF plays a critical, though not yet well-understood role in the complex process of X-chromosome inactivation (Lee, 2003; Pugacheva et al., 2005) and in the insulation of "escape" genes on inactive X-chromosomes (Filippova et al., 2005). Finally, a growing body of evidence supports the assertion that CTCF contributes to the structural and spatial organization of chromatin within the nucleus as a component of complex, multipartite boundary elements (Gombert et al., 2003; Yusufzai et al., 2004; Yusufzai and Felsenfeld, 2004; Ishihara and Morohashi, 2005).

CTCF orthologues with insulator-binding activity have recently been described in three dipteran species: *Drosophila melanogaster* (Moon et al., 2005), *Ae. aegypti* and *An. gambiae* (Gray and Coates, in review). Furthermore, there is bioinformatics evidence that CTCF is widely conserved across multiple insect species (data not shown). Here we present expression data here that support the role of CTCF as a conserved chromatin insulator protein, potentially associated with nuclear matrix scaffolds that are likely involved in chromatin domain organization. Additionally, the data implicate mosquito CTCF in developmental gene regulation and perhaps in the establishment and/or maintenance of imprinted chromatin.

4.2. Materials and Methods

4.2.1. Cell fractionation

Anopheles gambiae Sua4 cells, ~1.8 x 10^7 cells/mL, were washed twice with 1X PBS and fractionated essentially as previously described (Sun et al., 2001). Cells were

resuspended in TNM buffer (100 mM NaCl, 300 mM sucrose, 10 mM Tris-HCl, pH7.4, 2 mM $MgCl_2$, 1% thiodiglycol) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) and Complete Protease Inhibitor Cocktail (Roche, Indianapolis, IN 46250). Total cell protein (TCP) was obtained by sonicating a sample of the cells for three, 10 second cycles at 30% output power (Sonic Dismembranator 50, Fisher Scientific, Pittsburgh, PA 15275). Samples were cooled on ice between sonication cycles. The remaining cells were lysed in a 15 mL dounce homogenizer with 100 strokes of pestle "B". The resulting lysate was centrifuged at 4500 x *g* for 10 minutes to separate the cytoplasmic and nuclear fractions. The cytoplasmic fraction was aliquoted and snap-frozen, while the nuclei were resuspended in TNM buffer and extracted by adding 0.5% Triton X-100 and incubating on ice for 5 minutes. The extracted nuclei were centrifuged at 4500 x *g* for 10 minutes to separate the Triton-soluble nuclear fraction (NS) from the Triton-insoluble nuclear fraction (NP). The pellet was resuspended in an equal volume of TNM buffer with 0.5% Triton X-100, and both fractions were aliquoted and snap-frozen.

4.2.2. Immunoblotting

Bacterial pellets were resuspended and lysed as described in 3.2.12. Cultured cells were washed two times in 1X PBS and lysed by resuspending the pellets in 1M Tris-Cl, 5M NaCl and 1% IGEPAL CA-630 (Sigma Aldrich Chemical Company, St. Louis, MO 63178), incubating 10 minutes on ice and sonicating at 30% power for 3 cycles of 10 seconds each. Samples were cooled on ice between sonication cycles.

Total protein was quantitated using a BCA Protein Assay Kit (Pierce, Rockford, IL 61105). Equivalent amounts of total protein were separated on a denaturing gel and

immunoblotting was performed as described in 3.2.12. Rabbit anti-sera against *An. gambiae* CTCF (Texas A&M Laboratory Animal Resources and Research Facility) was cleared by centrifugation and then diluted 1:250 in blocking buffer. Anti-Rabbit IgG (Fc) AP conjugate (Promega, Madison, WI 53711) was used at a 1:7500 dilution.

4.2.3. Pre-adsorption of antibodies

Pre-immune rabbit sera or polyclonal rabbit anti-sera against a C-terminal fragment of *An. gambiae* CTCF (see 3.2.11) was diluted 1:15 in 1X BBT (10 mM Tris-HCl, 55 mM NaCl, 40 mM KCl, 7 mM $MgCl_2$, 5 mM $CaCl_2$, 20 mM sucrose, 0.1% BSA, 0.1% Tween-20), added to a confluent (80 cm^2) flask of pre-washed Sua4 cells, and incubated with gentle rocking at 4 ºC for 24 hours. The antibody solution was removed from the cells, cleared by centrifugation and stored at 4 ºC until use.

4.2.4.. Immunocytochemistry

Sua4 cells (Muller et al., 1999) were seeded on single-well chamber slides (Nalge Nunc International, Rochester, NY 14625) in 2 mL of Schneider's medium supplemented with 10% Fetal Bovine Serum (FBS). *Aedes aegypti* ATC10 cells were seeded as above, but in 2 mL of L-15 medium supplemented with 10% FBS. Cells were grown at 28 ºC for ~24 hours, yielding layers that were ~50% confluent. This protocol was adapted from (Muller et al., 1999). Cells were washed once with 1 volume of 1X PBS. All washes were done at room temperature, for ten minutes, on an orbital shaker. The wash solution was removed, and the cells were fixed in 1X PBS with 4% formaldehyde for 20 minutes at room temperature with gentle rocking. After the fixative was removed, cells were washed once more with one volume of 1X PBS, exposed for 2

minutes to 0.2% Triton X-100 in 1X PBS, and washed two more times in 1X PBS. The slides were blocked at room temperature for 2 hours in 1X PBS with 3% BSA, 0.2% IGEPAL CA-630 (Sigma), 0.2% Tween-20, 10% NFDM. After removing the blocking solution, slides were incubated overnight at 4 ºC with either pre-adsorbed rabbit pre-immune or CTCF antisera diluted 1:16 in 1X PBS with 1% BSA. The final dilution of the serum was 1:250. After three washes in 1X PBS, slides were incubated, in the dark, for 1 hour with Rhodamine Red™-X goat anti-rabbit IgG (H + L) (Invitrogen, Carlsbad, CA 92008), diluted 1:1000 in 1X PBS with 1% BSA. All of the remaining steps were done in the dark to prevent fading of the fluorescent conjugate. After three washes in 1X PBS, samples were mounted in Slow Fade® Light Antifade mounting media with DAPI (Invitrogen) and visualized and photographed on a Zeiss Axiovert 135 microscope with a CARV confocal module (Zeiss Micro-Imaging and Atto Bioscience). The slides were sealed with clear nail polish and stored at 4 ºC until image collection.

4.2.5. Whole-mount immunohistochemistry of embryos

Ae. aegypti embryos were collected for ~2 hours on damp filter paper and fixed essentially as described by Goltsev et al. (2004). Eggs were bleached for 75 seconds in 25% household bleach and immediately rinsed with copious amounts of deionized water. Approximately 100 embryos were placed in a scintillation vial with 1 part heptane and 1 part 9% formaldehyde, pH = 7 and incubated at room temperature for 25 minutes on an orbital shaker. Most of the fixative was then removed and replaced with 1 part heptane and 1 part deionized water and the embryos incubated an additional 30 minutes with gentle shaking. The water phase was removed and the vial filled to the top with boiling,

deionized water and incubated for 30 seconds. The water phase was removed once again, replaced with ice-cold deionized water and incubated on ice for 15 minutes. The water phase was removed, the old heptane was replaced with fresh heptane, an equal volume of methanol was added and the mixture was incubated for 10-15 minutes at room temperature. The embryos were washed several times with methanol and stored, ~25 per 1 mL microfuge tube, at -20 ºC.

After removing the endochorion of each embryo using a pair of fine forceps and rolling over double-stick tape, most of the methanol was removed and the embryos were rehydrated by three, 20 minute washes in 1 mL 1X BBT with Protease Inhibitor Cocktail (Roche). All washes were done by gently rocking the microfuge tubes at room temperature. *In situ* localization of proteins in whole mount embryos was performed essentially as described by Gonzalez-Gaitan and Jackle (1997). Rehydrated embryos were incubated overnight with 300 µL of pre-adsorbed primary antibody (pre-immune sera or polyclonal anti-sera against *An. gambiae* CTCF) diluted 1:16 in 1X BBT at 4 ºC. All remaining steps were carried out at room temperature. The antibody solution was removed and the embryos were washed for 20 minutes in 1 mL of 1X BBT. Non-specific binding was reduced by washing the embryos twice in 0.5 mL of 1X BBT with 3% goat serum. All remaining steps were carried out in the dark to prevent photobleaching of the fluorescent conjugate. Secondary antibody, Rhodamine Red™-X goat anti-rabbit IgG (H + L) (Invitrogen), was diluted 1:750 in 1X BBT and incubated with the embryos for 2 hours at room temperature with gentle shaking. After removal of the secondary antibody solution, the embryos were washed four times for 10 minutes in

1 mL of 1X PBST (1X PBS with 0.1% Tween-20). Embryos were mounted in Slow Fade® Light Antifade mounting media with DAPI (Invitrogen) on glass slides, in a narrow channel created by the edges of two coverslips adhered to the slide by double-stick tape. The embryos were covered by a third coverslip, the edges sealed with clear nail polish and the samples stored at 4 ºC. Images were obtained using a Zeiss Axiovert 135 microscope with a CARV confocal module (Zeiss Micro-Imaging and Atto Bioscience).

4.2.6. Whole-mount immunohistochemistry of ovaries

This procedure was performed essentially as described by Suter and Steward (1991). Ovaries were dissected in 1X PBS with Complete Protease Inhibitor Cocktail (Roche), taking care to separate individual ovarioles from each other and to remove as much of the connective tissue as possible, and fixed in 4% paraformaldehyde (Polysciences EM grade) in PBST for 15-20 minutes at room temperature. One half of the ovarioles from each female were used in the pre-immune control experiments while the remaining ovarioles were incubated with anti-sera against *An. gambiae* CTCF. The fixative was removed by three brief rinses followed by three, 5 minute washes with 1X PBST. Ovarioles were then incubated four times for 1 hour intervals with 1X PBST containing 1% BSA and 0.1% Triton X-100. Pre-adsorbed antibody solution (pre-immune sera or polyclonal anti-sera against *An. gambiae* CTCF) was diluted 1:2 in 1X PBST with 1% BSA and 0.1% Triton X-100 and incubated overnight at 4 ºC with gentle rocking. The final serum dilution was 1:30. After incubating for an additional 2 hours at room temperature, the primary antibody solution was removed, and the ovarioles were

washed three times for 10 minutes in 1X PBST. The remaining steps were completed in the dark. The secondary antibody, Rhodamine Red™-X goat anti-rabbit IgG (H + L) (Invitrogen), was diluted 1:1500 in 1X PBST and incubated with the ovarioles at room temperature for 1 hour. The secondary antibody solution was removed, and the ovarioles were washed three times with 1X PBST at room temperature. The ovarioles were mounted in Slow Fade® Light Antifade mounting media with DAPI (Invitrogen) on glass slides in a narrow channel bounded by two strips of double stick tape. A coverslip was overlain, the edges sealed with clear nail polish and the slides stored at 4 ºC until imaged on a Zeiss Axiovert 135 microscope with a CARV confocal module (Zeiss Micro-Imaging and Atto Bioscience).

4.3. Results

4.3.1. CTCF is expressed in cultured mosquito cell lines and in early Ae. aegypti *embryos*

Since the *CTCF* transcript was constitutively expressed in all life stages, though at higher levels in early embryos and in the ovaries of blood-fed females (Gray and Coates, in review), we investigated whether the protein product was present at detectable levels in two different mosquito cell lines, *An. gambiae* Sua4 and *Ae. aegypti* ATC10, and in *Ae. aegypti* embryos, ≤ 1 hour post-oviposition (Fig. 20). Immunoblot analysis reveals a single band for each lysate sample, migrating at ~84 kD for *An. gambiae* CTCF (680 amino acid residues) and at ~81 kD for *Ae. aegypti* CTCF (615 amino acid residues).

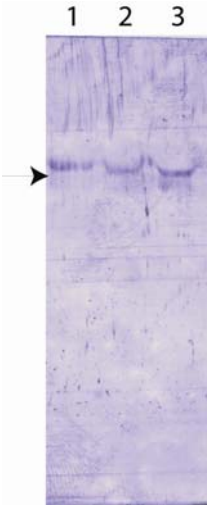


Fig. 20. CTCF protein is expressed in the *An. gambiae* Sua4 and *Ae. aegypti* ATC10 cell lines and in *Ae. aegypti* early embryos. Equivalent amounts of total protein from lysates of (1) *An. gambiae* Sua4 nuclei, (2) *Ae. aegypti* ATC10 total cell lysate, (3) *Ae. aegypti* embryos (<1hr post-oviposition) were fractionated on an 8% SDS-PAGE gel, electroblotted to a nylon membrane and incubated with rabbit antisera raised against a C-terminal fragment of *An. gambiae* CTCF. A goat anti-rabbit IgG (Fc) AP conjugate was used to detect the antibody-protein complex. The arrowhead indicates the 81.1 kD size marker.

4.3.2. CTCF is expressed in both the cytoplasmic and nuclear fractions of An. gambiae *Sua4 cultured cells, and is enriched in the Triton X-100-insoluble nuclear fraction*

In order to determine the subcellular distribution of CTCF, Sua4 cells were lysed in TNM buffer (Dunn et al., 2003) and separated into cytoplasmic (C) and nuclear (N) fractions. The nuclei were then extracted with 0.5% Triton X-100 to produce Triton X-100-insoluble (NP) and Triton X-100-soluble (TS) fractions. Triton X-100 removes loosely bound nuclear proteins, but does not disrupt the association of proteins tightly-bound to nuclear structures such as the matrix. All of the fractions were then subjected to immunoblot analysis (Fig. 21). CTCF was detected in all fractions; however,

proportionally greater amounts were detected in the Triton X-100 insoluble fraction, suggesting that significant amounts of CTCF are tighly-bound to nuclear structures.

Fig. 21. CTCF protein is expressed in both the cytoplasmic and nuclear fractions of cultured cells, and is enriched in the Triton X-insoluble fraction associated with the nuclear matrix. Equivalent amounts of total protein from the indicated cell fractions, TCP (total cell protein), C (cytoplasmic), N (nuclear), NP (Triton X-100-insoluble), NS (Triton X-100-soluble), were fractionated, electroblotted and the CTCF protein detected using a CTCF-specific polyclonal antibody as described in Fig. 20.

4.3.3. CTCF shows both cytoplasmic expression and localization to distinct nuclear foci in Ae. aegypti *ATC10 and* An. gambiae *Sua4 cell lines*

If significant amounts of CTCF protein are indeed bound to nuclear structures, immunocytochemistry should reveal distinct nuclear foci rather than more homogenous staining. Both *Ae. aegypti* ATC10 cells and *An. gambiae* Sua4 cells show this predicted pattern of CTCF distribution (Figs. 22 and 23). Unlike vertebrates, mosquito cells also show some cytoplasmic localization, though most of this appears localized close to the nuclear membrane.

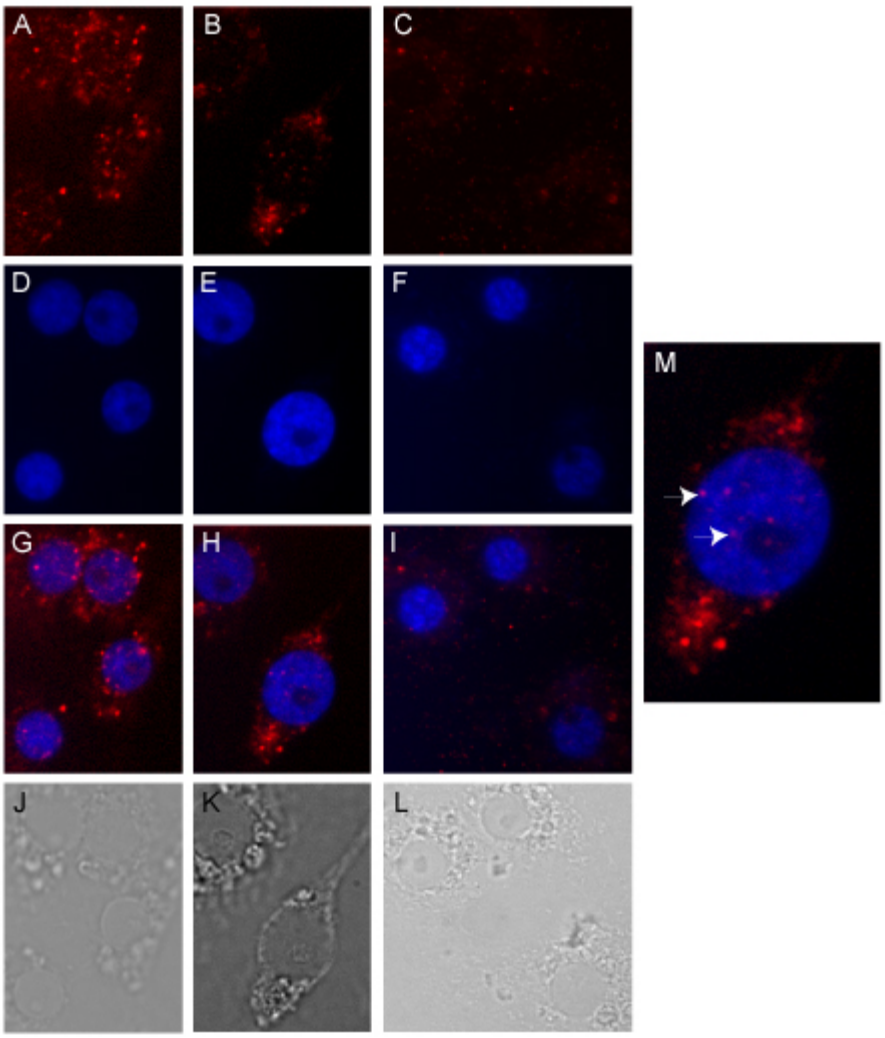


Fig. 22. Immunofluorescent analysis of CTCF expression in *Ae. aegypti* ATC10 cultured cells. Along with cytoplasmic expression, distinct foci within nuclei were seen after immunostaining using rabbit anti-CTCF polyclonal antibodies. CTCF was detected with a Rhodamine-conjugated 2º antibody (red) while nuclei were counterstained with DAPI (blue). Magnification X 630. (A and B) CTCF only, (C) preimmune control only, (D, E and F) DAPI only, (G and H) CTCF/DAPI merge, (I) preimmune control/DAPI merge, (J, K and L) bright-field, (M) CTCF/DAPI merge enlarged.

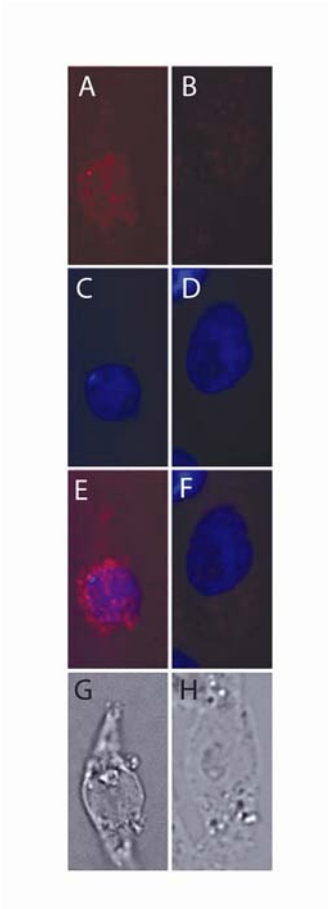


Fig. 23. Immunofluorescent analysis of CTCF expression in *An. gambiae* Sua4 cultured cells. As in Fig. 22, distinct foci of expression are observable in the nucleus with most of the cytoplasmic expression directly adjacent to the nucleus. Magnification X 630. (A) CTCF only, (B) preimmune control only, (C and D) DAPI only, (E) CTCF/DAPI merge, (F) preimmune control/DAPI merge, (G and H) bright-field.

4.3.4. CTCF protein is expressed throughout Ae. aegypti *syncytial, preblastoderm embryos, but is restricted to the follicle cells of mature ovarioles*

In order to further investigate the pattern of protein distribution during oogenesis and embryogenesis, CTCF protein was localized in *Ae. aegypti* mid-vitellogenic ovarioles, 36-48 hours after a second blood meal (Figs. 24 and 25), and in embryos collected ~4 hours post-oviposition (Fig. 26). Consistent with a maternal transcript,

mid- to late-stage vitellogenic ovarioles show no detectable CTCF expression in the oocyte. CTCF was, however, expressed markedly in regions occupied by both nurse cells and the single layer of follicular cells that envelops the oocyte.

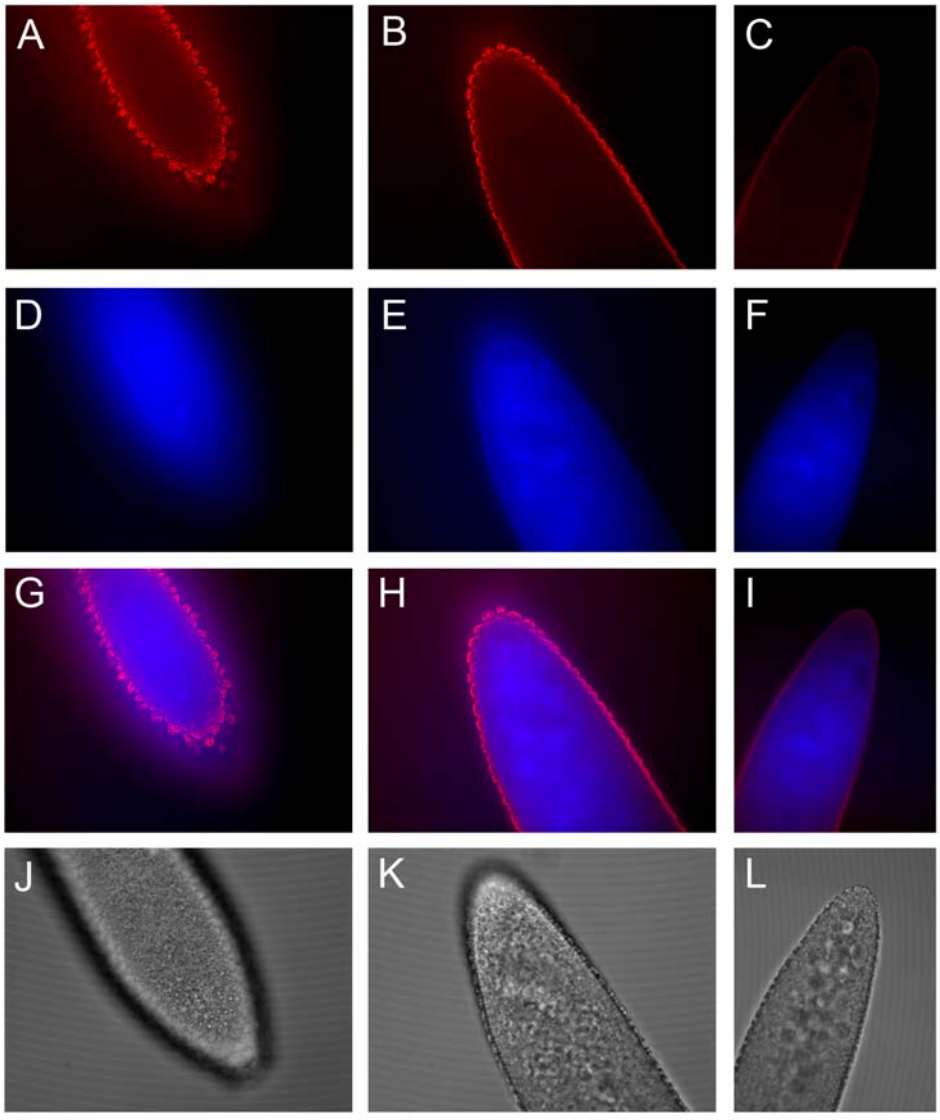


Fig. 24. CTCF expression in *Ae. aegypti* mid- to late-stage vitellogenic ovarioles. CTCF protein was immunolocalized in individually-fixed, mature ovarioles using rabbit anti-CTCF polyclonal antibodies and a Rhodamine-conjugated 2º antibody (red) as described in Fig. 3. Magnification X 630. (A and B) CTCF only, (C) preimmune control only, (D, E and F) DAPI only, (G and H) CTCF/DAPI merge, (I) preimmune control/DAPI merge, (J, K and L) bright-field.

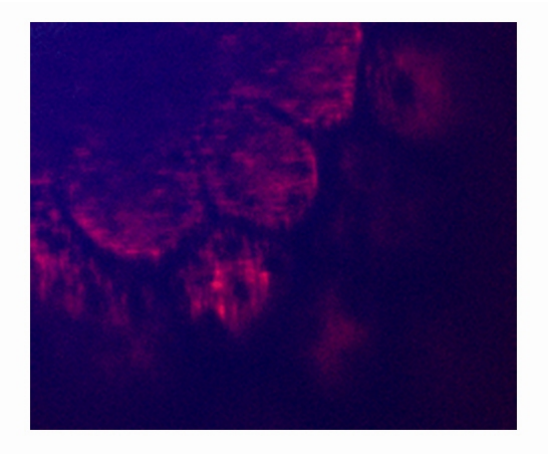


Fig. 25. Enlarged support cells from the tip of the ovariole pictured in Fig. 24G.

Earlier attempts at immunostaining of whole ovaries met with limited success due to the presence of a proteinaceous sheath of connective tissue that became very "sticky" after formaldehyde fixation. Earlier-stage ovarioles proved too fragile to separate, so we focused upon later stages of ovarian development. In these stages, the tunica propria portion of this sheath thins markedly during vitellogenesis (Chapman, 1998), making it more amenable to removal. In addition, more mature ovarian tissues possess less fragile follicles where the oocyte takes up more of the follicular space, allowing the complete removal of the connective tissue, while leaving the ovarioles intact.

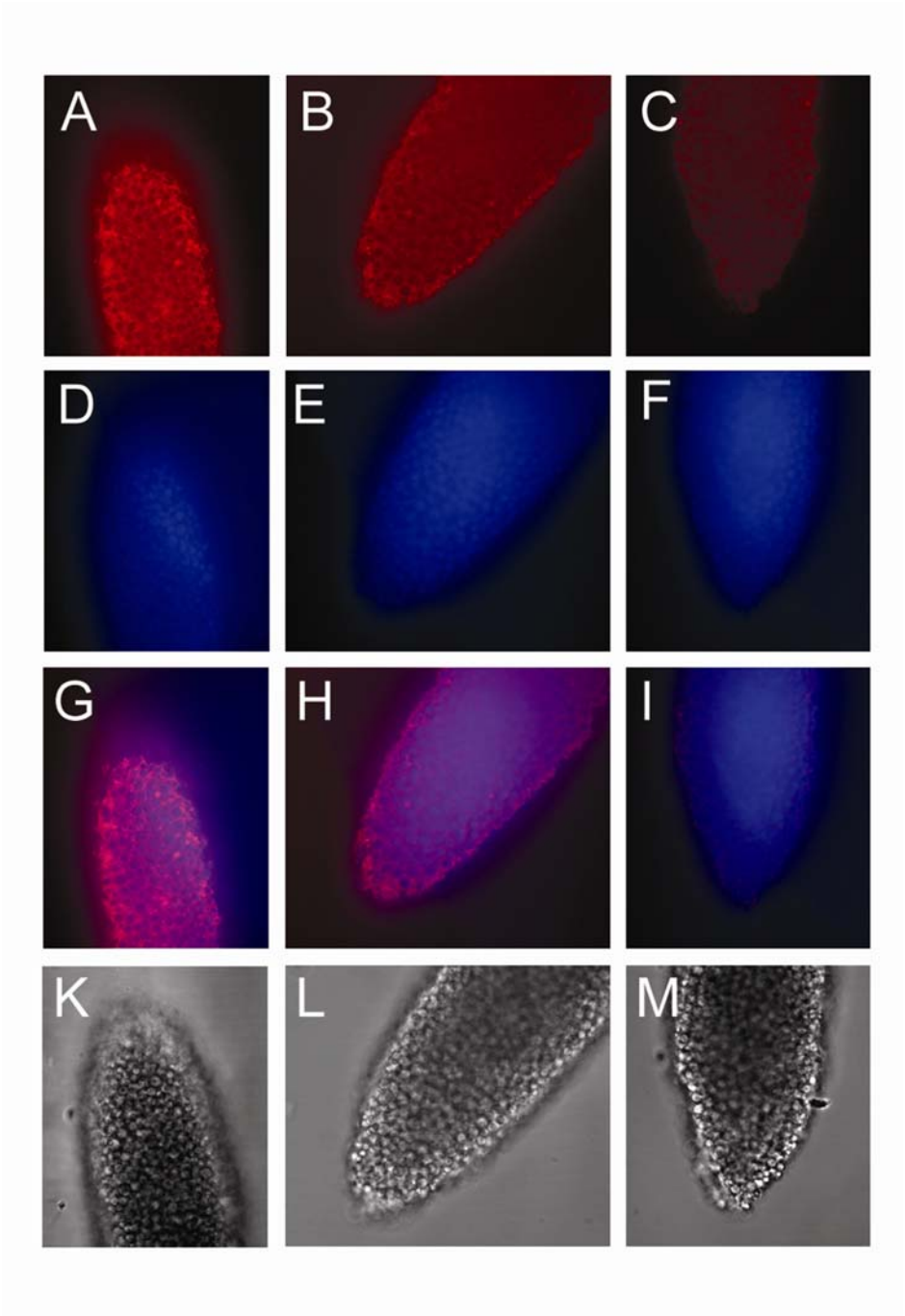


Fig. 26. CTCF expression in early *Ae. aegypti* embryos. Embryos (2 to 4 hours post-oviposition) were fixed and incubated with rabbit polyclonal anti-sera raised against a C-terminal fragment of *An. gambiae* CTCF. The antibody-protein complex was detected with Rhodamine-conjugated 2º antibody (red) while nuclei were counterstained with DAPI (blue). Magnification X 630. (A and B) CTCF only, (C) preimmune control only, (D, E and F) DAPI only, (G and H) CTCF/DAPI merge, (I) preimmune control/DAPI merge, (K, L and M) bright-field. (A, D, G and K) show the anterior end and (B, E, H and L) show the posterior end of the same embryo. (C, F, I and M) show the posterior end of a representative control embryo.

Because the eggs of this species develop tough, opaque endochorions, embryos must be hand-peeled prior to immunostaining. With no vitelline membrane at oviposition (Raminani and Cupp, 1975), and with the cell membrane essentially fused to the endochorion, there is very little internal structure—only a few nuclei in a "sea" of cytoplasm, so younger embryos proved impossible to peel in the preparation stages. The earliest time point at which embryos could be successfully manipulated for whole-mount immunolocalization was ~4 hours post-oviposition. At this syncytial stage, roughly equivalent to Bownes' stage 5 in *Drosophila* development, nuclei begin to migrate to the periphery of the embryo (Raminani and Cupp, 1975). At this time, CTCF appears ubiquitously expressed throughout the entire embryo in both the nuclei and the common cytoplasm.

4.4. Discussion

4.4.1. Mosquito CTCFs migrate aberrantly in SDS-PAGE, consistent with those of Drosophila *and vertebrates*

The mosquito CTCFs (81-84 kD) are considerably smaller than that of the *D. melanogaster* CTCF (~140 kD and 818 amino acid residues) previously reported (Moon et al., 2005), but they are consistent with the shorter polypeptides predicted for the mosquito transcripts (680 and 615 amino acid residues). All three insect CTCF proteins, however, exhibit greater relative molecular mass than is predicted by their sequence alone, much like the CTCFs of chicken and mammals (Klenova et al., 1997). This suggests that common structural features or post-translational modifications may be shared by all CTCF proteins.

4.4.2. Mosquito CTCF expression is consistent with nuclear matrix association

Similar cell fractionation experiments with human breast cancer cells revealed that CTCF expression is almost exclusively nuclear, and that the protein is tightly bound to nuclear structures, with none detectable in the Triton-soluble fraction (Dunn et al., 2003). These researchers also determined that CTCF binds specifically to nuclear matrix proteins, suggesting that CTCF may be part of a functional MAR (matrix attachment region) involved in the 3-D spatial organization of chromatin within the nucleus. An independent study subsequently confirmed that the 5'HS4 chicken β-globin insulator associates with the nuclear matrix in a CTCF-dependent manner (Yusufzai and Felsenfeld, 2004). While we detected some expression in the cytoplasmic and Triton X-100-soluble fractions, significantly greater amounts of CTCF were detected in the Triton X-100-insoluble fractions. This suggests that the association of CTCF with the nuclear matrix may be conserved in mosquitoes. Other researchers determined that CTCF expression is nuclear during interphase in HeLa cells, but that significant accumulations occur at the centrosomes and in the midbodies during mitotic cell division (Zhang et al., 2004). This data would be consistent with some cytoplasmic expression in actively dividing, asynchronous mosquito cells. Alternatively, the differences in subcellular localization may simply be due to fundamental differences in the role of CTCF in mammalian cancer cells compared to insect cells.

4.4.3. Vertebrate CTCF and other nuclear matrix-associated proteins show a similar pattern of nuclear localization

Multiple new binding sites for vertebrate CTCF were recently identified across the mouse genome, in both euchromatic and heterochromatic domains, using a ChIP-on-ChIP analysis (Mukhopadhyay et al., 2004). This same group observed scattered nuclear foci of CTCF expression that colocalize with the expression of the heterochromatin protein HP1β. HP1, identified as a mutant suppressor of PEV (position-effect-variegation) in *Drosophila* (Eissenberg et al., 1990), also interacts with the lamin B receptor (LBR), an integral membrane protein of the inner nuclear membrane (Ye and Worman, 1996), implicating HP1 in a chromatin tethering role. Finally, the *gypsy* insulator binding protein, suppressor of Hairy wing (su(Hw)), has been shown to be capable of bringing together otherwise distinct and distant chromatin domains by forming numerous, intensely-staining foci termed "insulator bodies", at the nuclear periphery to form looped domains (Gerasimova et al., 2000; Byrd and Corces, 2003). These observations are consistent with proteins that function to form chromatin boundaries, some of which may be anchored to nuclear structures.

4.4.4. CTCF expression in mature ovarioles and during early embryogenesis is consistent with that of a maternal gene

Expression of CTCF in the support cells rather than in the developing oocyte, suggests that it may be involved as a global transcription factor in these active cells and/or in the rearrangement of chromatin in order to facilitate the expression of genes specifically involved in oocyte development and the production of extraembryonic

membranes. The absence of the protein in the oocyte is also consistent with the expression pattern of maternal genes that are deposited in the developing oocyte as mRNA transcripts and are first translated post-fertilization, during early development. This observation appears to differ from mice where CTCF is expressed in both the germinal vesicle and in the cumulus cells (follicle cells) (Fedoriw et al., 2003). These researchers also used an RNAi strategy to elucidate the central role CTCF plays in developmental competence, presumably by protecting hypomethylated regions of the genome, such as the imprinted *H19* gene, from *de novo* methylation. Although dipteran insect genomes do exhibit methylation, major differences occur: the greatest levels of methylation are present during early embryonic development rather than in late development, the majority of methylation is asymmetrical CpA or CpT methylation rather than symmetrical CpG methylation, flies possess just one functional methyltransferase (Dmnt2) while vertebrates have at least four (Dmnt1, Dmnt2, Dmnt3a and Dmnt3b), and no parent-of-origin-specific imprinted genes have been described in insects, versus multiple genes in vertebrates (reviewed by Marhold et al., 2004). Collectively, the evidence suggests that CTCF acquired its function in imprinting after the divergence of protostomes and deuterostomes.

Despite no observable CTCF protein expression in the mature oocyte, significant amounts of CTCF expression were detected throughout early syncytial *Ae. aegypti* embryos (Fig. 18). Such a dramatic and rapid increase likely results from the translation of abundant maternal RNA transcripts following fertilization and oviposition. This observation is corroborated by immunolocalization of CTCF to the nuclei of stage 3

syncytial blastoderm *Drosophila* embryos (Moon et al., 2005). RNA transcript levels in oocytes and/or mature ovarian tissues, as measured by RT-PCR and/or *in situ* hybridization, are consistently high in *Ae. aegypti* (Gray and Coates, in review), *D. melanogaster* (Moon et al., 2005) and *X. laevis* (Burke et al., 2002). Maternal gene transcripts like *bicoid* and *nanos* often act as morphogens that impact the expression of downstream genes in a concentration dependent fashion, thus they localize specifically to different regions of the early embryo. CTCF appears to be distributed in a more uniform fashion, consistent with a different role. One specific insight into the developmental role of CTCF is the observation that CTCF regulates Pax6 expression in eye-development by binding to a repressor element, which in turn blocks the effect of an ectoderm enhancer tied to Pax6 expression in the eye tissues (Li et al., 2004). Many developmental genes must be regulated in both the proper temporal and spatial context. Perhaps CTCF plays a pivotal role in mediating the expression these genes by binding "conditional" insulators that facilitate the proper interaction of enhancers with their cognate promoters. These insulators may be regulated by binding of a ligand, such as a growth hormone. This type of CTCF-regulated insulator occurs in vertebrates at CTCF/TRE (thyroid response element) composite sites that are bound by CTCF and either a thyroid hormone receptor (TR) homodimer or a heterodimer of TR and retinoid-X-receptor (TR/RXR) in the absence of the thyroid hormone T3 (Lutz et al., 2003). Perhaps not coincidentally, insects possess an analogous type of ligand-activated system in the ecdysone receptor and ultraspiracle (EcR/USP) heterodimer that binds to

ecdysteroid-response elements (EcREs) within the regulatory regions of several early genes, which in turn regulate late gene expression (reviewed by Martin et al., 2001).

4.5. Conclusions

Newly identified in insect species, very little is known about the role(s) that CTCF may play in these species. Our expression data support roles for CTCF in chromatin organization and in insulation via multiple possible mechanisms. Mosquitoes impact human health worldwide by their superb ability to vector multiple pathogens. The identification of endogenous insulator sequences, with the ability to function early in development, could greatly benefit efforts to create transgenic strains refractory to pathogen transmission.

5. PCR-ASSISTED IDENTIFICATION OF *Anopheles gambiae* CTCF DNA BINDING SITES

5.1 Introduction

CTCF (CTCCC-binding factor) has been well-characterized in vertebrates as a ubiquitously-expressed, multivalent transcription factor and a key component of all known vertebrate insulators (reviewed by West et al., 2002). It binds an array of distinct DNA sequences with critical contact guanine residues within chicken, human, mouse and rat via combinatorial use of its 11 highly-conserved zinc fingers and, unlike any other multi-zinc finger protein known, is also capable of protein-protein interaction (YB-1, YY1 and RNP-K) via this same zinc-finger region (reviewed by Ohlsson et al., 2001). Recently, CTCF has been shown to bind differentially methylated DNA and to be key in both the maintenance of imprinted loci and in the process of X chromosome inactivation. In both instances, CTCF seems to act as a regulatable transcriptional switch within a chromatin boundary element (reviewed by Lee, 2003). Human CTCF has been shown to be associated with the nuclear matrix, a structure bound by DNA sequences known as MARs (matrix attachment regions) and enriched in chromatin remodeling enzymes (Dunn et al., 2003). To date, little is known about the role(s) of CTCF outside vertebrates, however recent evidence suggests that the role of CTCF in enhancer-blocking insulators is conserved in dipteran insects (Moon et al., 2005) and (Gray and Coates, in review).

Insulators are DNA-protein complexes that protect genes from the influences of neighboring regulatory elements and functionally possess one or both of the following

key properties: the ability to prevent the interaction between an enhancer and a promoter when juxtaposed between the two (enhancer-blocking) and the ability to protect genes from position effects by inhibiting the encroachment of heterochromatin into expressed euchromatic regions (barrier activity) (reviewed by Bell et al., 2001). Insulators clearly act via multiple, not always mutually exclusive, yet poorly understood mechanisms that may include: enhancer blockers that mimic competent promoter complexes and interact with the enhancer to prevent its association with the neighboring promoter; tethering multiple chromatin fibers to an "insulator body" to effectively dilute the effects of an enhancer by bringing multiple targets into close proximity; tethering enhancers near the base of chromatin loops, sterically preventing them from interacting with promoters outside their own loop domain; tethering chromatin to fixed nuclear structures such as the nuclear membrane, nucleolus or nuclear scaffolds thus imposing a fixed barrier to polymerase tracking; masking nucleosomes whereby insulator-associated proteins compete with histone modifying enzymes; nucleosome displacement where the presence of insulator protein complexes excludes nucleosome deposition, effectively removing the substrate for heterochromatin nucleation; and/or recruitment of histone-modifying enzymes to actively maintain an open chromatin conformation (reviewed by West and Fraser, 2005). Increasingly, insulators are characterized as complex, multipartite regulatory elements with separable functions. Two examples of insulators with separable CTCF-dependent enhancer-blocking and CTCF-independent barrier activities are the chicken 5'HS4 β-globin insulator (Recillas-Targa et al., 2002) and the MINE element at the 5' end of the *c-myc* locus in humans and mice (Gombert et al.,

2003). Furthermore, not all insulators are constitutive—many are regulated in developmental or tissue-specific contexts by mechanisms such as binding to differentially methylated sequences, poly-ADP-ribosylation by poly-ADP-ribose polymerase (PARP), and interaction with numerous protein co-factors (reviewed by West and Fraser, 2005). Insulators likely play a central role in the spatial organization of chromatin within the nucleus and are an integral part of ensuring the dynamic balance of interaction and segregation necessary for proper gene expression during an organisms life cycle.

DNA binding proteins with tandem zinc finger domains comprise one of the largest and most-diverse families of eukaryotic transcription factors. Many efforts have been made to elucidate a general "binding site code" for C2H2 zinc finger proteins (reviewed by Pabo and Nekludova, 2000), however multiple variables such as side-chain size, stereochemistry, orientation with respect to the major groove of the DNA, orientation within the α helix and interaction with neighboring residues, both within and adjoining the α helix, make this task extremely difficult. Additionally, recognition of zinc finger binding sites by multiple fingers in tandem array introduces the problem of neighboring domain influence, known as target site overlap (Segal, 2002). Finally, combinatorial use of subsets of their zinc finger domains makes zinc-finger proteins extraordinarily versatile (Ladomery and Dellaire, 2002). CTCF is one such protein, shown to bind divergent target sites in this manner (Filippova et al., 1996). It is not surprising then that CTCF binding sites have been elusive to predict.

CTCF-like cDNAs from the medically important mosquito species *Aedes aegypti* (AY935523) and Anopheles gambiae (AY939827) were recently cloned and characterized (Gray and Coates, in review). Putative DNA binding sites for the *An. gambiae* orthologue were determined using two independent, PCR-assisted, *in vitro* binding site selection experiments. Though not identical to core consensus sequences reported in vertebrates, the mosquito binding sites are C/G-rich and appear to be 14 bp in length, as reported for human, mouse and chicken CTCF (Chao et al., 2002; Ishihara and Sasaki, 2002; Bulger et al., 2003). *In silico* analysis of the selected random-primed genomic fragments indicates distribution in intergenic sequences as well as within introns of genes, consistent with regulatory function. The eventual identification of endogenous insulators associated with CTCF binding would provide valuable elements for protecting transgenes from the silencing effects of neighboring regulatory elements and unfavorable chromatin structure.

5.2. Materials and Methods

5.2.1. Generation of input fragment pools

Two separate input fragment pools were generated (Fig. 27). The first pool, termed R76, consists of a collection of radiolabeled, double-stranded oligonucleotides with a 26 base pair core of random DNA sequence, flanked by specific adaptor sequence with a *BamH*I restriction enzyme site at the 5' end and an *EcoR*I restriction enzyme site at the 3' end. Generally, the protocol outlined by (Pollock, 1996) was followed. To generate this pool, the R76 oligonucleotide (Table 5) was made double-stranded by primer extension using primer F and simultaneously labeled to a specific activity of 3200 Ci/mmol

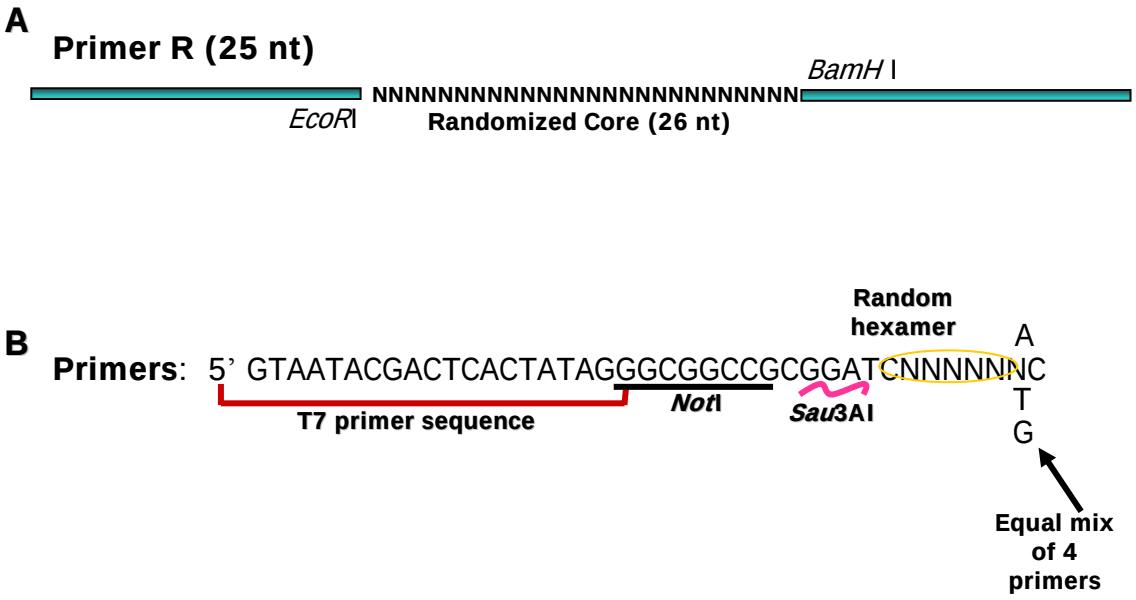


Fig. 27. Input pools generated for PCR-assisted binding site selection. (A) R76 oligo with known priming sequences and restriction sites, to allow PCR-amplification and subsequent cloning, flanking a randomized core of 26 nucleotides. (B) Bind-select primers consisting of a T7 primer sequence to allow PCR-amplification, a *Not*I restriction enzyme site for use in cloning, and the *Sau*3AI site which binds frequently within the *An. gambiae* genome, followed by a randomized hexamer. Four pools of bind-select primers were generated and then mixed in equimolar amounts to ensure that the amplification was not biased due to the presence of a single base at the 3'-most position in the primer.

with [α-^{32}P]CTP in the following 20 µL reaction: 100 ng R76, *Taq* DNA polymerase buffer (Continental Lab Products, San Diego, CA 92111), 1.5 mM $MgCl_2$, 50 µM 3dNTP mix (minus dCTP), 4 µM dCTP, 2 µL 80 ng/µL Primer F, 2 µL 10 Ci/mL (800 Ci/mmol) [α-^{32}P]dCTP (Amersham Pharmacia Biotech, Piscataway, NJ 08855), 5 units *Taq* DNA polymerase (CLP). The reaction conditions were as follows: 1' at 94 °C, 3' at 62 °C and 9' at 72 °C. The reaction was chased by incubating an additional 9' at 72 °C after adding 50 µM dCTP. Finally, the labeled DNA was purified on an 8% PAGE gel as described in 5.2.2.

Table 5
Primers for section 5

Identifier	Sequence
Primer F	5' gctgcagttgcactgaattcgcctc 3'
Primer R	5' caggtcagttcagcggatcctgtcg 3'
R76	5' caggtcagttcagcggatcctgtcg(N_{26})gaggcgaattcagtgcaactgcagc 3'
Bind-select "T"	5' taatacgactcactatagggcggccgcgatc(N_6)t 3'
Bind-select "G"	5' taatacgactcactatagggcggccgcgatc(N_6)g 3'
Bind-select "C"	5' taatacgactcactatagggcggccgcgatc(N_6)c 3'
Bind-select "A"	5' taatacgactcactatagggcggccgcgatc(N_6)a 3'
T7	5' taatacgactcactataggg 3'

Bind-select primers were mixed in equimolar amounts prior to use.

The second pool consisted of random genomic fragments generated by Klenow-mediated extension of an equimolar mix of "bind-select" primers (Table 5) hybridized to total *An. gambiae* genomic DNA in the following 50 µL reaction: *Taq* DNA polymerase buffer (CLP), 2 mM $MgCl_2$, 0.3 mM dNTPs, 2 µL of "bind-select" primer mix at a concentration of 80 pmol/µL, 1 µg genomic DNA and 5 units Klenow Polymerase (Promega, Madison, WI 53711). The reaction was assembled, minus the Klenow polymerase, and incubated at room temperature for 30 minutes to allow the primers to bind. The polymerase was then added and the reaction was incubated at room temperature for an additional 30 minutes to extend the primers and produce double-stranded templates for PCR-amplification. Finally, the reaction was denatured at 95 ºC for 5 minutes, 5 units *Taq* DNA polymerase (CLP) was added, and the newly-generated templates were PCR-amplified using the following reaction conditions: 95 ºC for 5'; 5 cycles of 95 ºC for 30", 45 ºC for 15", 72 ºC for 15"; 20 cycles of 95 ºC for 30", 65 ºC for 15", 72 ºC for 15". A negative control was performed in an identical reaction with no genomic DNA template. A single microliter of each reaction, experimental and

control, was subject to further amplification using T7 primer (Table 5). Reactions were assembled as follows: *Taq* polymerase buffer (CLP), 2 mM $MgCl_2$, 0.3 mM dNTPs, 2 µL 20 pmol/µL T7 primer and 2.5 units *Taq* DNA polymerase (CLP). PCR-amplification was performed according to the following reaction conditions: 95 ºC for 3'; 30 cycles of 95 ºC for 30", 55 ºC for 15", 72 ºC for 15". The reaction products were separated on a 1.5% agarose gel in 1X TBE buffer, and two gel slices containing amplified DNA fragments corresponding to 300-400 base pairs (bp) and 400-500 bp were excised and gel extracted using the QIAquick® Gel Extraction Kit (Qiagen, Valencia, CA 91355).

5.2.2. Purification of amplified and selected DNA fragments

Amplified or selected dsR76 fragments were purified on an 8% nondenaturing PAGE gel in 1X TBE buffer, dried and visualized by autoradiography (Fig. 28A). Typically, a clearly visible band was observed after 1.5 to 5 minutes of exposure. The gel slice containing labeled dsR76 fragments was excised with a clean blade, cut finely, placed in a microcentrifuge tube containing 250 µL of elution buffer (0.5 M ammonium acetate, 1 mM EDTA, 0.1% SDS), and incubated overnight at 37 ºC. After a brief centrifugation at 14,000 x g, the eluate was transferred to a new microcentrifuge tube, 1 µg of glycogen added and the DNA precipitated with 3M sodium acetate and 100% ethanol. The pellet was washed in 70% ethanol, dried and then resuspended in 10 µL of TE buffer.

Amplified or selected genomic DNA fragments (Fig. 28B) were purified on a 1.5% TBE agarose gel and the gel exposed to film covered with a single sheet of plastic

wrap for ~5-10 minutes. The labeled band was excised, gel-purified and ethanol precipitated in the presence of 1 µg glycogen as described above.

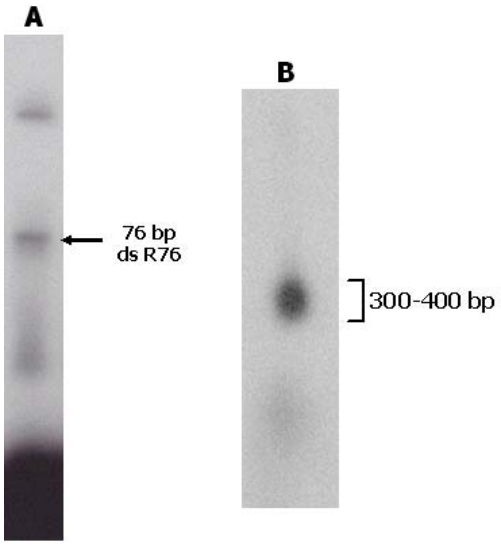


Fig. 28. Purification of input pools for PCR-assisted binding site selection. (A) R76 oligonucleotides purified by 5% non-denaturing PAGE. (B) *An. gambiae* genomic fragments purified on a 1.5% agarose gel. Unincorporated nucleotides ran off the gel.

5.2.3. Quantification of amplified and selected DNA fragments

Following purification, 1 µL of DNA was added to 1 mL of scintillation fluid and the counts per minute (cpm) measured in a scintillation counter (LS6000IC, Beckman Coulter, Fullerton, CA 92834). The amount of probe was calculated knowing that the molecular weight of R76 is 23,487 µg/µmol, the specific activity of the probe is 3200 Ci/mmol, and 10^6 cpm ~ 1 µCi.

5.2.4. PCR-assisted binding site selection

The protocol was followed as summarized in Fig. 29 and as outlined by (Pollock, 1996). Pre-swelled Protein G PLUS-Agarose beads (Oncogene Research Products, San Diego, CA 92121) were washed twice in 50 volumes of wash buffer (20 mM HEPES, pH 7.9; 100 mM KCl; 0.1 mM EDTA; 0.2 mM EGTA; 20% (v/v) glycerol) and then suspended in an equal volume of wash buffer containing 50 µg/mL BSA to produce a 50% (v/v) slurry and allowed to equilibrate for 2-3 hours at 4 ºC.

Rabbit reticulocyte lysate (RRL, Promega) was used to transcribe and translate *An. gambiae CTCF* in a coupled reaction using AgaCTCFpET (described in 3.2.3.) linearized by digestion with *Sph*I. The reaction was performed according to the manufacturer's instructions. A parallel set of reactions, with and without plasmid template, were carried out in the presence of ^{35}S-methionine (Amersham) and the products analyzed on an SDS-PAGE gel and visualized by autoradiography, in order to confirm the presence of a protein of the appropriate size (~81 kD).

The following binding reaction was assembled in a 1.5 mL microcentrifuge tube on ice: 20 µL binding buffer with BSA, 2 µL 100 ng/µL poly (dI-dC)·poly (dI-dC), 2 µL protein lysate (RRL), 2 µL 0.2 ng/µL radiolabeled dsR76 probe oligonucleotides or genomic fragments, and 1 µL rabbit polyclonal antisera raised against a C-terminal fragment of *An. gambiae* CTCF. For subsequent rounds of selection, half of the amount of probe was used. Protein-DNA complexes were allowed to form on ice for 20-30 minutes.

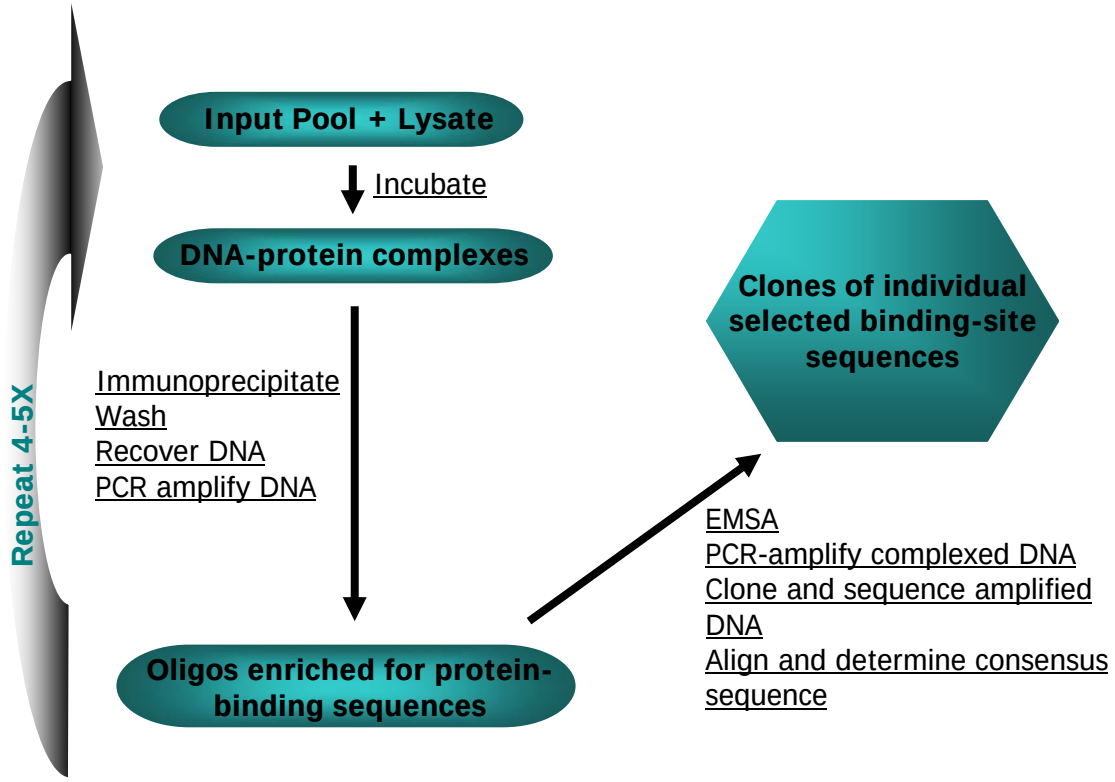


Fig. 29. Schematic of PCR-assisted binding site selection.

Beads were packed by adding 20 µL of 50% protein G-agarose slurry to a 1.5 mL microcentrifuge tube containing 250 µL of cold wash buffer without BSA, centrifuging at maximum speed for 15 seconds and aspirating off the wash buffer. The complexes were immunoprecipitated by mixing with 10 µL of packed beads and rocking the tube overnight at 4 ºC. The beads were washed three times with 250 µL cold binding buffer without BSA. For each wash, the binding buffer and complexes were vortexed and the tube inverted twice followed centrifugation at maximum speed for 15 seconds. The bound DNA was eluted from the protein G-agarose bead pellet by resuspending in 200 µL of recovery buffer (50 mM Tris-Cl, pH 8; 100 mM sodium acetate; 5 mM EDTA;

0.5% SDS) and incubating at 45 ºC for 1 hour. The DNA was purified by phenol/chloroform extraction followed by ethanol precipitation in the presence of 1 µg of glycogen as described in 5.2.2. The recovered DNA was quantified as described in 5.2.3.

The selected and purified DNA was PCR-amplified in a 20 µL reaction with *Taq* DNA polymerase buffer (CLP), 1.5 mM $MgCl_2$, 80 µM 3dNTP mix (minus dCTP), 4 µM dCTP, 2 µL 80 ng/µL Primer F, 2 µL 80 ng/µL Primer R, 1 µL 10 Ci/mL (800 Ci/mmol) [α-^{32}P]dCTP (Amersham Pharmacia Biotech, Piscataway, NJ 08855), 2.5 units *Taq* DNA polymerase (CLP). Reaction conditions were as follows: 15 cycles of 94 ºC for 1', 62 ºC for 1', 72 ºC for 1'. Unincorporated nucleotides were removed by gel filtration on a Sephadex G-50 column. The R76 oligonucleotide products were then PAGE purified, while the genomic fragments were fractionated and purified on an agarose gel as described in 5.2.2. Binding site selection and PCR-amplification were repeated for a total of five rounds of selection for the R76 oligonucleotides and four rounds for the random genomic fragments.

5.2.5. Electromobility Shift Assay (EMSA)

The binding of the selected oligonucleotides or genomic fragments to *An. gambiae* CTCF was then confirmed by mobility shift analysis using the same binding buffer used in the binding site selection process. Binding reactions were assembled on ice: 20 µL binding buffer with BSA, 2 µL 100 ng/µL poly (dI-dC)·poly (dI-dC), 2 µL 0.2 ng/µL radiolabeled dsR76 probe oligonucleotide, 2 µL protein lysate (RRL). One of three reactions was set up with just buffer, non-specific competitor DNA and the probe.

After incubation at room temperature for 10 minutes, 1 µL antiserum was added to one of the two remaining reactions. All reactions were incubated for an additional 20 minutes at room temperature before loading onto a gel. The R76 oligonucleotides were analyzed on a 5% non-denaturing page gel prepared with 1X TBE buffer and 2.5% glycerol, and pre-run for 20-30 minutes at 5V/cm. The gel was run at 5V/cm and 4 ºC for ~2.5 hours, dried and exposed to film overnight. The genomic fragments were analyzed on a 1.5% agarose gel with 2.5% glycerol. The gel was run at 7.5 V/cm for ~2 hours, dried and exposed to film overnight.

5.2.6. Extraction, amplification, cloning and sequencing of selected DNA fragments confirmed by EMSA

The region of the gel containing the DNA-protein complexes was excised and divided into four squares. Two squares were stored at -20 ºC. The paper was removed from the remaining two squares before placing the dried gel slices into a PCR tube and the following 50 µL reaction assembled: *Taq* DNA polymerase buffer (CLP), 1.5 mM $MgCl_2$, 80 µM 3dNTP mix (minus dCTP), 4 µM dCTP, 5 µL 80 ng/µL primer F, 5 µL 80 ng/µL primer R, 2.5 µL 10 Ci/mL (800 Ci/mmol) [α^{32}-P]dCTP (Amersham), 7.5 units *Taq* DNA polymerase (CLP). PCR was carried out for 17 cycles: 1' at 94 ºC, 1' at 62 ºC, 1' at 72 ºC. The reaction products were cleaned up by gel filtration, PAGE-purified (R76) or agarose-purified (genomic fragments), eluted and precipitated as described in 5.2.3. Finally, the selected DNA (R76 or genomic fragments) and 1 µg of the plasmid pBCKS+ (Stratagene, LaJolla, CA 92037) were double-digested with *EcoR*I and *BamH*I restriction endonucleases, phenol/chloroform extracted and ethanol precipitated in the

presence of 1 μg glycogen before ligating in a 10 μL reaction with 7 μL selected DNA, 1 μL pBCKS+, 1 μL 10X buffer and 1 μL T4 DNA ligase (Promega). The ligation was incubated at 4 ºC overnight and then transformed into electrocompetent DH10B cells. White and light blue colonies were PCR-screened for the presence of an insert. Insert-positive clones were cultured overnight, plasmid DNA mini-preps performed using the Promega mini-prep kit and the sequence of the insert DNA determined on an ABI 3100 capillary sequencer.

5.2.7. Motif searching and consensus representation

Several strategies were employed to determine a "consensus" binding site for the selected DNA oligonucleotides or fragments. The selected pools were treated separately, because the R76 oligonucleotides are entirely synthesized, so they may not actually be represented in the *An. gambiae* genome, whereas the genomic fragments were directly amplified from genomic DNA.

For the selected R76 oligonucleotides, common sequences were removed, obvious A/T-rich stretches were trimmed and a ClustalW alignment (http://ch.embnet.org/software/ClustalW.html) was performed. This alignment was manually inspected and edited with some of the flanking core sequence added back. This flanking sequence was available to be bound by CTCF in the selection process, but was removed initially so not to skew the alignment of the variable regions in the selected oligonucleotides. Because several different subsets of sequences have been characterized for vertebrate CTCFs, a subset of the most C/G-rich oligonucleotides were chosen for further analysis. The resultant alignment was submitted to WebLogo (Crooks

et al., 2004). The WebLogo profile was then used to formulate the 14 base pair "consensus" sequence 5' CNCCTCMSCMNSMM 3'. When there were only two nucleotides typically present at a given position, they were both incorporated (M = A or C, or S = C or G), while N was incorporated where there was no clear consensus.

For the larger genomic DNA fragments, the "consensus" sequence was used to identify similar sequences using the motif searching function of the Vector NTI™ Suite (InforMax, Inc., 1999). Each fragment contained at least one sequence with >55% identity to the "consensus". Each fragment was then used as a BLAST query against the *An. gambiae* whole genome sequence at Ensembl (http://www.ensembl.org/Anopheles_gambiae/). Two motif-searching algorithms, AlignACE (Roth et al., 1998) and MAST (Bailey and Gribskov, 1998) were employed to attempt to identify common elements within the selected genomic DNA fragments.

WebLogo was employed to visually represent the "consensus" sequences derived for the selected R76 oligonucleotides and genomic fragments. For comparison, a WebLogo profile was generated for vertebrate CTCFs by searching the literature for experimentally verified CTCF binding sites and submitting these to build a "consensus".

5.2.8. In silico *identification of potential CTCF binding sites in the* An. gambiae *genome using the consensus sequence derived from selected R76 oligonucleotides*

A JAVA script was kindly written by Andrea Julian, a collaborating lab member, to extract 5000 base pairs upstream and downstream of ~700 known genes in the Ensembl database (http://www.ensembl.org./Anopheles_gambiae/). Another JAVA

script was then written to search this flanking sequence for the "consensus" sequence derived from the selected R76 oligonucleotides.

5.3. Results

5.3.1. Many selected, double-stranded R76 oligonucleotides contain C/G-rich stretches

The final gel shift with the selected R76 oligonucleotide pool yielded two bands, a and b (Fig. 30A). When these were PCR-amplified, only band b produced detectable product (Fig. 30B). This product was excised, extracted and cloned. In all, 36 out of 40 clones submitted for sequencing resulted in unambiguous, unique sequences. Hence, the selected pool was diverse. These are presented in Table 6. The unboxed band was not followed as the sequences were presumed to be oligonucleotide dimers.

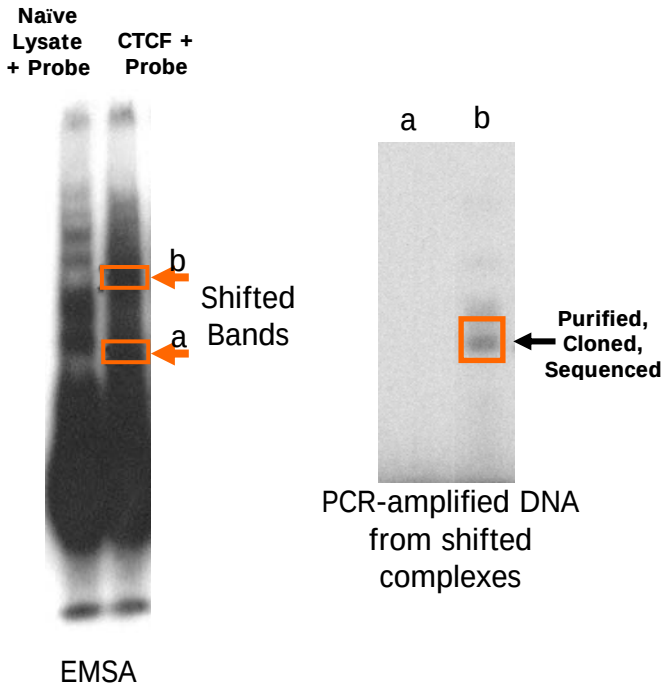


Fig. 30. EMSA with the final, amplified pool of selected R76 oligonucleotides produces one amplifiable band. (A) Lane 1, probe only; Lane 2, final pool of amplified, selected R76 oligonucleotides. (B) PCR-amplification of bands a and b excised from Lane 2 of panel A. * indicates the band excised, amplified and cloned into pBCKS+.

The ClustalW alignment (Fig. 31) was done with a subset of the R76 oligonucleotides in Table 6 as described in 5.2.7. These fragments are indicated by ∞ in Table 6. Some of the difficulties that arose in data analysis, and the assumptions that were consequently made are discussed in 5.4.1. Of the R76 oligonucleotides in the ClustalW alignment, several were subsequently shown to bind *An. gambiae* CTCF in competitive EMSA experiments (Fig. 32). These EMSA were performed as described in 3.2.12, using conditions not yet published when these binding-site selection experiments were done. Oligonucleotides shown to bind in these EMSAs are indicated by * in the ClustalW alignment (Fig. 31) and in Table 6.

```
*BS10      CGCCTCCCGCCCCA
 BS24      CTCTTCAATCCCAA
*BS11      CGCCTCAGACCGCT
 BS28      CGCCTCAGAACGCC
 BS32      CGCCTCAGCCTGTC
 BS15      CCTCTCAGCCACAC
*BS25      GAAGGCTCCCAAAC
*BS23      CTCTGCACCCAGAC
*BS16      CATCACACCACGAC
 BS27      CGCCTCACCCTGAA
*BS36      CGCCTCCCCATGCA
 BS38      CGCCTCCCCTATCA
Consensus  CNCCTCMSCMNSMM
```

Fig. 31. ClustalW alignment of a subset of selected R76 oligonucleotides. * indicates oligonucleotides that specifically bind *An. gambiae* CTCF as confirmed by EMSA. The consensus sequence was used to search the sequence flanking ~700 known *An. gambiae* genes.

Table 6
R76 oligonucleotides selected in PCR-assisted binding site selection

Identifier	Sequence—Forward strand
BS1	5’ agccatacatacaacttcgataaaaa 3’
BS3	5’ aatgacaacgcacgcacaaacaaaac 3’
BS4	5’ tcctcatacttgcaagattaattgc 3’
BS5	5’ acaaatgtaaactcgtttaaccagca 3’
BS6	5’ cctacgaaaatgccaatattaaataa 3’
BS7	5’ catccagcagggactgaagtcaagaa 3’
BS9	5’ atcacaccacaaccatgatacaataa 3’
BS10*∞	5’ ccgccccacaataacaaacaccaata 3’
BS11*∞	5’ agaccgctgtatgaatcaataatgag 3’
BS12	5’ gtcagaaagactagcatcaatattaa 3’
BS13	5’ taaaattacatttaaagatacaaca 3’
BS15∞	5’ aatcagccacaccacactaaggccaa 3’
BS16*∞	5’ acactctaacacatcacacca 3’
BS17	5’ cacaattacaaaagctttaaagaata 3’
BS18	5’ tgcaagtctaacaatgctgctaagat 3’
BS19	5’ aacacgtattcctataaagcaggaat 3’
BS20	5’ gatacagacacacatgaatataacaa 3’
BS21	5’ cacagatggatgcctattagacagac 3’
BS23*∞	5’ tgcacccagacaatgagtagcatgac 3’
BS24∞	5’ ttcaatcccaacattaaccccataag 3’
BS25*∞	5’ gaaggctcccaaacaacaacaagtaa 3’
BS27∞	5’ accctgaaacacactagaaataataa 3’
BS28∞	5’ agaacgccatcacttcgataaaataa 3’
BS29	5’ atcaaacagctagcaaaacaacttaa 3’
BS30	5’ cacaaacgaagcaaaacaataggtcc 3’
BS31	5’ tagacggttaagaaaataactgaata 3’
BS32∞	5’ agcctgtctaaagtaataatgcaca 3’
BS33	5’ actacgtcactcataaaaaatgaaat 3’
BS34	5’ agcatgaacacctaaaataacaaaga 3’
BS35	5’ tgggaacatagaaaaaataaagcaag 3’
BS36*∞	5’ cccatgcagctacatgaaaaattttc 3’
BS37	5’ acaaagatcctaactagacacaaaat 3’
BS38∞	5’ ccctatcacccgaacgttaattataa 3’
BS39	5’ tctagcaaatcatcagaatacgatac 3’

* indicates that binding to *An. gambiae* CTCF was confirmed by competitive EMSA.
∞ indicates the oligonucleotide was used to determine the “consensus” binding site used for *in silico* analysis.

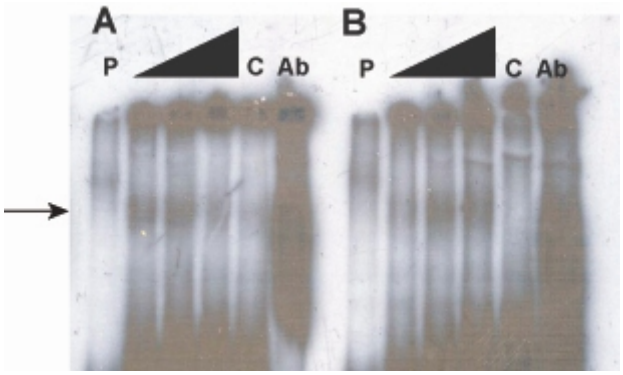


Fig. 32. Competitive EMSA for two of the selected R76 oligonucleotides. Probe only lane is indicated by "P"; >500-fold specific cold competitor is indicated by "C"; Polyclonal antisera against *An. gambiae* CTCF is indicated by "Ab". Increasing amounts of Sua4 cell lysate are indicated by the triangle. (A) BS25; (B) BS36.

5.3.2. Selected genomic DNA fragments are distributed across the genome and occur in several different contexts

In an effort to get a better representation of sites that *An. gambiae* CTCF binds *in vivo*, the PCR-assisted binding-site selection experiment was repeated with an input pool generated from *An. gambiae* genomic DNA. This time the input consisted of sequences that occur naturally in the genome rather than all possible combinations of nucleotides within the variable region of R76. Only five unique fragments were selected from the random *An. gambiae* genomic fragment pool, however. This is reflected in the poor EMSA result and subsequent amplification at the conclusion of this experiment (Fig. 33). The selected fragments were used to query the *An. gambiae* genome. The selected fragments and their respective genomic contexts as determined by BLAST results are summarized in Table 7.

Fig. 33. EMSA with the final, amplified pool of random genomic DNA fragments yields a single, faint amplified band of products. (A) EMSA after the final round of selection. Lane 1 is probe only; Lane 2 shows a slight increase in the intensity of the shifted band. (B) A single, PCR-amplified band of products corresponding to the band of the same size in A-2.

Table 7
Summary of genomic DNA selection

	Intergenic	Intron	CpG Island	Class II Transposon	Position in Genome
BS1a*	√	√		√	X, 2L, UNKN ¶
BS1b		√			2L
BS3	√	√		√	2L, 2R, 3L, 3R, X ¶
BS5	√			√	2L, 2R, 3L, 3R, X ¶
BS17			√		3R

* indicates that this sequence was represented 8 times in the selected pool. ¶ indicates multiple hits within the genome.

5.3.3. Consensus WebLogo profiles reveal C/G-rich core binding sites in vertebrates and in the An. gambiae*-selected fragment pools*

In order to compare the *An. gambiae* CTCF selected core binding sites from both PCR-assisted *in vitro* binding site selection experiments presented in 5.3.1. and 5.3.2. with each other and with known CTCF core binding sites in vertebrates, WebLogo

profiles were generated. In order to generate a vertebrate profile, the literature was searched for known vertebrate CTCF core binding sites. These are summarized in Table 8. WebLogo profiles for vertebrate CTCF core binding sites and for both binding site selection experiments are presented in Fig. 34. Though not identical, all appear C/G-rich and approximately 14 base-pairs in length. This is particularly significant given the >60% A/T composition of the *An. gambiae* genome.

Table 8
Vertebrate CTCF DNA binding sites

Binding-site Identifier	Sequence	Reference
CTCF site "A"	5' ccgccccctagcgg 3'	(Chao et al., 2002)
CTCF site "B"	5' ctgccccctagcgg 3'	(Chao et al., 2002)
CTCF site "C"	5' ctgccgccgtgcgg 3'	(Chao et al., 2002)
CTCF site "D"	5' ctgccaccacgcgg 3'	(Chao et al., 2002)
Mouse H19m1	5' ctgccaccggggac 3'	(Hark et al., 2000)
Mouse H19m2	5' ctgccgcataacgg 3'	(Hark et al., 2000)
β-globin	5' ctgccgccgccagg 3'	(Farrell et al., 2002)
DM1-site 1	5' ccgccccctagcgg 3'	(Filippova et al., 2001)
DM1-site 2	5' cccccacctatcgt 3'	(Filippova et al., 2001)
Human Igf2-H19	5' ctgccgccgcgcgg 3'	(Bell and Felsenfeld, 2000)
Chicken 5'HS4 FII	5' ctgccccctagcgg 3'	(Bell et al., 1999)
Mouse HS-85.5	5' ctgccctctcctgg 3'	(Farrell et al., 2000)
Mouse HS-62.5	5' cttccccctggtgg 3'	(Farrell et al., 2000)
Human HS-III	5' gctccccctggtgg 3'	(Farrell et al., 2000)
Mouse 5'HS5	5' cttccctctagtgg 3'	(Farrell et al., 2002)
Human 5'HS5	5' cttccctctagtgg 3'	(Farrell et al., 2002)
Mouse 3'HS1	5' ctgccccctactgg 3'	(Farrell et al., 2002)
Human 3'HS1	5' ctgaccccctagtgg 3'	(Farrell et al., 2002)

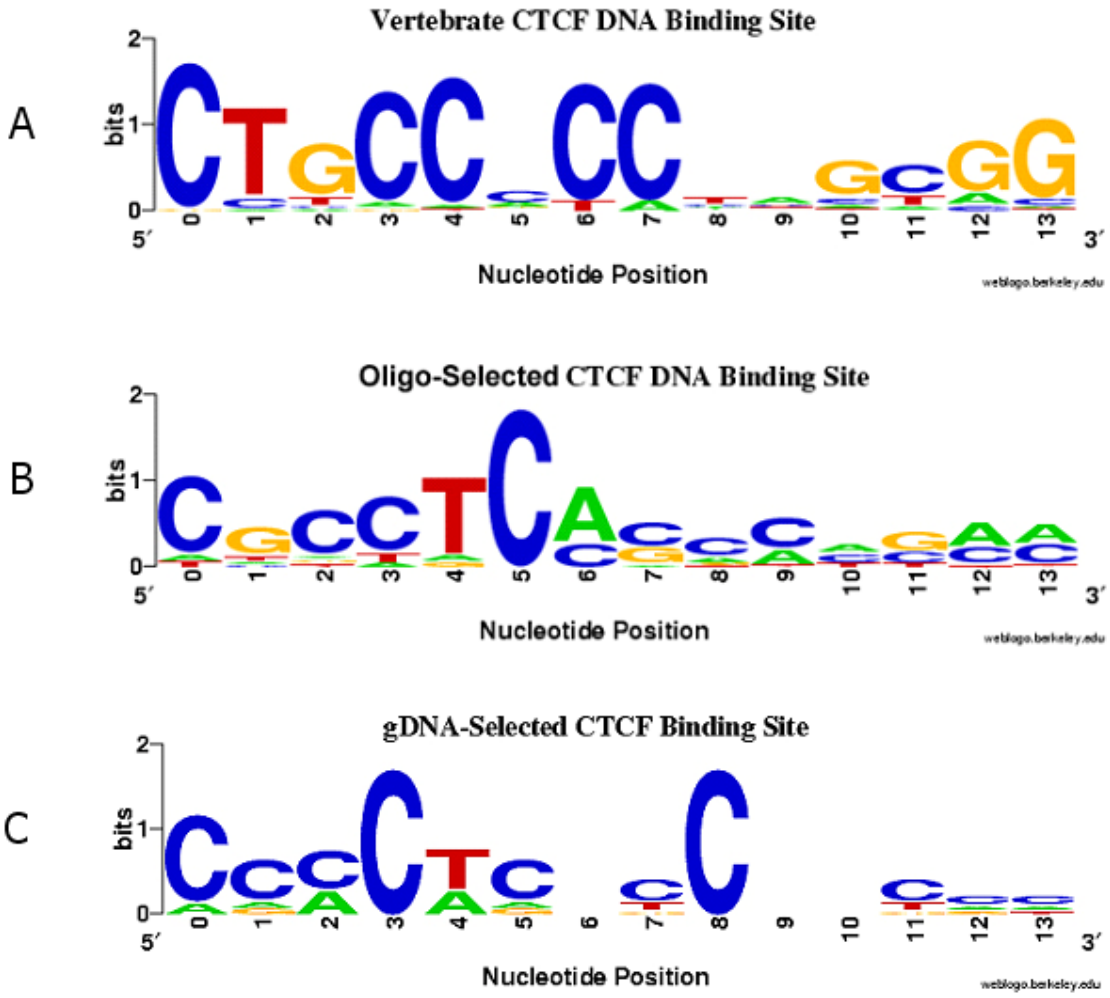


Fig. 34. WebLogo profiles of CTCF "consensus" binding-sites. (A) Vertebrate sequences presented in Table 8; (B) Selected R76 oligonucleotides presented in Table 6; (C) Selected genomic fragments presented in Table 7. The WebLogo software utilized was developed by (Crooks et al., 2004).

5.3.4. In silic*o mining of the* An. gambiae *genome yields potential CTCF binding sites in the flanking regions of many genes*

Though it is likely representative of only a subset of binding sites for *An. gambiae* CTCF, the "consensus" site derived from 12 of the selected R76 oligonucleotides shown in Fig. 31 and Fig. 34B was used to search sequence flanking known *An. gambiae* genes in an effort to identify sequences with potential insulating activity. Known insulators, as discussed in sections 3 and 4, protect genes within their boundaries. A set of 124 putative CTCF targets was generated. These are presented in Appendix B. Several binding sites flank genes that encode proteins that are regulated in conjunction with specific biological processes that are mediated by the sequential and/or coordinate expression of multiple genes. Insulators have recently been associated with a number of such genes. These sites are summarized in Table 9.

Table 9

Putative, *in silico* CTCF binding-sites associated with specific biological processes

Major Function	Putative, *in silico* CTCF binding-sites
Cytochrome P450s	CTCCTCACCCACCA CTCCTCCGCCGGAA CTCCTCCGCAACAA CACCTCCGCCCGCA CCCCTCCCCCCCCC CCCCTCCCCCCCCC
Caspases	CCCCTCCCCACCCA CTCCTCCCCACCAC CTCCTCCCCACCAC CTCCTCCCCACCAC CTCCTCCCCAGGCC
GCPR—odorant	CGCCTCCGCCGGAA CCCCTCCCCCCCCC CCCCTCCCCCCCCC CTCCTCCCCCGGAC CCCCTCCCCACCCA CCCCTCCCCCTCCA CTCCTCACCACCCA CACCTCAGCCGCCA CCCCTCCCCCTCCC
Pheromone/general odorant	CCCCTCACCAGCCA CCCCTCCCCCCCCC CGCCTCCGCATGAC CGCCTCACCATGCC CACCTCCCCAACAA
Immunity	CACCTCCGCCACCA CCCCTCCCCCTCCA CGCCTCCGCATGAC
Blood feeding	CCCCTCACCCACAC CCCCTCCCCACCCC CTCCTCCGCAACAA CCCCTCCGCCCCCC
GCPR—opsin	CACCTCCCCACGCA CTCCTCCGCACCAC CCCCTCCCCACCCA
Development	CGCCTCCCCCCGCC CCCCTCCCCCTCCC CGCCTCACCAGGAA CGCCTCCCCCCGAA
Serpins	CGCCTCCCCAGGCC CTCCTCCCCACCCC CCCCTCCCCCACAC

5.4. Discussion

5.4.1. Selected, short oligonucleotides appear biased by the common adaptor sequence at the 5' end of the input fragments

Alignment of the selected R76 oligonucleotides has proven to be a daunting task. There are common sequences that flank the 26 base pair randomized core of the oligonucleotides that were subjected to the selection process. This is a consequence of the technique as these sequences are needed for both PCR-amplification and cloning. Nonetheless, if they are not removed prior to alignment, they will skew the alignment so much that the actual binding site will either appear significantly altered or will be missed altogether. Most of the C/G-rich sites are extensions of one of the two flanking regions. This flanking region consists of an *EcoR*I restriction enzyme site (GAATTC) followed by GCCTC(N_{26}). Omitting the flanking sequence prior to analysis allowed identification of oligonucleotides such as BS16, BS23 and BS25 that have internal binding sites for CTCF. All of these bind CTCF in competitive EMSA experiments (Fig. 32 and data not shown). Nonetheless, oligonucleotides with significant stretches of C/G-rich sequence adjacent to this flanking sequence, such as BS10, BS11 and BS36, also bind CTCF specifically. In retrospect, the known flanking sequences should probably have been designed to contain >60% A/T as zinc-finger proteins typically bind targets that are >50% C/G-rich (Ladomery and Dellaire, 2002).

5.4.2. The presence of endogenous CTCF in the rabbit reticulocyte lysate interferes with the identification of DNA-mosquito CTCF complexes in agarose gels

CTCF is highly-conserved among vertebrate organisms, thus rabbits would be predicted to possess a CTCF gene and to produce a CTCF protein of similar size to that seen in other vertebrates. In SDS-PAGE gels, this protein migrates at ~140 kD (Dunn et al., 2003). Added difficulty arises with the variable sizes of the genomic fragments selected (300-400 bp). This required the use of 1.5% agarose gels rather than 5% acrylamide to resolve the DNA-CTCF complexes. The resulting EMSA did not give a distinct shifted band that was unique to the programmed lysate. The shifted band from the *An. gambiae* CTCF lane did amplify by PCR, but this does not exclude the possibility that the rabbit CTCF protein could have bound the mosquito genomic DNA fragments. Even so, the varied contexts of the selected genomic fragments raise a number of intriguing possibilities that warrant further investigation.

5.4.3. Possible association of CTCF with transposable elements could reveal yet another function of CTCF in mosquitoes

Of particular interest is the number of times that several of the selected genomic fragments were found at or near TE boundaries (Table 7). Mosquitoes possess a diverse array of DNA TEs, many of which are uncharacterized (Tu and Coates, 2004). TE regulation is currently a "black box" in many organisms, but it is evident that some type of negative regulation must be employed by host genomes in order to prevent the mutagenic effects of excessive TE movement. Perhaps CTCF is functioning to block the binding of proteins necessary for transposition, is recruiting other proteins that block the

propagation of heterochromatin formation often associated with TEs, and/or is blocking the interaction of transcription factor necessary for expression of the transposase that catalyzes excision and movement of the TE. Flanking transgenes with known endogenous CTCF sites could provide a powerful means of shielding them from the very silencing that the TE-associated elements required for genome integration may evoke.

5.4.4. Putative, in silico *sites are consistent with CTCF function as a boundary element protein*

Boundary elements are often positioned between two differentially regulated genes or on the flanking edges of a group of coordinately-regulated genes. Both situations are illustrated by the *in silico* data presented in Table 9 and in Appendix B. A specific example is the cluster of serpin genes found on *An. gambiae* chromosome 2L (Fig. 35). There are five known serpins and at least one predicted serpin arranged in two sub-groups separated by ~9.2 megabases (Mb) of DNA. At either end of the entire cluster is an array of several putative CTCF binding sites, identified by the *in silico* JAVA script employed in this study and by visual inspection of the surrounding sequence for closely-related sites. Just upstream of the 5'-most putative CTCF binding site, lies a putative connective tissue growth factor orthologue, while several kilobases downstream of the 3'most putative CTCF binding site, lies a putative Syntaxin 1A orthologue. Syntaxin 1A is believed to be involved in neurotransmitter transport (http://www.ensembl.org/). Clearly, both the upstream and the downstream genes would be expected to exhibit expression profiles distinct from that of serine-protease inhibitors (serpins). The upstream sub-cluster is thought to be involved in embryonic patterning

(Rushlow, 2004), while the downstream sub-cluster contains serpins known to be involved in innate immunity in dipteran insects (DeGregorio et al., 2002; Danielli et al., 2003). The annotation of the *An. gambiae* genome is very much a work in progress, and will likely be furthered by the impending release of the *Ae. aegypti* complete genome sequence. Further analysis of the *in silico* dataset in light of improved annotation will likely reveal other gene clusters that may also be insulated by CTCF.

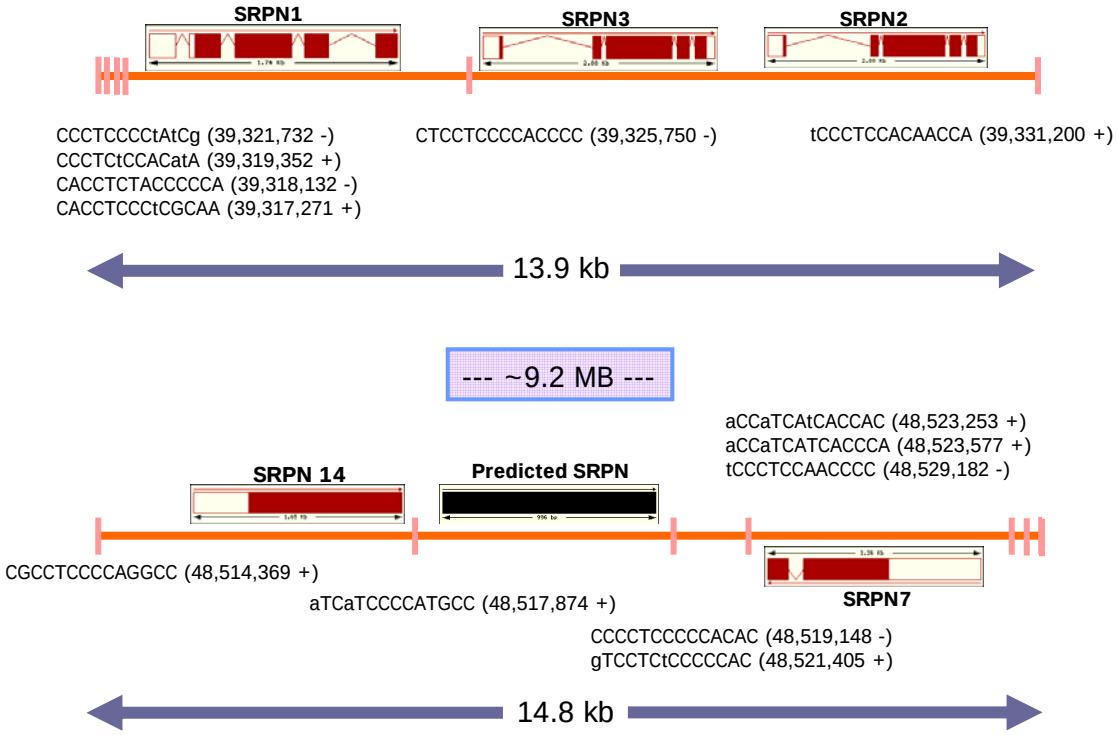


Fig. 35. Putative, *in silico* CTCF binding-sites within the serpin gene cluster of *An. gambiae*.

5.5. Conclusions

Vertebrate CTCF binds a diverse array of binding sites *in vivo*, not always with the same affinity seen *in vitro*. This is likely due to the potential role of CTCF in the

arrangement of chromatin into domains within the three-dimensional space of the nucleus. Mosquito CTCF appears capable of binding similar C/G-rich sites that are likewise scattered across the genome. Ultimately, these predictions must be confirmed with *in vivo* experiments. Though very preliminary, the data presented provide promising avenues for further investigation. Understanding the native context of CTCF function in mosquitoes will be important in developing effective methods to employ it for transgene insulation.

6. GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

6.1 General Conclusions

Over the past six years, much effort has been put forth to develop germ-line transformation in non-drosophilid insects, in particular for several mosquito species that vector pathogens responsible for a great deal of human morbidity and mortality. It has become increasingly apparent that each target species poses unique challenges arising from different genome organization, habitat impact and evolutionary pressures imposed both by its pathogen(s) and human control efforts. Much remains to be done before genetic transformation in *Ae. aegypti* and *An. gambiae* becomes routine. Routine transformation would provide not only a powerful tool for genetic control, but would enable gene discovery and characterization on a scale that would rival what is currently possible in *Drosophila*. It is certainly a work in progress.

In this study, it has been shown that exogenous promoter/enhancer combinations can be employed to increase the expression of a reporter gene in cultured mosquito cells. Furthermore, a viral *trans*activating protein can be used to further increase expression. Data presented in this work confirms the hypothesis that some exogenous regulatory elements may function as predicted in a mosquito background. Unanticipated increases in expression resulting from the presence of the IE1 *trans*activating protein also highlight the necessity of testing exogenous regulatory elements in each new target species. Cultured mosquito cells provide one powerful tool to evaluate novel regulatory regions for potential transgene expression.

Mosquito CTCFs were cloned from the mosquitoes, *Ae. aegypti* and *An. gambiae.* They were shown to be expressed at detectable levels in all stages of development, with increased expression in early embryonic development and during oogenesis. The mosquito CTCFs were shown to bind two known insulator sequences in a manner similar to that of *Drosophila* CTCF. Thus, it is likely that the insulator function of CTCF is conserved in mosquitoes. This is the first potential insulator identified in mosquitoes.

Studies of CTCF expression in *Ae. aegypti* reveal high levels of protein in the support cells of the ovarian follicle and thoughout early syncytial embryos, consistent with the RT-PCR transcriptional profile. Cell lines show expression in both the cytoplasmic and nuclear fractions, with enrichment of the protein in the insoluble nuclear fraction. This is consistent with CTCF association with a nuclear structure, perhaps the nuclear matrix. Further investigation of CTCF and its native binding sites may yield valuable insight into the patterns of chromatin organization in these mosquitoes.

Finally, CTCF appears to bind C/G-rich sites in *An. gambiae* that are very similar to those bound by vertebrate CTCF. *In silico* analysis reveals many putative targets for CTCF, consistent with what has been observed in humans. Further studies will be needed to characterize the binding preferences of mosquito CTCF *in vivo.* Once known, these binding preferences may be exploited to craft effective insulating sequences for transgenes, potentially increasing the number of usable stable lines recovered per experiment.

6.2 For Future Consideration

A caveat of using exogenous promoter sequences is that they must always be tested in the species of interest. This is not trivial when transgenic lines are difficult to generate. Cell lines can be useful when considering general promoters, but many cell lines do not transcribe from tissue-specific or temporal-specific promoters very well, if at all. The *trans*activating protein, IE1, is a mixed "bag". It is effective in upregulating expression from its target enhancer/promoter, however it appears to be promiscuous in its binding activity. This could result in the misregulation of important endogenous genes. Of particular concern are those involved in development. For expression of the transposase, more may actually not be better, due to the tendency of host genomes to negatively regulate active TEs. With the whole genome sequencing efforts in both *An. gambiae* and *Ae. aegypti*, endogenous genes will be much easier to identify and characterize in their native contexts. Better candidates for transgene expression and transposase regulation would probably be stage- and tissue-specific endogenous promoters.

Given the number of different insulators described in *Drosophila*, it is reasonable to assume that there will also be additional insulators within mosquito genomes. In general, Dipteran genomes are much more compact than vertebrate genomes, so insulators may actually be more critical for appropriate gene expression in "tighter quarters". Some of these insulators may be conserved in mosquitoes, others may be uncovered that are novel, perhaps related to the management of the cadre of genes

specific to anantogeny. It will be important to exploit the newly available sequence information to identify other potential candidates for investigation.

Identification of endogenous binding sites for CTCF would provide valuable clues to additional functions of this protein in mosquitoes. The limiting factor in this study was the polyclonal antibody produced. The antibody was produced prior to the realization that there were an additional 33 amino acid residues at the C-terminal end of the protein not predicted by the gene annotation algorithms. Perhaps this additional sequence or other regions of the endogenous protein partially obscure the dominant epitopes for the antibody when the protein folds. Vertebrate CTCF has been shown to bind DNA via combinatorial use of subsets of its zinc finger domains, so it was assumed that mosquito CTCF probably bound in a similar fashion. Another group also communicated that they had produced a good antibody in *Drosophila* against this region of the protein. However, little is known about the function of the C-terminal region of CTCF in any species. The antibody produced does recognize a protein of the expected molecular mass in Western blots, however the binding was relatively weak, even with embryo lysates. Also, the antibody was able to shift DNA-CTCF complexes only weakly in gel shift assays. Typically, the shifted band weakened or disappeared, rather than producing a more typical supershift, perhaps indicating that the major epitope recognized by the antibody overlapped part or all of the DNA binding domain when the protein was in its native conformation. If this is true, then the protein can potentially bind DNA and the antibody, but not both very efficiently. The binding site selection experiments were dependent upon the formation of a stable DNA-protein-antibody

complex. Chromatin Immunoprecipitation (ChIP) assays were repeatedly attempted with some of the genomic fragments and the putative serpin cluster targets. These did not produce consistent results. Finally, attempts to immunolocalize CTCF on meiotic chromosomes in *Ae. aegypti* with this antibody did not produce detectable signal. Taken together, the data indicate that this particular antibody may not be appropriate for investigating the DNA binding of mosquito CTCF.

In the future, the first priority would be to produce an antibody that binds the DNA-protein complex more efficiently. In the short term, the N-terminus of the protein could be tagged with an epitope and subjected to Western blot analysis with both cell and embryo lysates. The EMSAs with the known boundary elements, chicken 5'HS and *Drosophila Fab8*, could be repeated to assess its ability to shift the DNA-protein complex. If successful, a peptide antibody could be generated against this region of the protein. This antibody would be a valuable reagent to investigate the *in vivo* targets of mosquito CTCF via genome-wide ChIP experiments such as those done in *Drosophila* with Engrailed (Solano et al., 2003) and in humans with CTCF (Mukhopadhyay et al., 2004). Alternatively, the chromatin profiling method DamID could be used to identify CTCF targets in a manner similar to that recently done with the transcription factor Hairy in *Drosophila* (Bianchi-Frias et al., 2004). Confirmation of any identified targets however, will require an antibody capable of tightly binding the entire DNA-protein complex. Ultimately, it would be advantageous to identify specific genomic targets for this protein that might then be used to insulate a transgene.

Despite the difficulties outlined above, the identification of mosquito CTCF orthologues provides an avenue of much promise. Once thought to be exclusive to vertebrates, this protein likely possesses critical functions in mosquito development and in the establishment of chromatin domains via an insulator function. Knowledge of this protein will allow identification of target DNA insulating sequences for future use in mosquito transgenesis. Better transgenesis is pivotal in designing new control strategies for these potent disease vectors.

REFERENCES

Adelman, Z.N., Jasinskiene, N., Vally, K.J.M., Peek, C., Travanty, E.A., et al., 2004. Formation and loss of large, unstable tandem arrays of the *piggyBac* transposable element in the yellow fever mosquito, *Aedes aegypti.* Transgenic Res. 13, 411-425.

Akio, M., Chadee, D.D., Graham, D.H., Severson, D.W., 2004. Reinvestigation of an endogenous meiotic drive system in the mosquito, *Aedes aegypti* (Diptera: Culicidae). J. Med. Entomol. 41 (6), 1027-1033.

Allen, M.L., O'Brochta, D.A., Atkinson, P.W., Levesque, C.S., 2001. Stable, germ-line transformation of *Culex quinquefasciatus* (Diptera: Culicidae). J. Med. Entomol. 38 (5), 701-710.

Alphey, L., Andreasen, M., 2002. Dominant lethality and insect population control. Mol. Biochem. Parasitol. 121, 173-178.

Antes, T.J., Namciu, S.J., Fournier, R.E.K., Levy-Wilson, B., 2001. The 5' boundary of the human apolipoprotein B chromatin domain in intestinal cells. Biochemistry 40, 6731-6742.

Anxolabehere, D., Kidwell, M.G., Periquet, G., 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P*-elements. Mol. Biol. Evol. 5 (3), 252-269.

Arbeitman, M.N., Furlong, E., E. M., Imam, F., Johnson, E., Null, B.H., et al., 2002. Gene expression during the life cycle of *Drosophila melanogaster*. Science 297, 2270-2275.

Atkinson, P.W., James, A.A., 2002. Advances in Genetics. Elsevier Science, San Diego, CA.

Atkinson, P.W., Warren, W.D., O'Brochta, D.A., 1993. The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element in maize. Proc. Natl. Acad. Sci. U. S. A. 90, 9693-9697.

Bailey, T.L., Gribskov, M., 1998. Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14, 48-54.

Baniahmad, A., Steiner, C., Kohne, A.C., Renkawitz, R., 1990. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell 61, 505-514.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., et al., 2004. The Pfam protein families database. Nucleic Acids Res. 32 (Database Issue), D138-D141.

Becker, N., Ascher, K.R.S., 1998. The use of *Bacillus thuringiensis* subsp. *israelensis* (*Bti*) against mosquitoes, with special emphasis on the ecological impact. Isr. J. Entomol. 32, 63-69.

Beerntsen, B.T., James, A.A., Christensen, B.M., 2000. Genetics of mosquito vector competence. Microbiol. Mol. Biol. Rev. 64, 115-137.

Bell, A.C., Felsenfeld, G., 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. Nature 405, 482-485.

Bell, A.C., West, A.G., Felsenfeld, G., 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98, 387-396.

Bell, A.C., West, A.G., Felsenfeld, G., 2001. Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. Science 291, 447-450.

Berger, M.P., Munson, P.J., 1991. A novel randomized iterative strategy for aligning multiple protein sequences. Comput. Appl. Biosci. 7, 479-484.

Bhalla, S.C., 1968. White eye, a new sex-linked mutant of *Aedes aegypti*. J. Am. Mosq. Control Assoc. 28, 380-385.

Bianchi-Frias, D., Orian, A., Delrow, J.J., Vazquez, J., Rosales-Nieves, A.E., et al., 2004. Hairy transcriptional repression targets and cofactor recruitment in *Drosophila*. PLoS Biol. 2 (7), 975-990.

Blackman, R.K., Meselson, M., 1986. Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the *Drosophila hsp82* gene. J. Mol. Biol. 188, 499-515.

Blair, C.D., Adelman, Z.N., Olson, K.E., 2000. Molecular strategies for interrupting arthropod-borne virus transmission by mosquitoes. Clin. Microbiol. Rev. 13 (4), 651-661.

Braig, H.R., Yan, G., 2002. The spread of genetic constructs in natural insect populations. In: Letourneau, D.K., Burrows, B.E. (Eds.), Genetically Engineered Organisms, CRC Press, Boca Raton, FL, pp. 251-314.

Brooke, B.D., Hunt, R.H., Chandre, F., Carnevale, P., Coetzee, M., 2002. Stable chromosomal inversion polymorphisms and insecticide resistance in the malaria vector mosquito *Anopheles gambiae* (Diptera: Culicidae). J. Med. Entomol. 39 (4), 568-573.

Brown, A.E., Bugeon, L., Crisanti, A., Catteruccia, F., 2003. Stable and heritable gene silencing in the malaria vector *Anopheles stephensi*. Nucleic Acids Res. 31 (15), e85.

Bulger, M., Schubeler, D., Bender, M.A., Hamilton, J., Farrell, C.M., et al., 2003. A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse *β-globin* locus. Mol. Cell. Biol. 23 (15), 5234-5244.

Burke, L.J., Hollemann, T., Pieler, T., Renkawitz, R., 2002. Molecular cloning and expression of the chromatin insulator protein CTCF in *Xenopus laevis*. Mech. Dev. 113 (1), 95-98.

Byrd, K., Corces, V.G., 2003. Visualization of chromatin domains created by the *gypsy* insulator of *Drosophila*. J. Cell. Biol. 162 (4), 565-574.

Carlson, J., Afanasiev, B., Suchman, E., 2000. Densonucleosis viruses as transducing vectors for insects. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications, CRC Press, Boca Raton, FL, pp. 139-159.

Carson, D.D., Summers, M.D., Guarino, L.A., 1991. Transient expression of the *Autographa californica* nuclear polyhedrosis virus intermediate-early gene, *IE-N*, is regulated by three viral elements. J. Virol. 65, 945-951.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540-552.

Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., et al., 2000. Stable germ-line transformation of the malaria mosquito *Anopheles stephensi*. Nature 405, 959-962.

Centers for Disease Control, 2005. Dengue Fever. http://www.cdc.gov/ncidod/dvbid/dengue/.

Chandre, F., Darriet, F., Manga, L., Akogbeto, M., Faye, O., et al., 1999. Status of pyrethroid resistance in *Anopheles gambiae*, sensu lato. Bull. World Health Organ. 77, 230-234.

Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S., Lee, J.T., 2002. CTCF, a candidate trans-acting factor for X-inactivation choice. Science 295 (5553), 345-347.

Chapman, H.C., 1974. Biological control of mosquito larvae. Annu. Rev. Entomol. 19, 33-59.

Chapman, R.F., 1998. The Insects: Structure and Function. 4th Edition. Cambridge University Press, Cambridge, UK.

Choi, J., Guarino, L.A., 1995. The baculovirus transactivator *IE1* binds to viral enhancer elements in the absence of insect cell factors. J. Virol. 69, 4548-4551.

Chung, J.H., Whiteley, M., Felsenfeld, G., 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against postion effect in *Drosophila*. Cell 74 (3), 505-514.

Coates, C.J., 2005. Stable transformation of vector species. In: Marquardt, W.C., Black, I., William C., Freier, J.E., Hagedorn, H.H., Hemingway, J., Higgs, S., James, A.A., Kondratieff, B., Moore, C.G. (Eds.), Biology of Disease Vectors, Elsevier Academic Press, San Diego, CA, pp. 613-624.

Coates, C.J., Howells, A.J., O'Brochta, D.A., Atkinson, P.W., 1996. The 5' regulatory region from the *Drosophila pseudoobscura hsp82* gene results in a high level of reporter gene expression in *Lucilia cuprina* embryos. Gene 175, 199-201.

Coates, C.J., Jasinskiene, N., James, A.A., 1998. *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. Proc. Natl. Acad. Sci. U. S. A. 95, 3748-3751.

Coates, C.J., Jasinskiene, N., Morgan, D., Tosi, L.R., Beverley, S.M., et al., 2000. Purified mariner (*Mos*I) transposase catalyzes the integration of marked elements into the germ-line of the yellow fever mosquito, *Aedes aegypti*. Insect Biochem. Mol. Biol. 30, 1003-1008.

Coates, C.J., Jasinskiene, N., Pott, G.B., James, A.A., 1999. Promoter-directed expression of recombinant fire-fly luciferase in the salivary glands of *Hermes*-transformed *Aedes aegypti*. Gene 226, 317-325.

Collins, F.H., James, A.A., 1996. Genetic modification of mosquitoes. Science Med. 3, 52-61.

Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16 (22), 10881-10890.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 6, 1188-1190.

Curtis, C.F., 1976. Population replacement by means of cytoplasmic incompatibility, 2. Field cage experiments with overlapping generations. Bull. World Health Organ. 53, 107-119.

Curtis, C.F., Sinkins, S.P., 1998. *Wolbachia* as a means of driving genes into populations. Parasitol. 116, S111-S115.

Cuvier, O., Hart, C.M., Laemmli, U.K., 1998. Identification of a class of chromatin boundary elements. Mol. Cell Biol. 18 (12), 7478-7486.

Danielli, A., Kafatos, F.C., Loukeris, T.G., 2003. Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion. J. Biol. Chem. 278 (6), 4184-4193.

Davidson, A.E., Balciunas, D., Mohn, D., Shaffer, J., Hermanson, S., et al., 2003. Efficient gene delivery and gene expression in zebrafish using the *Sleeping Beauty* transposon. Dev. Biol. 263, 191-202.

DeGregorio, E., Han, S.-J., Lee, W.-J., Baek, M.-J., Osaki, T., et al., 2002. An immune-responsive serpin regulates the melanization cascade in *Drosophila*. Dev. Cell 3, 581-592.

Du, M., Beatty, L.G., Winjing, Z., Lew, J., Schoenherr, C., et al., 2003. Insulator and silencer sequences in the imprinted region of human chromosome 11p15.5. Hum. Mol. Genet. 12 (15), 1927-1939.

Dunn, K.L., Zhao, H., Davie, J.R., 2003. The insulator binding protein CTCF associates with the nuclear matrix. Exp. Cell Res. 288, 218-223.

Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., et al., 1990. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 87, 9923-9927.

Eldridge, B.F., 2005. Mosquitoes, the Culicidae. In: Marquardt, W.C., Black, I., William C., Freier, J.E., Hagedorn, H.H., Hemingway, J., Higgs, S., James, A.A., Kondratieff, B., Moore, C.G. (Eds.), Biology of Disease Vectors, Elsevier Academic Press, San Diego, CA, pp. 95-111.

Engels, W.R., 1989. *P*-elements in *Drosophila melanogaster*. In: Berg, D.E., Howe, M.M. (Eds.), Mobile DNA, American Society for Microbiology, Washington, DC, pp. 437-489.

Farrell, C., West, A.G., Felsenfeld, G., 2002. Conserved CTCF insulator elements flank the mouse and human *β–globin* loci. Mol. Cell. Biol. 22 (11), 3820-3831.

Farrell, C.M., Grinberg, A., Huang, S.P., Chen, D., Pichel, J.G., et al., 2000. A large upstream region is not necessary for gene expression or hypersensitive site

formation at the mouse *β–globin* locus. Proc. Natl. Acad. Sci. U. S. A. 97 (26), 14554-14559.

Fedoriw, A.M., Stein, P., Svoboda, P., Schultz, R.M., Bartolomei, M.S., 2003. Transgenic RNAi reveals essential function for CTCF in *H19* gene imprinting. Science 303 (5655), 238-240.

Felsenstein, J., 1989. PHYLIP-phylogeny interference package (Version 3.2). Cladistics 5, 164-166.

Felsenstein, J., 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol. 266, 418-427.

Filippova, G.N., Cheng, M.K., Moore, J.M., Truong, J.-P., Hu, Y.J., et al., 2005. Boundaries between chromosomal domains of X-inactivation and escape bind CTCF and lack CpG methylation during early development. Dev. Cell 8, 31-42.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., et al., 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian *c-myc* oncogenes. Mol. Cell. Biol. 16 (6), 2802-2813.

Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., et al., 2001. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the *DM1* locus. Nat. Genet. 28 (4), 335-343.

Finnegan, D.J., 1985. Transposable elements in eukaryotes. Int. Rev. Cytol. 93, 281-326.

Fitch, W.M., Margoliash, E., 1967. Construction of phylogenetic trees. Science 155 (760), 279-284.

Franz, G., Savakis, C., 1991. *Minos*, a new transposable element from *Drosophila hydei*, is a member of the *Tc-1*-like family of transposons. Nucleic Acids Res. 19 (6646).

Fraser, M.J., Ciszczon, T., Elick, T., Bauser, C., 1996. Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (IFP2) and *tagalong* (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. Insect Mol. Biol. 5, 141-151.

Gerasimova, T.I., Byrd, K., Corces, V.G., 2000. A chromatin insulator determines the nuclear localization of DNA. Mol. Cell 6, 1025-1035.

Geyer, P.K., Corces, V.G., 1992. DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. Genes Dev. 6, 1865-1873.

Gilchrist, B.M., Haldane, J.B.S., 1947. Sex-linkage and sex-determination in a mosquito, *Culex molestus*. Heredity 33, 175-190.

Goltsev, Y., Hsiaong, W., Lanzaro, G.C., Levine, M., 2004. Different combinations of gap repressors for common stripes in *Anopheles* and *Drosophila* embryos. Dev. Biol. 275 (2), 435-446.

Gombert, W.M., Farris, S.D., Rubio, E.D., Morey-Rosler, K.M., Schubach, W.H., et al., 2003. The *c-myc* insulator element and matrix attachment regions define the *c-myc* chromosomal domain. Mol. Cell. Biol. 23 (24), 9338-9348.

Gonzalez-Gaitan, M.A., Jackle, H., 1997. *In situ* localization of proteins in whole mounted tissue. In: Crampton, J.M., Beard, C.B., Louis, C. (Eds.), Molecular Biology of Insect Disease Vectors: A Methods Manual, Chapman & Hall, London, pp. 283-294.

Grewal, S.I.S., Elgin, S.C.R., 2002. Heterochromatin: new possibilities for the inheritance of structure. Curr. Opin. Genet. Dev. 12, 178-187.

Grossman, G.L., Rafferty, C.S., Clayton, J.R., Stevens, T.K., Mukabayire, O., et al., 2001. Germ-line transformation of the malaria vector, *Anopheles gambiae*, with the *piggyBac* transposable element. Insect Mol. Biol. 10, 597-604.

Guarino, L.A., Dong, W., 1991. Expression of an enhancer-binding protein in insect cells transfected with the *Autographa californica* nuclear polyhedrosis virus *IE1* gene. J. Virol. 65.

Guarino, L.A., Dong, W., 1994. Functional dissection of the *Autographa californica* nuclear polyhedrosis enhancer element *Hr5*. Virology 200, 328-335.

Guarino, L.A., Summers, M.D., 1986. Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. J. Virol. 60, 215-223.

Gubler, D.J., 1998. Factors influencing the emergence/resurgence of infectious diseases. Infect. Dis. Rev. 1 (1), 50-52.

Gubler, D.J., Clark, G.G., 1999. Dengue/Dengue hemorrhagic fever: the emergence of a global health problem. Dispatches 1 (2), 1-6.

Handler, A.M., Harrell, R.A., 2001. Transformation of the Caribbean fruit fly, *Anastrepha suspensa*, with a *piggyBac* vector marked with polyubiquitin-regulated GFP. Insect Biochem. Mol. Biol. 31, 199-205.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., et al., 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. Nature 405, 486-489.

Hartl, D.L., Hiraizumi, Y., Crow, J.F., 1967. Evidence for sperm dysfunction as the mechanism of segregation distortion in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 58, 2240-2245.

Hemingway, J., 2005. Biological control of mosquitoes. In: Marquardt, W.C., Black, I., William C., Freier, J.E., Hagedorn, H.H., Hemingway, J., Higgs, S., James, A.A., Kondratieff, B., Moore, C.G. (Eds.), Biology of Disease Vectors, Elsevier Academic Press, San Diego, CA, pp. 649-660.

Hemingway, J., Ranson, H., 2000. Insecticide resistance in insect vectors of human disease. Annu. Rev. Entomol. 45, 371-391.

Hemingway, J., Ranson, H., 2005. Chemical control of vectors and mechanisms of resistance. In: Marquardt, W.C., Black, I., William C., Freier, J.E., Hagedorn, H.H., Hemingway, J., Higgs, S., James, A.A., Kondratieff, B., Moore, C.G. (Eds.), Biology of Disease Vectors, Elsevier Academic Press, San Diego, CA, pp. 627-637.

Hickey, W.A., Craig, J., G. B., 1966. Genetic distortion of sex ratio in a mosquito, *Aedes aegypti*. Genetics 53, 1177-1196.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., et al., 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298, 129-149.

Horn, C., Wimmer, E.A., 2000. A versatile vector set for animal transgenesis. Dev. Genes Evol. 210, 630-637.

Horn, P.J., Peterson, C.L., 2002. Molecular biology. Chromatin higher order folding--wrapping up transcription. Science 297 (5588), 1824-1827.

Ishihara, K., Sasaki, H., 2002. An evolutionarily conserved putative insulator element near the 3' boundary of the imprinted *Igf2/H19* domain. Hum. Mol. Genet. 11 (14), 1627-1636.

Ishihara, S.L., Morohashi, K.-i., 2005. A boundary for histone acetylation allows distinct expression patterns of the *Ad4BP/SF-1* and *GCNF* loci in adrenal cortex cells. Biochem. Biophys. Res. Commun. 329, 554-562.

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, 452-455.

Jaga, K., Dharmani, C., 2003. Global surveillance of DDT and DDE levels in human tissues. Int. J. Occup. Med. Environ. Health 16 (1), 7-20.

James, A.A., 2000. Control of disease transmission through genetic modification of mosquitoes. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis, CRC Press, Boca Raton, FL, pp. 319-333.

Jasinskiene, N., Coates, C.J., Benedict, M.Q., Cornel, A.J., Salazar Rafferty, C., et al., 1998. Stable transformation of the yellow fever mosquito, *Aedes aegypti*, with the *Hermes* element from the housefly. Proc. Natl. Acad. Sci. U. S. A. 95, 3743-3747.

Kapetanaki, M.G., Loukeris, T.G., Livadaras, I., Savakis, C., 2002. High frequencies of *Minos* transposon mobilization are obtained in insects by using *in vitro* synthesized mRNA as a source of transposase. Nucleic Acids Res. 30, 3333-3340.

Kaya, H.K., Gaugler, R., 1993. Entomopathogenic nematodes. Annu. Rev. Entomol. 38, 181-206.

Kidwell, M.G., Frydryk, T., Novy, J.B., 1983. The hybrid dysgenesis potential of *Drosophila melanogaster* strains of diverse temporal and geographic origin. Drosophila Inf. Serv. 59, 63-69.

Kidwell, M.G., Lisch, D.R., 2000. Transposable elements and host genome evolution. Trends Ecol. Evol. 15 (3), 95-99.

Kim, W., Koo, H., Richman, A.M., Seeley, D., Vizioli, J., et al., 2004. Ectopic expression of a cecropin transgene in the human malaria vector mosquito *Anopheles gambiae* (Diptera: Culicidae): effects upon susceptibility to *Plasmodium*. J. Med. Entomol. 41 (3), 447-455.

Klenova, E.M., Nicolas, R.H., Carne, A.F., Lee, R.E., Lobanenkov, V.V., et al., 1997. Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS-PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene. Nucleic Acids. Res. 25, 466-474.

Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., et al., 1993. CTCF, a conserved nuclear factor required for chicken *c-myc* gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol. Cell. Biol. 13 (12), 7612-7624.

Knudson, D.L., Zheng, L., Gordon, S.W., Brown, S.E., Kafatos, F.C., 1996. Genome organization of vectors. In: Beatty, B.J., Marquardt, W.C. (Eds.), The Biology of Disease Vectors, University Press of Colorado, Fort Collins, CO, pp. 175-214.

Koella, J.C., Sorensen, F.L., Anderson, R.A., 1998. The malaria parasite, *Plasmodium falciparum*, increases the frequency of multiple feeding of its mosquito vector, *Anopheles gambiae*. Proc. Natl. Acad. Sci. U. S. A. 265, 763-768.

Kokoza, V., Ahmed, A., Wimmer, E.A., Raikhel, A.S., 2001. Efficient transformation of the yellow fever mosquito *Aedes aegypti* using the *piggyBac* transposable element vector pBac[3xP3-EGFP afm]. Insect Biochem. Mol. Biol. 31, 1137-1143.

Kovacs, G.R., Guarino, L.A., Summers, M.D., 1991. Novel regulatory properties of the *IE1* and *IE0* transactivators encoded by the baculovirus *Autographa californica* multicapsid nuclear polyhedrosis virus. J. Virol. 65, 5281-5288.

Kremer, A., Knebel-Mordsdorf, D., 1998. The early baculovirus *he65* promoter: on the mechanism of transcriptional activation by IE1. Virology 249, 336-351.

Ladomery, M., Dellaire, G., 2002. Multifunctional zinc finger proteins in development and disease. Ann. Hum. Genet. 66 (Pt 5-6), 331-342.

Larget, B., Simon, D., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetics trees. Mol. Biol. Evol. 16, 750-759.

Lee, C., Grasso, C., Sharlow, M., 2002. Multiple sequence alignment using partial order graphs. Bioinformatics 18, 452-464.

Lee, H.S., Simon, J.A., Lis, J.T., 1988. Structure and expression of ubiquitin genes of *Drosophila melanogaster*. Mol. Cell. Biol. 8, 4727-4735.

Lee, J.T., 2003. Molecular links between X-inactivation and autosomal imprinting: X-inactivation as a driving force for the evolution of imprinting? Curr. Biol. 13 (6), R242-254.

Legner, E.F., 1995. Biological control of Diptera of medical and veterinary importance. J. Vector Ecol. 20, 59-120.

Leisy, D.J., Rasmussen, C., Owusu, E.O., Rohrmann, G.F., 1997. A mechanism for negative gene regulation in *Autographa californica* multinucleocapsis nuclear polyhedrosis virus. J. Virol. 71, 5088-5094.

Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., et al., 2004. SMART 4.0: towards genomic data integration. Nucleic Acids Res. 32 (Database Issue), D142-D144.

Lewis, A., Murrell, A., 2004. Genomic imprinting: CTCF protects the boundaries. Curr. Biol. 14, R284-R286.

Li, T., Lu, Z., Lu, L., 2004. Regulation of eye development by transcriptional control of CCCTC binding factor (CTCF). J. Biol. Chem. 279 (26), 27575-27583.

Lindquist, D.A., Abusowa, M., Hall, M.J., 1992. The New World screwworm fly in Libya: a review of its introduction and eradication. Med. Vet. Entomol. 6 (1), 2-8.

Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., et al., 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken *c-myc* gene. Oncogene 5, 1743-1753.

Lobo, N.F., Hua-Van, A., Li, X., Nolen, B.M., Fraser Jr., M.J., 2002. Germ-line transformation of the yellow fever mosquito, *Aedes aegypti*, mediated by transpositional insertion of a *piggyBac* vector. Insect Mol. Biol. 11, 133-139.

Lu, M., Farrell, P.J., Johnson, R.R., Iatrou, K., 1997. A baculovirus (*Bombyx mori* nuclear polyhedrosis virus) repeat element functions as a powerful constitutive enhancer in transfected insect cells. J. Biol. Chem. 272, 30724-30728.

Lu, M., Johnson, R.R., Iatrou, K., 1996. *Trans*-activation of a cell housekeeping gene promoter by the *ie1* gene product of baculoviruses. Virology 218, 103-113.

Lutz, M., Burke, L.J., LeFevre, P., Myers, F.A., Thorne, A.W., et al., 2003. Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor. EMBO J. 22 (7), 1579-1587.

Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., et al., 2003. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res. 31, 383-387.

Marhold, J., Rothe, N., Pauli, A., Mund, C., Kuehle, K., et al., 2004. Conservation of DNA methylation in dipteran insects. Insect Mol. Biol. 13 (2), 117-123.

Martin, D., Wang, S.-F., Raikhel, A.S., 2001. The vitellogenin gene of the mosquito *Aedes aegypti* is a direct target of ecdysteroid receptor. Mol. Cell. Endocrinol. 173, 75-86.

Matyunina, L.V., Jordan, I.K., McDonald, J.F., 1996. Naturally occurring variation in *copia* expression is due to both element (*cis*) and host (*trans*) regulatory variation. Proc. Natl. Acad. Sci. U. S. A. 93 (7097-7102).

McClintock, B., 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. U. S. A. 36 (6), 344-355.

Medhora, M., Maruyama, K., Hartl, D.L., 1991. Molecular and functional analysis of the *mariner* mutator element *MosI* in *Drosophila*. Genetics 128, 311-318.

Meredith, S.E., James, A.A., 1990. Biotechnology as applied to vectors and vector control. Ann. Parasitol. Hum. Comp. 65 (Suppl. 1), 113-118.

Moon, H., Filippova, G.N., Loukinov, D.I., Pugacheva, E., Chen, Q., et al., 2005. CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the *Fab-8* insulator. EMBO Rep. 6 (2), 165-170.

Moreau, P., R., H., Wasylyk, B., Everett, R., Gaub, M.P., et al., 1981. The SV40 72 base pair repeat has a striking effect on gene expression in both SV40 and other chimeric recombinants. Nucleic Acids Res. 9, 6047-6068.

Moreira, L.A., Ito, J., Ghosh, A., Davenport, M., Zieler, H., et al., 2002. Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. J. Biol. Chem. 277 (43), 40839-40843.

Mukhopadhyay, R., Yusufzai, T.M., Yu, W., Whitehead, J., Xu, J., et al., 2004. The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. Genome Res. 14, 1594-1602.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., et al., 2003. The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res. 31, 315-318.

Muller, J.-M., Dimopoulos, G., Blass, C., Kafatos, F.C., 1999. A hemocyte-like cell line established from the malaria vector *Anopheles gambiae* expresses six prophenoloxidase genes. J. Biol. Chem. 274 (17), 11727-11735.

Namciu, S.J., Blochlinger, K.B., Fournier, R.E., 1998. Human matrix attachment regions insulate transgene expression from chromosomal position effects in *Drosophila melanogaster*. Mol. Cell. Biol. 18, 2382-2391.

Newton, M.E., Wood, R.J., Southern, D.I., 1976. A cytogenetic analysis of meiotic drive in the mosquito *Aedes aegypti* (L.). Genetica 48, 137-143.

Nirmala, X., James, A.A., 2003. Engineering *Plasmodium*-refractory phenotypes in mosquitoes. Trends Parasitol. 19, 384-387.

Nolan, T., Bower, T.M., Brown, A.E., Crisanti, A., Catteruccia, F., 2002. *piggyBac*-mediated germ-line transformation of the malaria mosquito *Anopheles stephensi* using the red fluorescent protein dsRED as a selectable marker. J. Biol. Chem. 277, 8759-8762.

Notredame, C., Higgins, S., Heringa, J., 2000. T-Coffee: a novel method for multiple sequence alignments. J. Mol. Biol. 302, 205-217.

O'Brochta, D.A., Sethuraman, N., Wilson, R., Hice, R.H., Pinkerton, A.C., et al., 2003. Gene vector and transposable element behavior in mosquitoes. J. Exp. Biol. 206, 3823-3834.

Ohlsson, R., Renkawitz, R., Lobanenkov, V., 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet. 17 (9), 520-527.

Olson, K.E., Carlson, J.O., Blair, C.D., 2005. Virus-induced, transient expression of genes in mosquitoes. In: Marquardt, W.C., Black, I., William C., Freier, J.E., Hagedorn, H.H., Hemingway, J., Higgs, S., James, A.A., Kondratieff, B., Moore, C.G. (Eds.), Biology of Disease Vectors, Elsevier Academic Press, San Diego, CA, pp. 601-611.

Pabo, C.O., Nekludova, L., 2000. Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? J. Mol. Biol. 301, 597-624.

Palla, F., Melfi, R., Anello, L., Di Bernardo, M., Spinelli, G., 1997. Enhancer blocking activity located near the 3' end of the sea urchin early *H2A* histone gene. Proc. Natl. Acad. Sci. U. S. A. 94, 2272-2277.

Palopoli, M.F., Wu, C.-I., 1996. Rapid evolution of a coadapted gene complex: evidence from the *Segregation Distorter* (*SD*) system of meiotic drive in *Drosophila melanogaster*. Genetics 143, 1675-1688.

Pant, V., Mariano, P., Kanduri, C., Mattsson, A., Lobanenkov, V., et al., 2003. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the *H19* imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. Genes Dev. 17, 586-590.

Perera, O.P., Harrell, R.A., Handler, A.M., 2002. Germ-line transformation of the South American malaria vector, *Anopheles albimanus*, with a *piggyBac/EGFP* transposon vector is routine and highly efficient. Insect Mol. Biol. 11, 291-297.

Peterson, K.J., Lyons, J.B., Nowak, K., S., Takacs, C.M., Wargo, M.J., et al., 2004. Estimating metazoan divergence times with a molecular clock. Proc. Natl. Acad. Sci. U. S. A. 101 (17), 6536-6541.

Pinkerton, A.C., Michel, K., O'Brochta, D.A., Atkinson, P.W., 2000. Green fluorescent protein as a genetic marker in transgenic *Aedes aegypti*. Insect Mol. Biol. 9, 1-10.

Pollock, R.M., 1996. Determination of protein-DNA sequence specificity by PCR-assisted binding-site selection. In: Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (Eds.), Current Protocols in Molecular Biology, John Wiley & Sons, Inc., Edison, NJ, pp. 12.11.11-12.11.11.

Pugacheva, E.M., Tiwari, V.K., Abdullaev, Z., Vostrov, A.A., Flanagan, P.T., et al., 2005. Familial cases of point mutations in the *XIST* promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. Hum. Mol. Genet. 14 (7), 953-965.

Pullen, S.S., Friesen, P.D., 1995. Early transcription of the *IE-1* transregulator gene of *Autographa californica* nuclear polyhedrosis virus is regulated by DNA sequences within its 5' noncoding leader region. J. Virol. 69, 156-165.

Raminani, L.N., Cupp, E.W., 1975. Early embryology of *Aedes aegypti* (L.) (Diptera: Culicidae). Int. J. Insect Morphol. Embryol. 4 (6), 517-528.

Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., et al., 2002. Position-effect protection and enhancer blocking by the chicken *β–globin* insulator are separable activities. Proc. Natl. Acad. Sci. U. S. A. 99 (10), 6883-6888.

Ribeiro, J.M.C., Kidwell, M.G., 1993. Transposable elements as population drive mechanisms: specification of critical parameter values. J. Med. Entomol. 31 (1), 10-16.

Rio, D.C., 1990. Molecular mechanisms regulating *Drosophila P*-element transposition. Annu. Rev. Genet. 24, 543-578.

Robinett, C.C., O'Connor, A., Dunaway, M., 1997. The repeat organizer, a specialized insulator element within the intergenic spacer of the *Xenopus* rRNA genes. Mol. Cell. Biol. 17, 2866-2875.

Robinson, A.S., Franz, G., Atkinson, P.W., 2004. Insect transgenesis and its potential role in agriculture and human health. Insect Biochem. Mol. Biol. 34, 113-120.

Rodems, S.M., Friesen, P.D., 1995. Transcriptional enhancer activity of *hr5* requires dual-palindrome half sites that mediate binding of a dimeric form of the baculovirus transregulator IE1. J. Virol. 69, 5368-5375.

Rose, T., Henikoff, J., Henikoff, S., 2003. CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. Nucleic Acids. Res. 31 (13), 3763-3766.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., et al., 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res. 26 (7), 1628-1635.

Roth, F.P., Hughes, J.D., Estep, P.W., Church, G.M., 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat. Biotechnol. 16, 939-945.

Rowan, K.H., Orsetti, J., Atkinson, P.W., O'Brochta, D.A., 2004. *Tn5* as an insect gene vector. Insect Biochem. Mol. Biol. 34, 695-705.

Rubin, G.M., Spradling, A.C., 1982. Genetic transformation of *Drosophila* with transposable element vectors. Science 218, 348-353.

Rushlow, C., 2004. Dorsoventral patterning: a serpin pinned down at last. Curr. Biol. 14, R16-R18.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4 (4), 406-425.

Sanchez-Vargas, I., Travanty, E.A., Keene, K.M., Franz, A.W.E., Beaty, B.J., et al., 2004. RNA interference, arthropod-borne viruses, and mosquitoes. Virus Res. 102, 65-74.

Sarker, A., Coates, C.J., Whyard, S., Willhoeft, U., Atkinson, P.W., et al., 1997. The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. Genetica 99, 15-29.

Schoenherr, C.J., Levorse, J.M., Tilghman, S.M., 2003. CTCF maintains differential methylation at the *Igf2/H19* locus. Nat. Genet. 33, 66-69.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. U. S. A. 95 (11), 5857-5864.

Segal, D.J., 2002. The use of zinc finger peptides to study the role of specific factor binding sites in the chromatin environment. Methods 26, 76-83.

Severson, D.W., Knudson, D.L., Soares, M.B., Loftus, B.J., 2004. *Aedes aegypti* genomics. Insect Biochem. Mol. Biol. 34, 715-721.

Shih, K.M., Gerenday, A., Fallon, A.M., 1998. Culture of mosquito cells in Eagle's medium. *In Vitro* Cell. Dev. Biol. Anim. 34, 629-630.

Sinkins, S.P., 2004. *Wolbachia* and cytoplasmic incompatibility in mosquitoes. Insect Biochem. Mol. Biol. 34, 723-729.

Sinkins, S.P., Hastings, I.M., 2004. Male-specific insecticide resistance and mosquito transgene dispersal. Trends Parasitol. 20 (9), 413-416.

Sinkins, S.P., O'Neill, S.L., 2000. *Wolbachia* as a vehicle to modify insect populations. In: Handler, A.M., James, A.A. (Eds.), Insect Transgenesis: Methods and Applications, CRC Press, Boca Raton, FL, pp. 271-287.

Slack, J.M., Blissard, G.W., 1997. Identification of two independent transcriptional activation domains in the *Autographa californica* multicapsid nuclear polyhedrosis virus IE1 protein. J. Virol. 71, 9579-9587.

Small, G.J., 2005. Environmental management for vector control. In: Marquardt, W.C., Black, I., William C., Freier, J.E., Hagedorn, H.H., Hemingway, J., Higgs, S., James, A.A., Kondratieff, B., Moore, C.G. (Eds.), Biology of Disease Vectors, Elsevier Academic Press, San Diego, CA, pp. 639-647.

Solano, P.J., Mugat, B., Martin, D., Girard, F., Huibant, J.-M., et al., 2003. Genome-wide identification of *in vivo Drosophila* Engrailed-binding DNA fragments and related target genes. Development 130, 1243-1254.

Spradling, A., 1986. *P*-element-mediated transformation. In: Roberts, D.B. (Ed.), *Drosophila*: A Practical Approach, IRL Press, Oxford, pp. 175-197.

Stephenson, J.R., 2005. Understanding Dengue pathogenesis: implications for vaccine design. Bull. World Health Organ. 83, 308-314.

Stoye, J., Perrey, S.W., Dress, A.W.M., 1997. Improving the divide-and-conquer approach to sum-of-pairs multiple sequence alignment. Appl. Math. Lett. 10 (2), 67-73.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964-969.

Sun, J.-M., Chen, H.Y., Davie, J.R., 2001. Effect of estradiol on histone acetylation dynamics in human breast cancer cells. J. Biol. Chem. 276 (52), 49435-49442.

Suter, B., Steward, R., 1991. Requirement for phosphorylation and localization of the Bicaudal-D protein in *Drosophila* oocyte differentiation. Cell 67, 917-926.

Suzuki, M., Gerstein, M., Yagi, N., 1994. Sterochemical basis of DNA recognition by Zn fingers. Nucleic Acids Res. 22 (16), 3397-3405.

Thompson, J.D., Thierry, J.C., Poch, O., 2003. RASCAL: rapid scanning and correction of multiple sequence alignment programs. Bioinformatics 19 (9), 1155-1161.

Thummel, C.S., Boulet, A.M., Lipshitz, H.D., 1988. Vectors for *Drosophila P*-element-mediated transformation and tissue culture transfection. Gene 74, 445-456.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., et al., 2002. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. Genome Biol. 3 (2), research0088.0081-0088.0014.

Travanty, E.A., Adelman, Z.N., Franz, A.W., Keene, K.M., Beaty, B.J., et al., 2004. Using RNA interference to develop Dengue virus resistance in genetically modified *Aedes aegypti*. Insect Biochem. Mol. Biol. 34 (7), 607-613.

Tu, Z., Coates, C.J., 2004. Mosquito transposable elements. Insect Biochem. Mol. Biol. 34, 631-644.

Udvardy, A., Maine, E., 1985. The *87A7* chromomere: identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. J. Mol. Biol. 185, 341-358.

Villalva, C., Touriol, C., Seurat, P., Trempat, P., Delsol, G., et al., 2001. Increased yield of PCR products by addition of T4 gene 32 protein to the SMART™ PCR cDNA synthesis system. Biotechniques 31, 81-86.

Warren, A.M., Crampton, J.M., 1991. The *Aedes aegypti* genome: complexity and organization. Genet. Res. 58, 225-232.

West, A.G., Fraser, P., 2005. Remote control of gene transcription. Hum. Mol. Genet. 14 (Review Issue 1), R101-R111.

West, A.G., Gaszner, M., Felsenfeld, G., 2002. Insulators: many functions, many mechanisms. Genes Dev. 16 (3), 271-288.

Wilson, C., Bellen, H.J., Gehring, W.J., 1990. Postion effects on eukaryotic gene expression. Annu. Rev. Cell Biol. 6, 679-714.

Wilson, S., Matyunina, L.V., McDonald, J.F., 1998. An enhancer region within the *copia* untranslated leader contains binding sites for *Drosophila* regulatory proteins. Gene 209, 237-246.

Wood, R.J., 2005. Genetic control of vectors. In: Marquardt, W.C., Black, I., William C., Freier, J.E., Hagedorn, H.H., Hemingway, J., Higgs, S., James, A.A., Kondratieff, B., Moore, C.G. (Eds.), Biology of Disease Vectors, Elsevier Academic Press, San Diego, CA, pp. 661-669.

Wood, R.J., Newton, M.E., 1991. Sex-ratio distortion carried by meiotic drive in mosquitoes. Am. Nat. 137, 379-391.

Wood, R.J., Ouda, N.A., 1987. The genetic basis of resistance and sensitivity to the meiotic drive gene (*D*) in the mosquito *Aedes aegypti*. Genetica 74, 69-79.

World Health Organization, 1982. Manual on Environmental Management for Vector Control, Geneva, Switzerland.

World Health Organization, 2002. Malaria Frequently Asked Questions. http://mosquito.who.int/docs/MalariaFAQ.htm.

World Health Organization, 2003. Malaria in Africa Infosheet. http://www.rbm.who.int/cmc_upload/0/000/015/370/RBMInfosheet_3.htm.

Wylie, A.A., Murphy, S.K., Orton, T.C., Jirtle, R.L., 2000. Novel imprinted *DLK1/GTL2* domain on human chromosome 14 contains motifs that mimic those implicated in *IGF2/H19* regulation. Genome Res. 10 (11), 1711-1718.

Wyss, J.H., 2000. Screwworm eradication in the Americas. Ann. N. Y. Acad. Sci. 916, 186-193.

Ye, Q., Worman, H.J., 1996. Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to *Drosophila* HP1. J. Biol. Chem. 271 (25), 14653-14656.

Yusufzai, T.M., Felsenfeld, G., 2004. The 5'-HS4 chicken *β-globin* insulator is a CTCF-dependent nuclear matrix-associated element. Proc. Natl. Acad. Sci. U. S. A. 101 (23), 8620-8624.

Yusufzai, T.M., Tagami, H., Nakatani, Y., Felsenfeld, G., 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell 13, 291-298.

Zhang, R., Burke, L.J., Rasko, J.E., Lobanenkov, V., Renkawitz, R., 2004. Dynamic association of the mammalian insulator protein CTCF with centrosomes and the midbody. Exp. Cell Res. 294, 86-93.

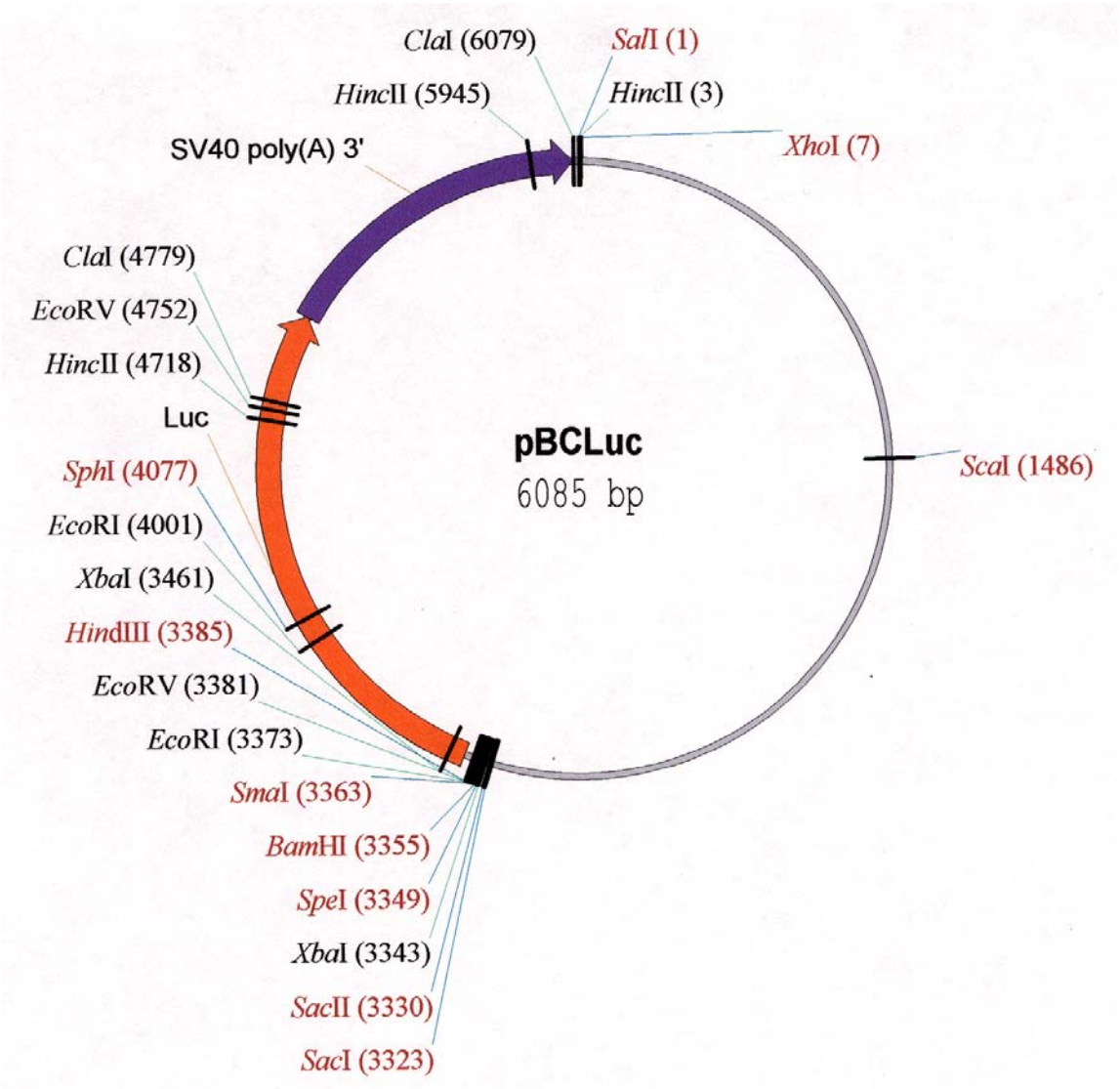
Zhong, X.P., Krangel, M.S., 1997. An enhancer-blocking element between α and δ gene segments within the human T cell receptor α/δ locus. Proc. Natl. Acad. Sci. U. S. A. 94, 5219-5224.

Zwiebel, L.J., Takken, W., 2004. Olfactor regulation of mosquito-host interactions. Insect Biochem. Mol. Biol. 34 (7), 645-652.

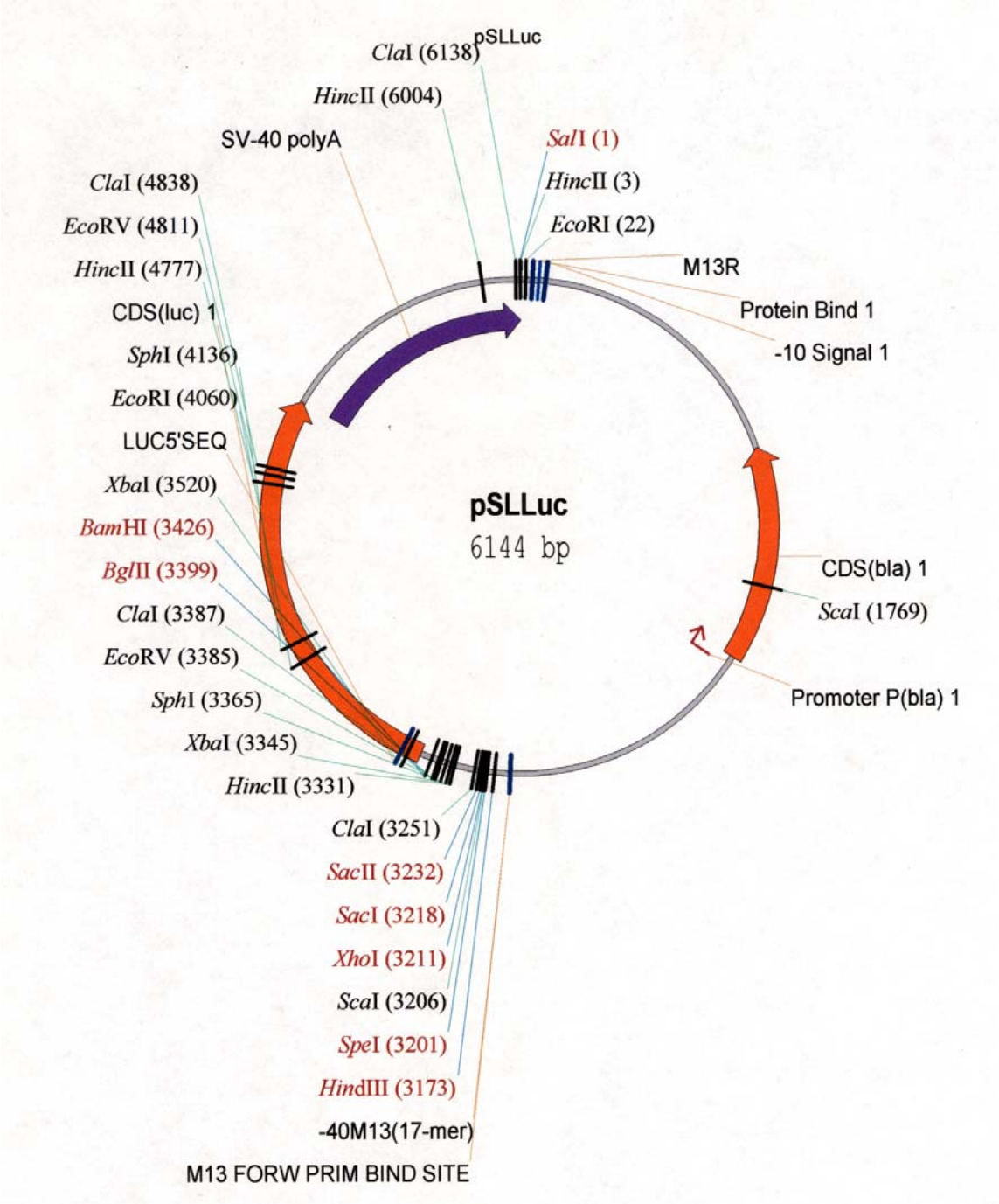
APPENDIX A

PLASMID MAPS

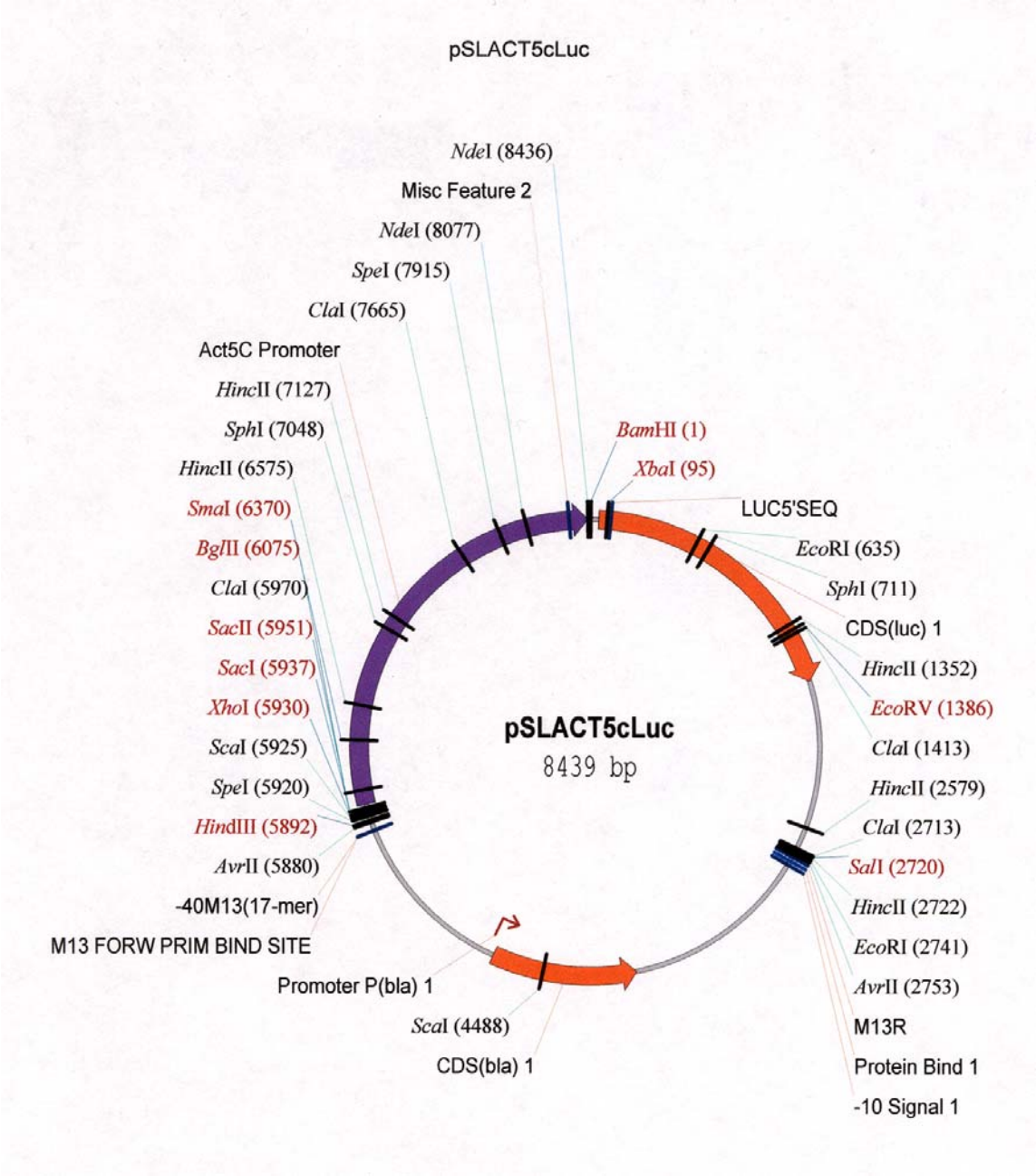
A-1


ClaI (6079)
SalI (1)
HincII (5945)
HincII (3)
SV40 poly(A) 3'
XhoI (7)
ClaI (4779)
EcoRV (4752)
HincII (4718)
Luc
pBCLuc
6085 bp
SphI (4077)
ScaI (1486)
EcoRI (4001)
XbaI (3461)
HindIII (3385)
EcoRV (3381)
EcoRI (3373)
SmaI (3363)
BamHI (3355)
SpeI (3349)
XbaI (3343)
SacII (3330)
SacI (3323)

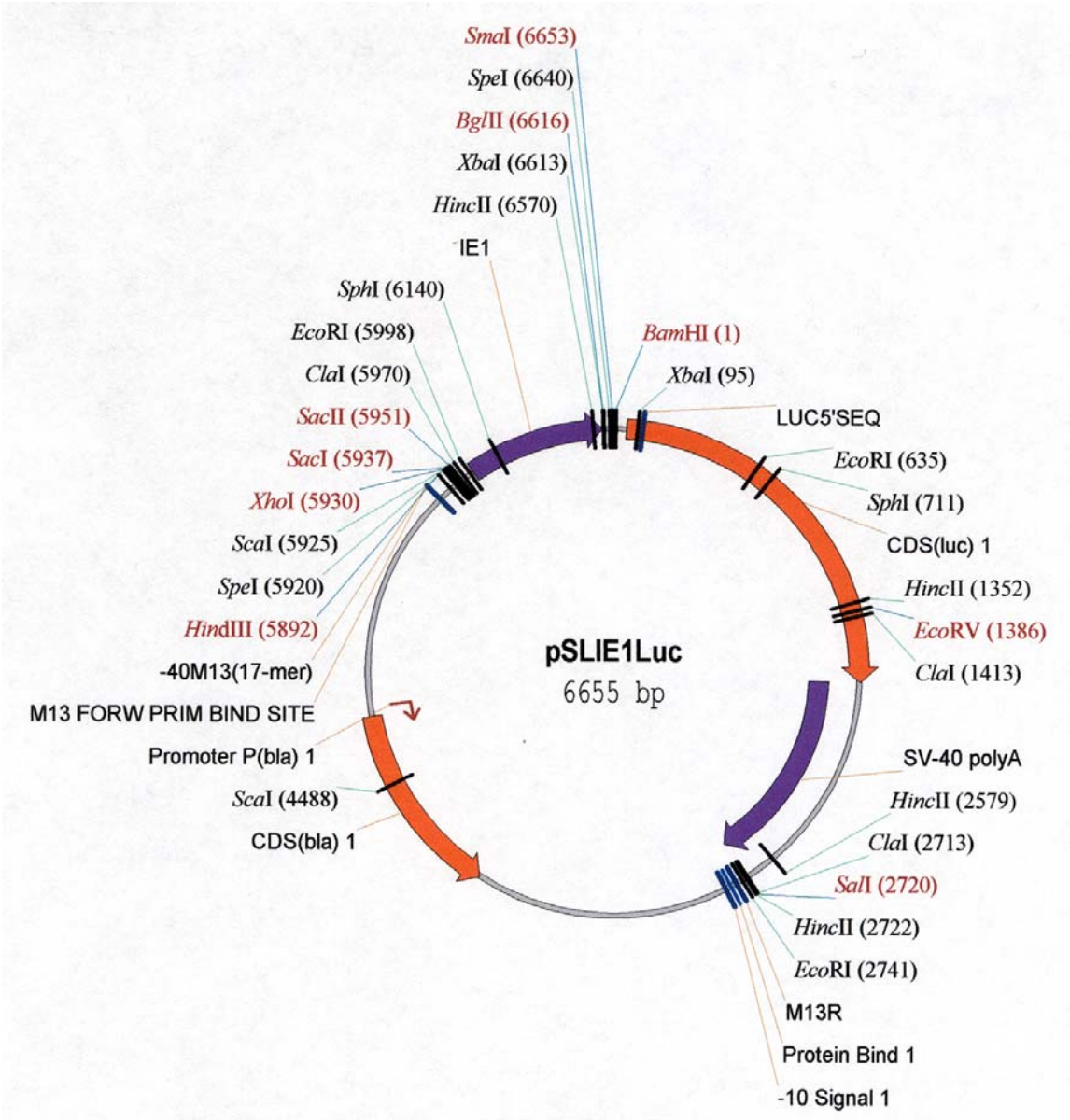
A-2


pSLLuc
ClaI (6138)
HincII (6004)
SV-40 polyA
SalI (1)
HincII (3)
EcoRI (22)
M13R
Protein Bind 1
-10 Signal 1
ClaI (4838)
EcoRV (4811)
HincII (4777)
CDS(luc) 1
SphI (4136)
EcoRI (4060)
LUC5'SEQ
XbaI (3520)
BamHI (3426)
BglII (3399)
ClaI (3387)
EcoRV (3385)
SphI (3365)
XbaI (3345)
HincII (3331)
pSLLuc
6144 bp
CDS(bla) 1
ScaI (1769)
Promoter P(bla) 1
ClaI (3251)
SacII (3232)
SacI (3218)
XhoI (3211)
ScaI (3206)
SpeI (3201)
HindIII (3173)
-40M13(17-mer)
M13 FORW PRIM BIND SITE

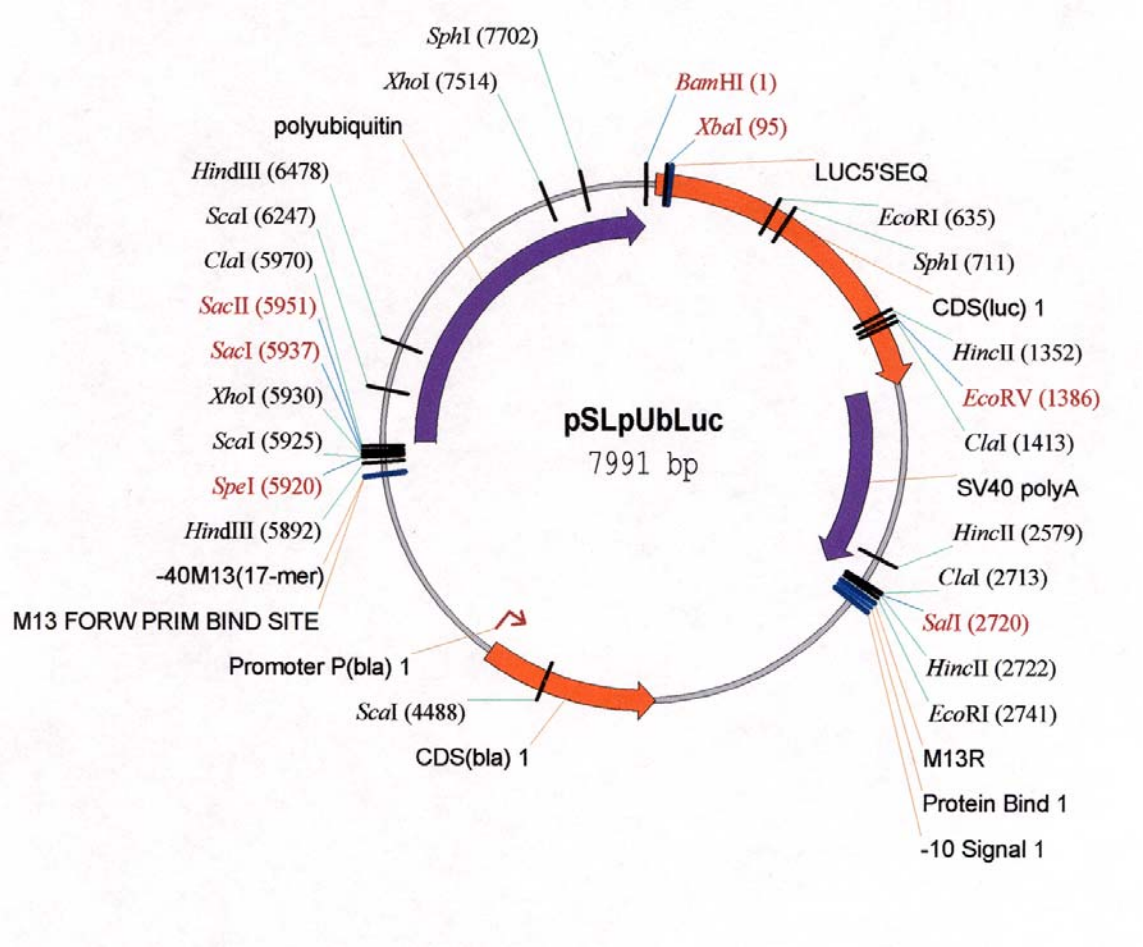
A-3


pSLACT5cLuc
NdeI (8436)
Misc Feature 2
NdeI (8077)
SpeI (7915)
ClaI (7665)
Act5C Promoter
HincII (7127)
SphI (7048)
HincII (6575)
SmaI (6370)
BglII (6075)
ClaI (5970)
SacII (5951)
SacI (5937)
XhoI (5930)
ScaI (5925)
SpeI (5920)
HindIII (5892)
AvrII (5880)
-40M13(17-mer)
M13 FORW PRIM BIND SITE
Promoter P(bla) 1
ScaI (4488)
CDS(bla) 1
BamHI (1)
XbaI (95)
LUC5'SEQ
EcoRI (635)
SphI (711)
CDS(luc) 1
HincII (1352)
EcoRV (1386)
ClaI (1413)
HincII (2579)
ClaI (2713)
SalI (2720)
HincII (2722)
EcoRI (2741)
AvrII (2753)
M13R
Protein Bind 1
-10 Signal 1
pSLACT5cLuc
8439 bp

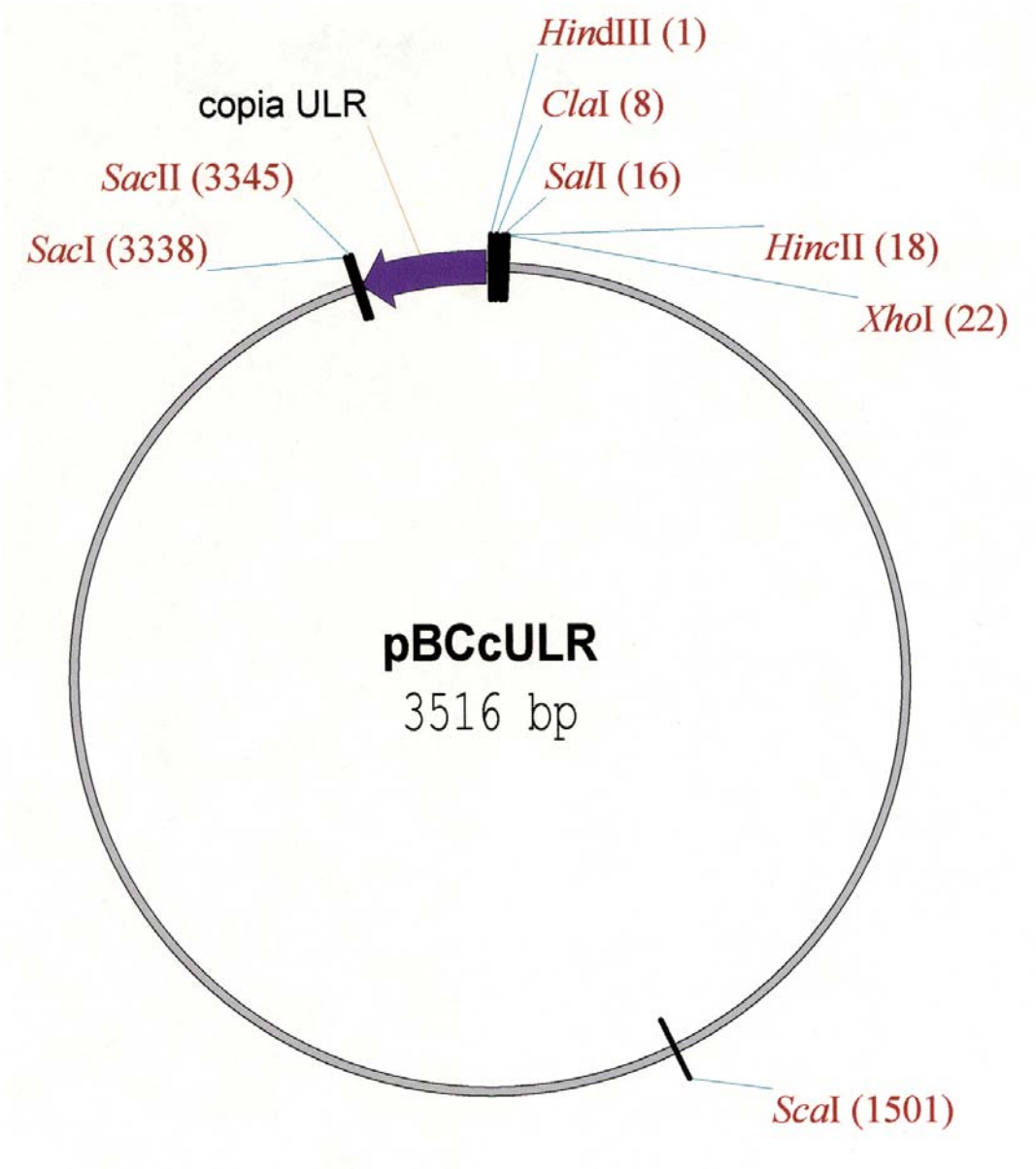
A-4


SmaI (6653)
SpeI (6640)
BglII (6616)
XbaI (6613)
HincII (6570)
IE1
SphI (6140)
EcoRI (5998)
ClaI (5970)
SacII (5951)
SacI (5937)
XhoI (5930)
ScaI (5925)
SpeI (5920)
HindIII (5892)
-40M13(17-mer)
M13 FORW PRIM BIND SITE
Promoter P(bla) 1
ScaI (4488)
CDS(bla) 1
BamHI (1)
XbaI (95)
LUC5'SEQ
EcoRI (635)
SphI (711)
CDS(luc) 1
HincII (1352)
EcoRV (1386)
ClaI (1413)
pSLIE1Luc
6655 bp
SV-40 polyA
HincII (2579)
ClaI (2713)
SalI (2720)
HincII (2722)
EcoRI (2741)
M13R
Protein Bind 1
-10 Signal 1

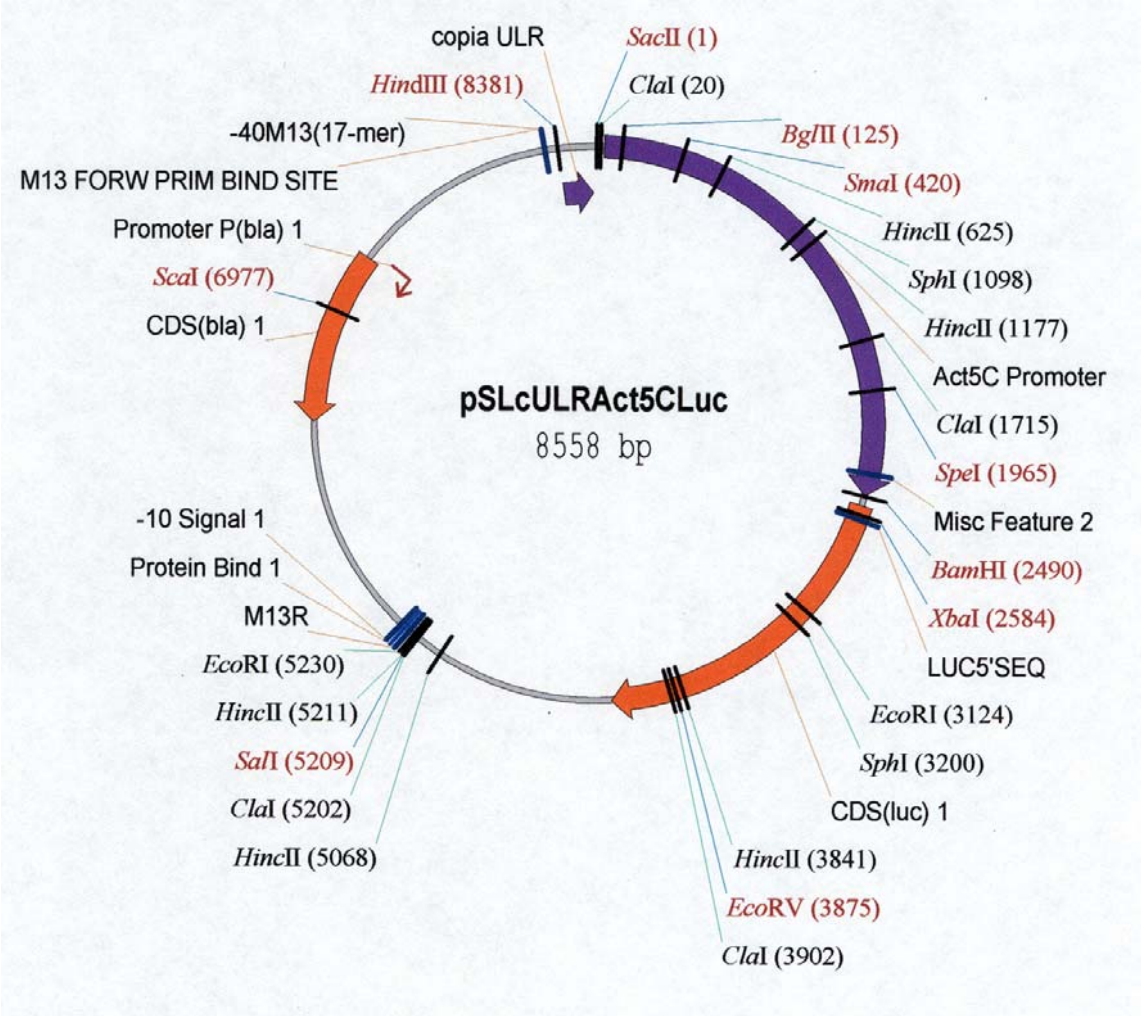
A-5


SphI (7702)
XhoI (7514)
BamHI (1)
XbaI (95)
polyubiquitin
HindIII (6478)
LUC5'SEQ
ScaI (6247)
EcoRI (635)
ClaI (5970)
SphI (711)
SacII (5951)
CDS(luc) 1
SacI (5937)
HincII (1352)
XhoI (5930)
EcoRV (1386)
pSLpUbLuc
ScaI (5925)
7991 bp
ClaI (1413)
SpeI (5920)
SV40 polyA
HindIII (5892)
HincII (2579)
-40M13(17-mer)
ClaI (2713)
M13 FORW PRIM BIND SITE
SalI (2720)
Promoter P(bla) 1
HincII (2722)
ScaI (4488)
EcoRI (2741)
CDS(bla) 1
M13R
Protein Bind 1
-10 Signal 1

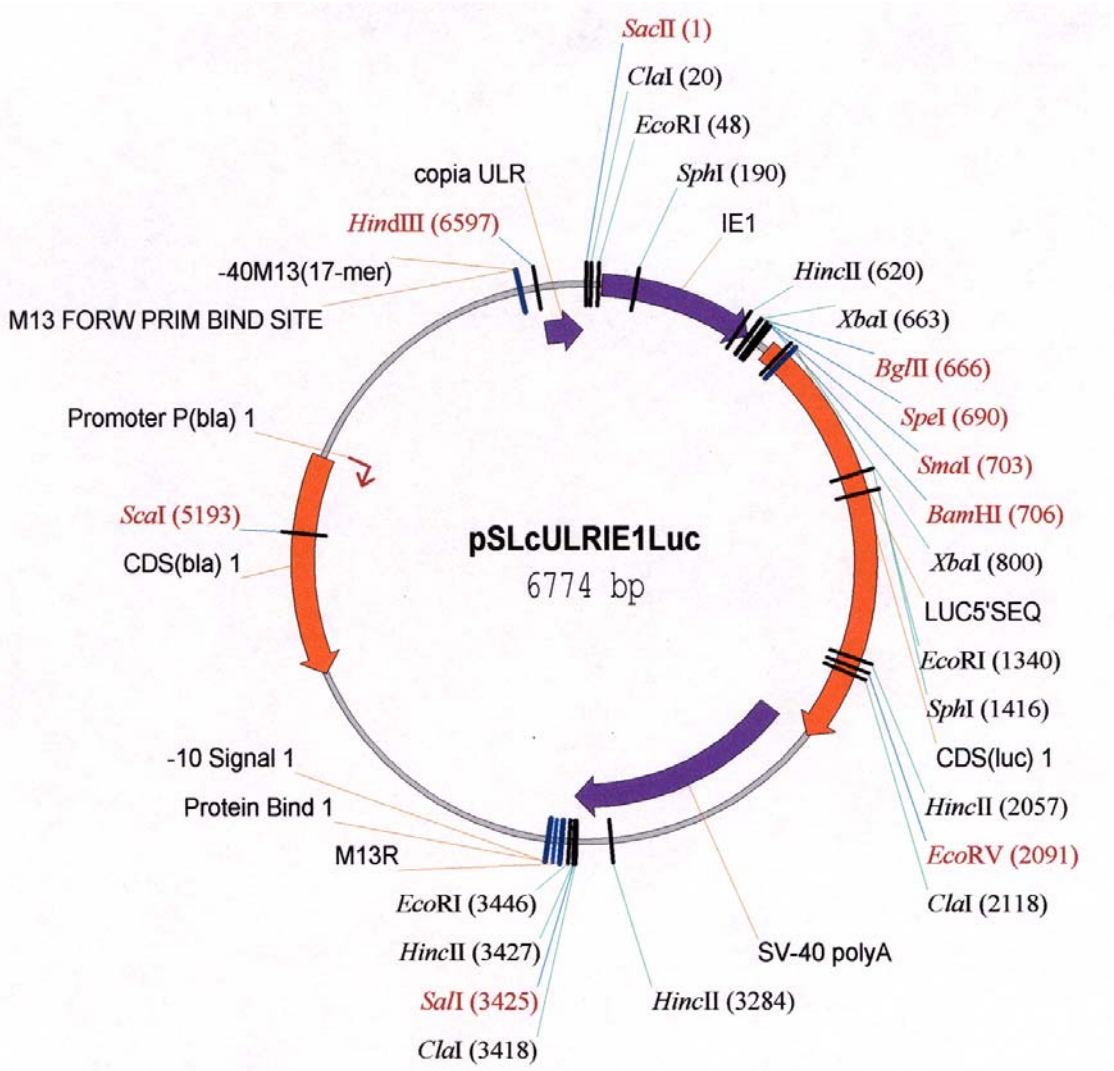
A-6


HindIII (1)
copia ULR
ClaI (8)
SacII (3345)
SalI (16)
SacI (3338)
HincII (18)
XhoI (22)
pBCcULR
3516 bp
ScaI (1501)

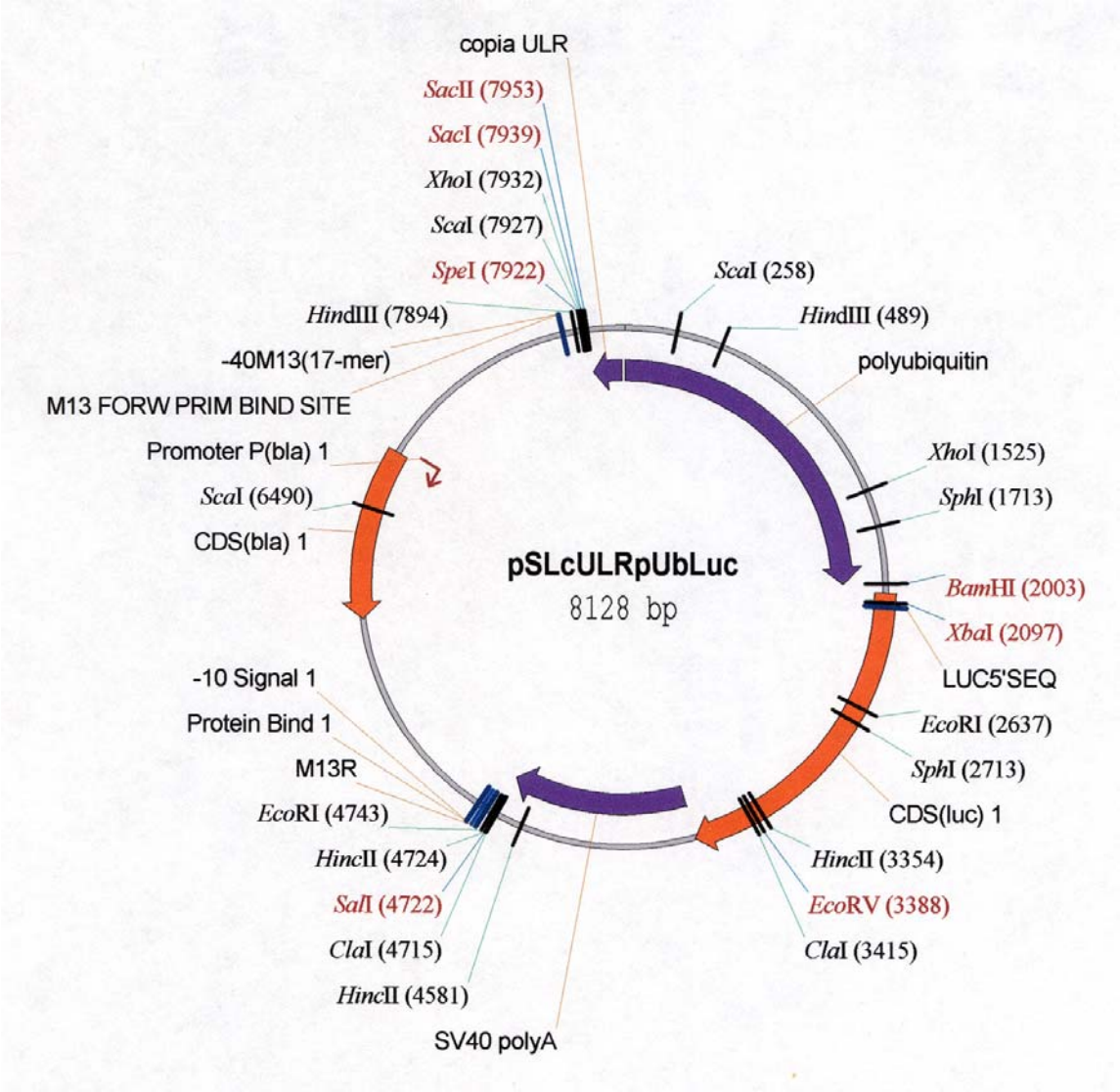
A-7


copia ULR
SacII (1)
HindIII (8381)
ClaI (20)
-40M13(17-mer)
BglII (125)
M13 FORW PRIM BIND SITE
SmaI (420)
Promoter P(bla) 1
HincII (625)
ScaI (6977)
SphI (1098)
CDS(bla) 1
HincII (1177)
Act5C Promoter
pSLcULRAct5CLuc
8558 bp
ClaI (1715)
SpeI (1965)
-10 Signal 1
Misc Feature 2
Protein Bind 1
BamHI (2490)
M13R
XbaI (2584)
EcoRI (5230)
LUC5'SEQ
HincII (5211)
EcoRI (3124)
SalI (5209)
SphI (3200)
ClaI (5202)
CDS(luc) 1
HincII (5068)
HincII (3841)
EcoRV (3875)
ClaI (3902)

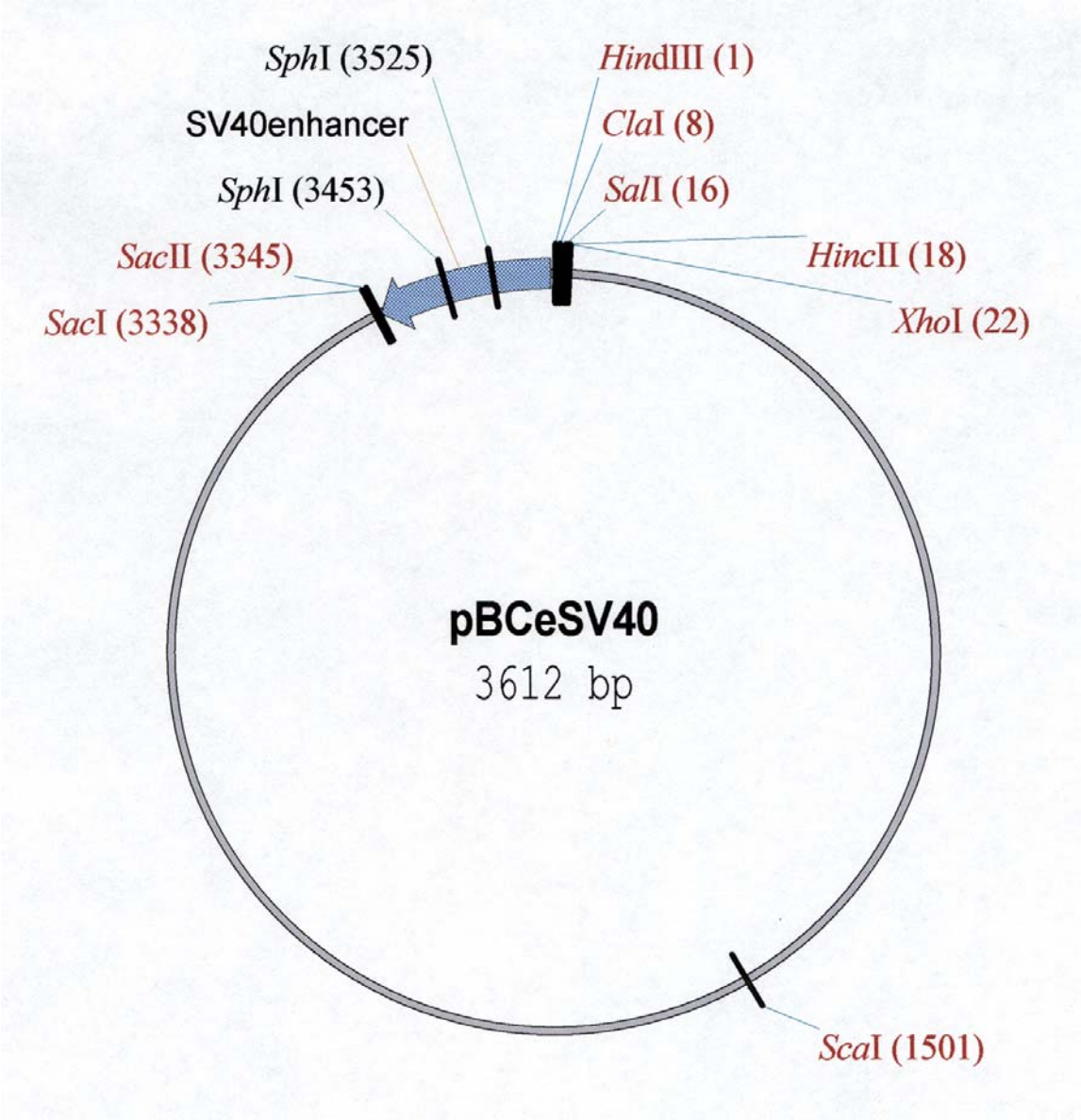
A-8


SacII (1)
ClaI (20)
EcoRI (48)
SphI (190)
IE1
HincII (620)
XbaI (663)
BglII (666)
SpeI (690)
SmaI (703)
BamHI (706)
XbaI (800)
LUC5'SEQ
EcoRI (1340)
SphI (1416)
CDS(luc) 1
HincII (2057)
EcoRV (2091)
ClaI (2118)
SV-40 polyA
HincII (3284)
SalI (3425)
ClaI (3418)
EcoRI (3446)
HincII (3427)
M13R
Protein Bind 1
-10 Signal 1
ScaI (5193)
CDS(bla) 1
Promoter P(bla) 1
M13 FORW PRIM BIND SITE
-40M13(17-mer)
HindIII (6597)
copia ULR
pSLcULRIE1Luc
6774 bp

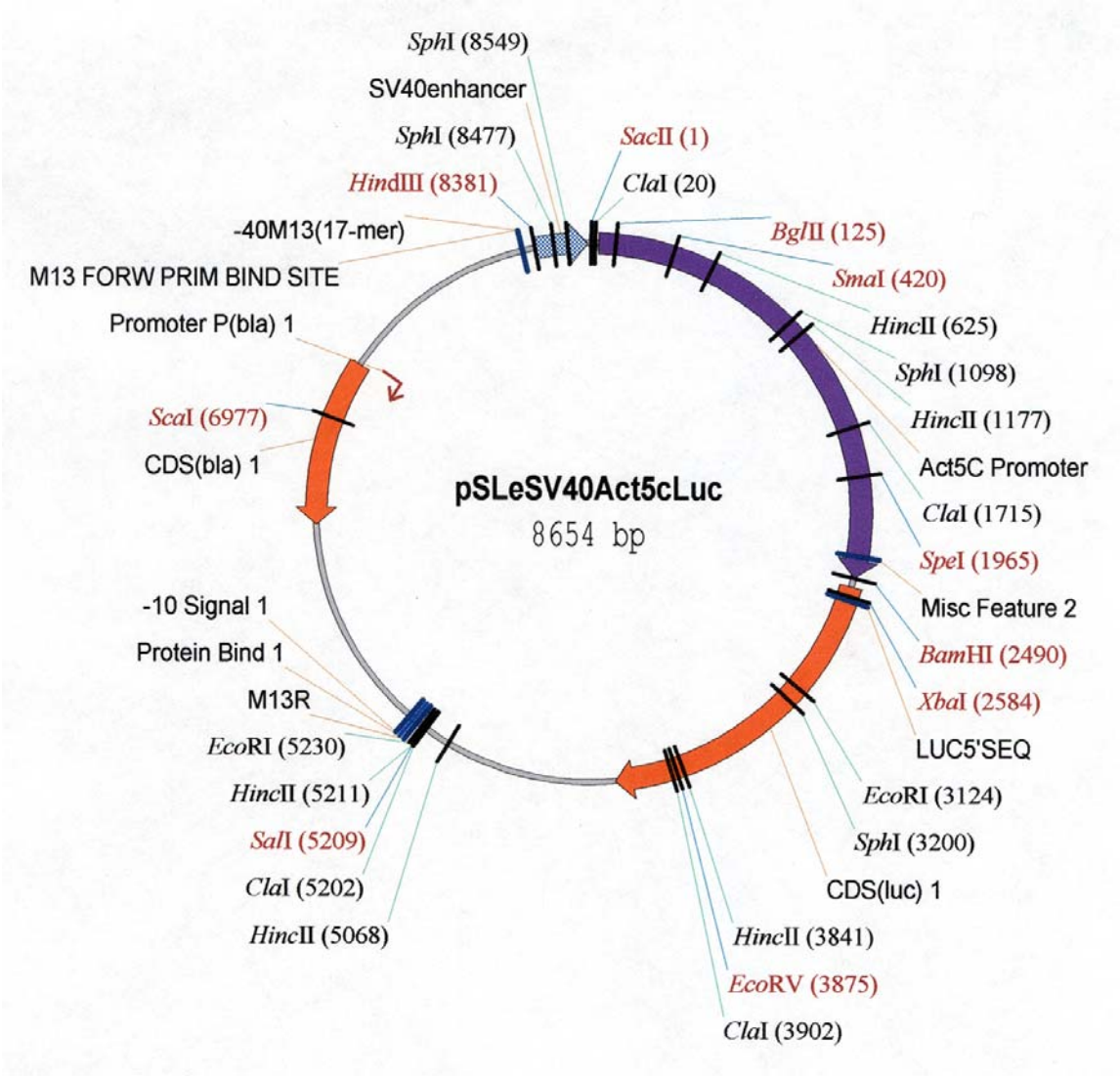
A-9


copia ULR
SacII (7953)
SacI (7939)
XhoI (7932)
ScaI (7927)
SpeI (7922)
HindIII (7894)
-40M13(17-mer)
M13 FORW PRIM BIND SITE
Promoter P(bla) 1
ScaI (6490)
CDS(bla) 1
-10 Signal 1
Protein Bind 1
M13R
EcoRI (4743)
HincII (4724)
SalI (4722)
ClaI (4715)
HincII (4581)
SV40 polyA
ScaI (258)
HindIII (489)
polyubiquitin
XhoI (1525)
SphI (1713)
BamHI (2003)
XbaI (2097)
LUC5'SEQ
EcoRI (2637)
SphI (2713)
CDS(luc) 1
HincII (3354)
EcoRV (3388)
ClaI (3415)
pSLcULRpUbLuc
8128 bp

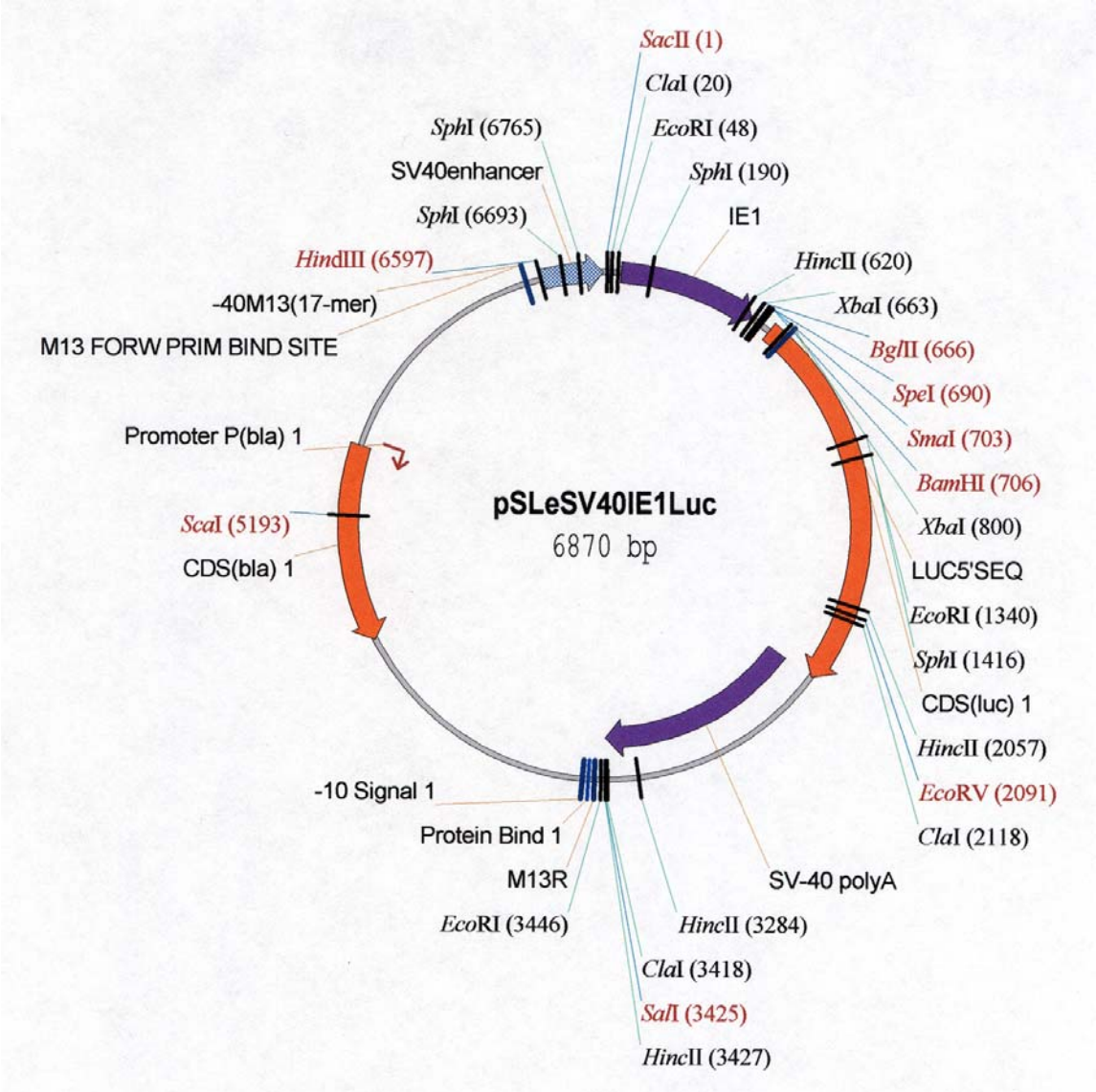
A-10


SphI (3525)
HindIII (1)
SV40enhancer
ClaI (8)
SphI (3453)
SalI (16)
SacII (3345)
HincII (18)
SacI (3338)
XhoI (22)
pBCeSV40
3612 bp
ScaI (1501)

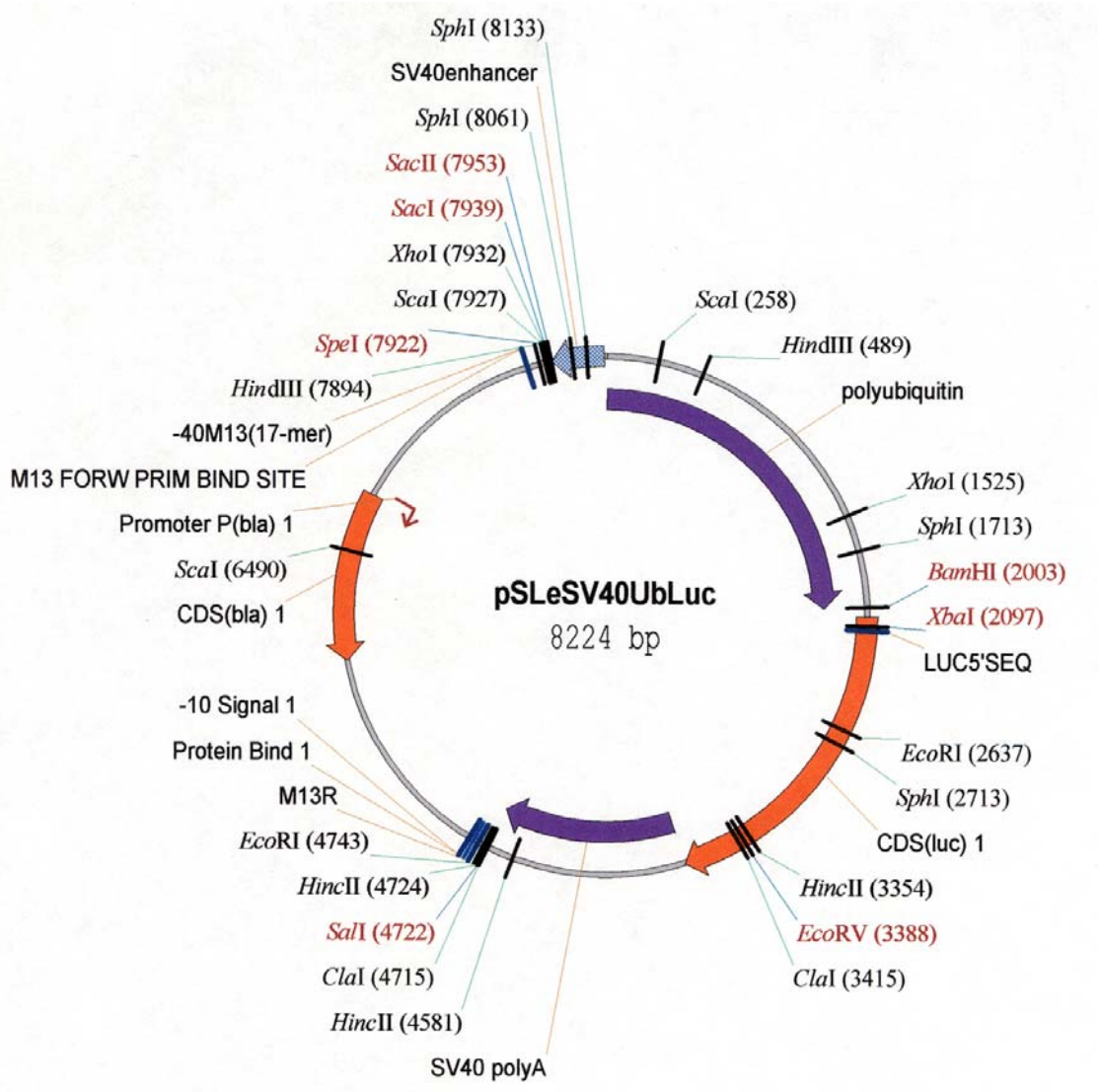
A-11


SphI (8549)
SV40enhancer
SphI (8477)
SacII (1)
HindIII (8381)
ClaI (20)
-40M13(17-mer)
BglII (125)
M13 FORW PRIM BIND SITE
SmaI (420)
Promoter P(bla) 1
HincII (625)
SphI (1098)
ScaI (6977)
HincII (1177)
CDS(bla) 1
Act5C Promoter
pSLeSV40Act5cLuc
8654 bp
ClaI (1715)
SpeI (1965)
Misc Feature 2
-10 Signal 1
BamHI (2490)
Protein Bind 1
XbaI (2584)
M13R
LUC5'SEQ
EcoRI (5230)
EcoRI (3124)
HincII (5211)
SphI (3200)
SalI (5209)
CDS(luc) 1
ClaI (5202)
HincII (3841)
HincII (5068)
EcoRV (3875)
ClaI (3902)

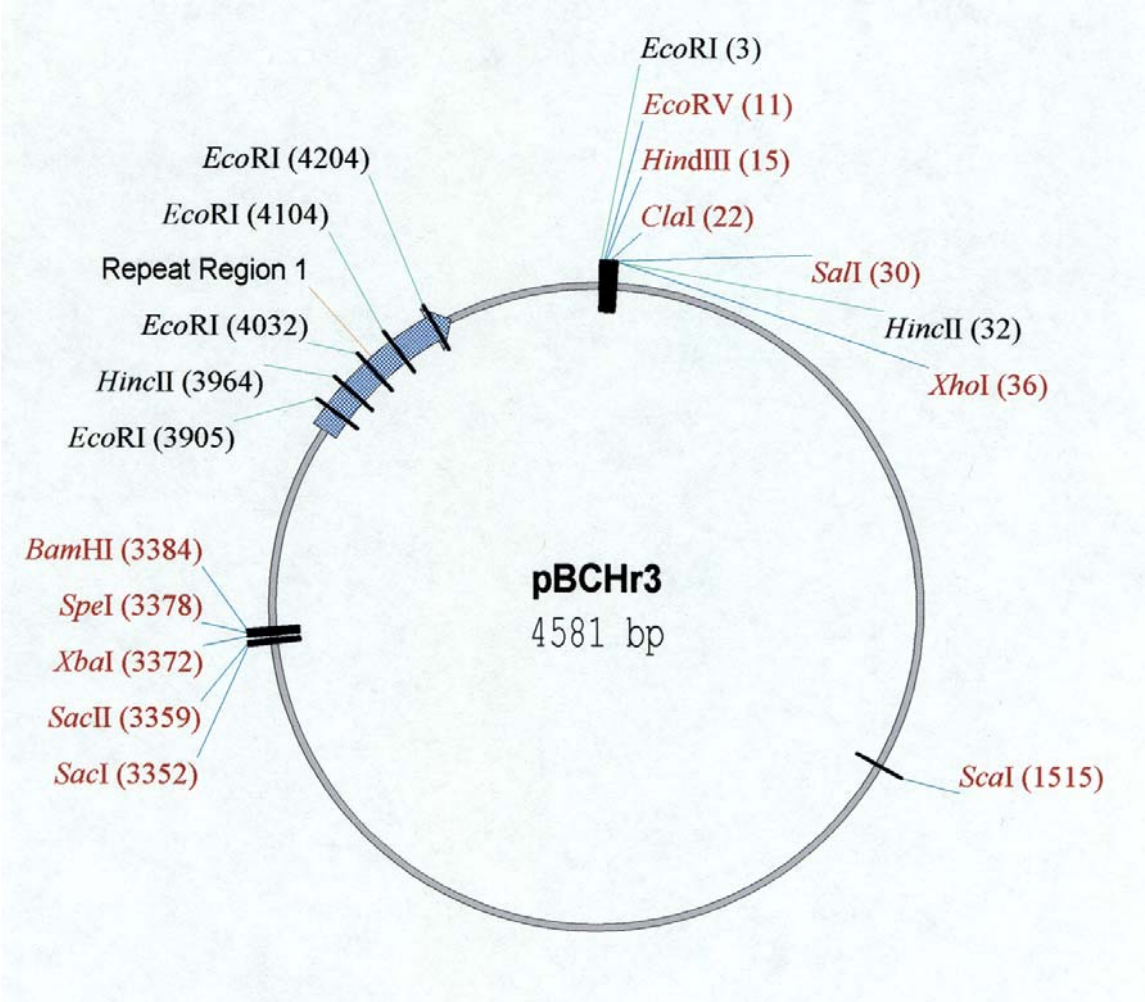
A-12


SacII (1)
ClaI (20)
EcoRI (48)
SphI (190)
IE1
HincII (620)
XbaI (663)
BglII (666)
SpeI (690)
SmaI (703)
BamHI (706)
XbaI (800)
LUC5'SEQ
EcoRI (1340)
SphI (1416)
CDS(luc) 1
HincII (2057)
EcoRV (2091)
ClaI (2118)
SV-40 polyA
HincII (3284)
ClaI (3418)
SalI (3425)
HincII (3427)
EcoRI (3446)
M13R
Protein Bind 1
-10 Signal 1
CDS(bla) 1
ScaI (5193)
Promoter P(bla) 1
M13 FORW PRIM BIND SITE
-40M13(17-mer)
HindIII (6597)
SphI (6693)
SV40enhancer
SphI (6765)
pSLeSV40IE1Luc
6870 bp

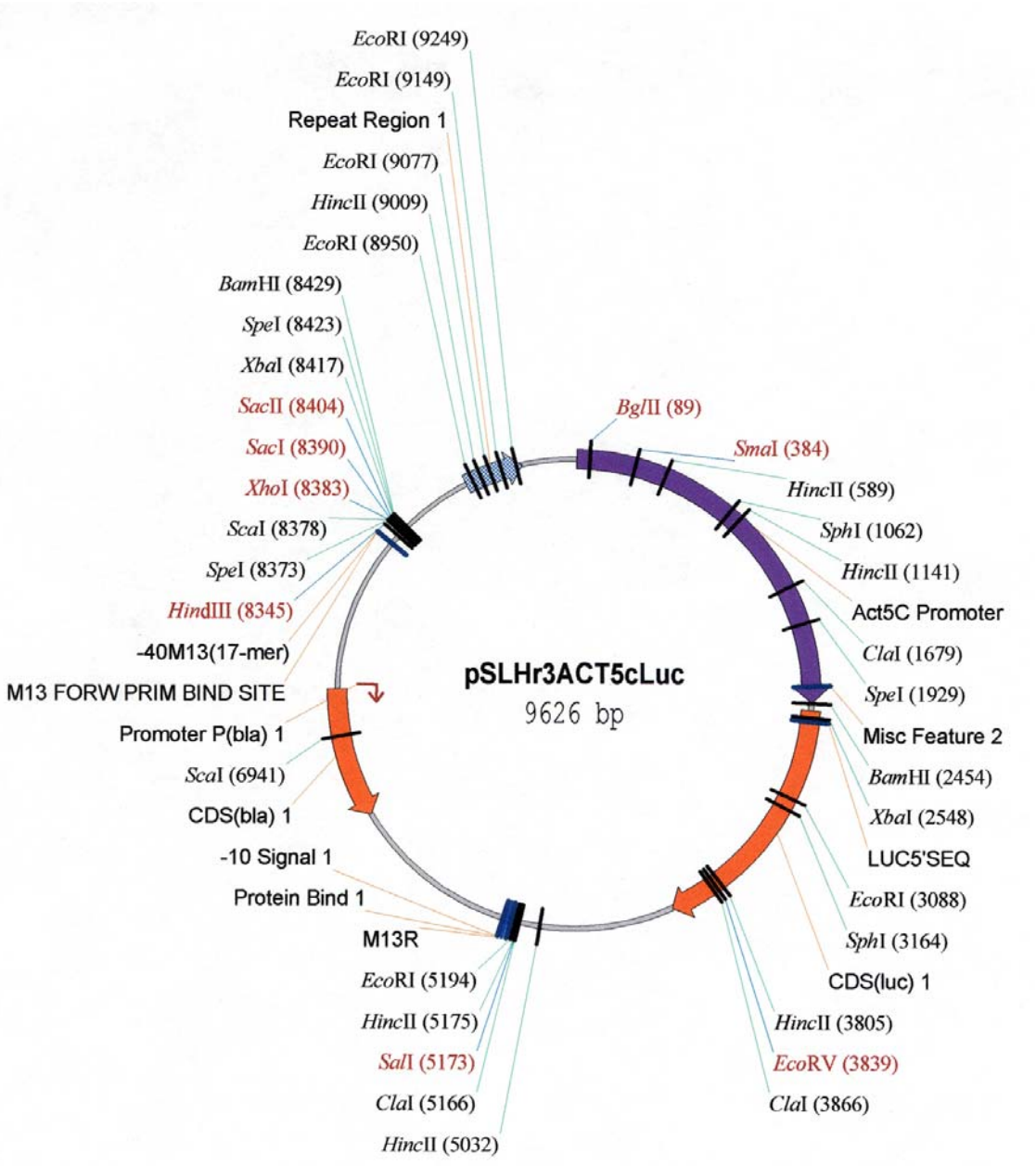
A-13


SphI (8133)
SV40enhancer
SphI (8061)
SacII (7953)
SacI (7939)
XhoI (7932)
ScaI (7927)
SpeI (7922)
HindIII (7894)
-40M13(17-mer)
M13 FORW PRIM BIND SITE
Promoter P(bla) 1
ScaI (6490)
CDS(bla) 1
-10 Signal 1
Protein Bind 1
M13R
EcoRI (4743)
HincII (4724)
SalI (4722)
ClaI (4715)
HincII (4581)
SV40 polyA
ScaI (258)
HindIII (489)
polyubiquitin
XhoI (1525)
SphI (1713)
BamHI (2003)
XbaI (2097)
LUC5'SEQ
EcoRI (2637)
SphI (2713)
CDS(luc) 1
HincII (3354)
EcoRV (3388)
ClaI (3415)
pSLeSV40UbLuc
8224 bp

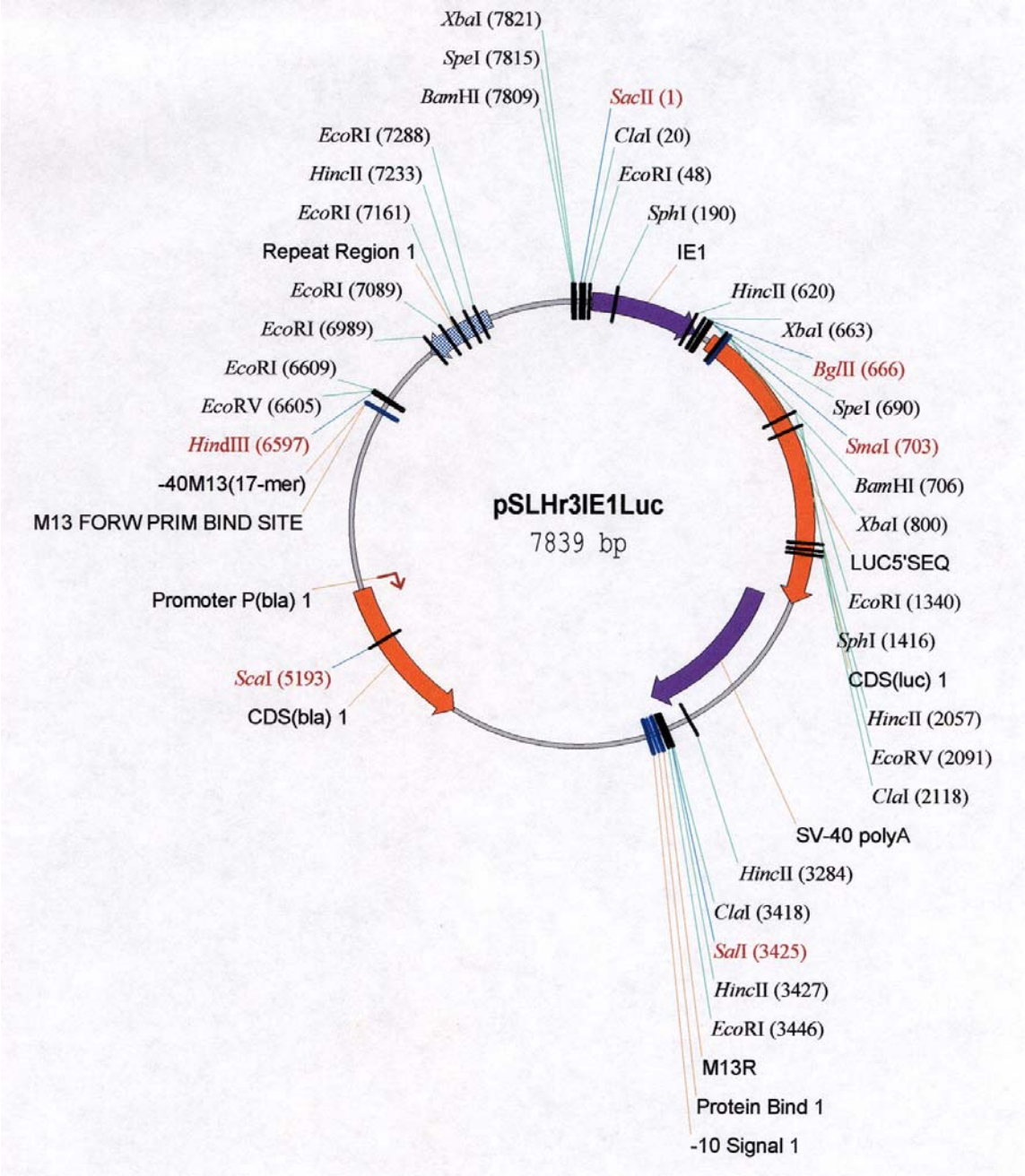
A-14


EcoRI (3)
EcoRV (11)
HindIII (15)
ClaI (22)
SalI (30)
HincII (32)
XhoI (36)
ScaI (1515)
SacI (3352)
SacII (3359)
XbaI (3372)
SpeI (3378)
BamHI (3384)
EcoRI (3905)
HincII (3964)
EcoRI (4032)
Repeat Region 1
EcoRI (4104)
EcoRI (4204)
pBCHr3
4581 bp

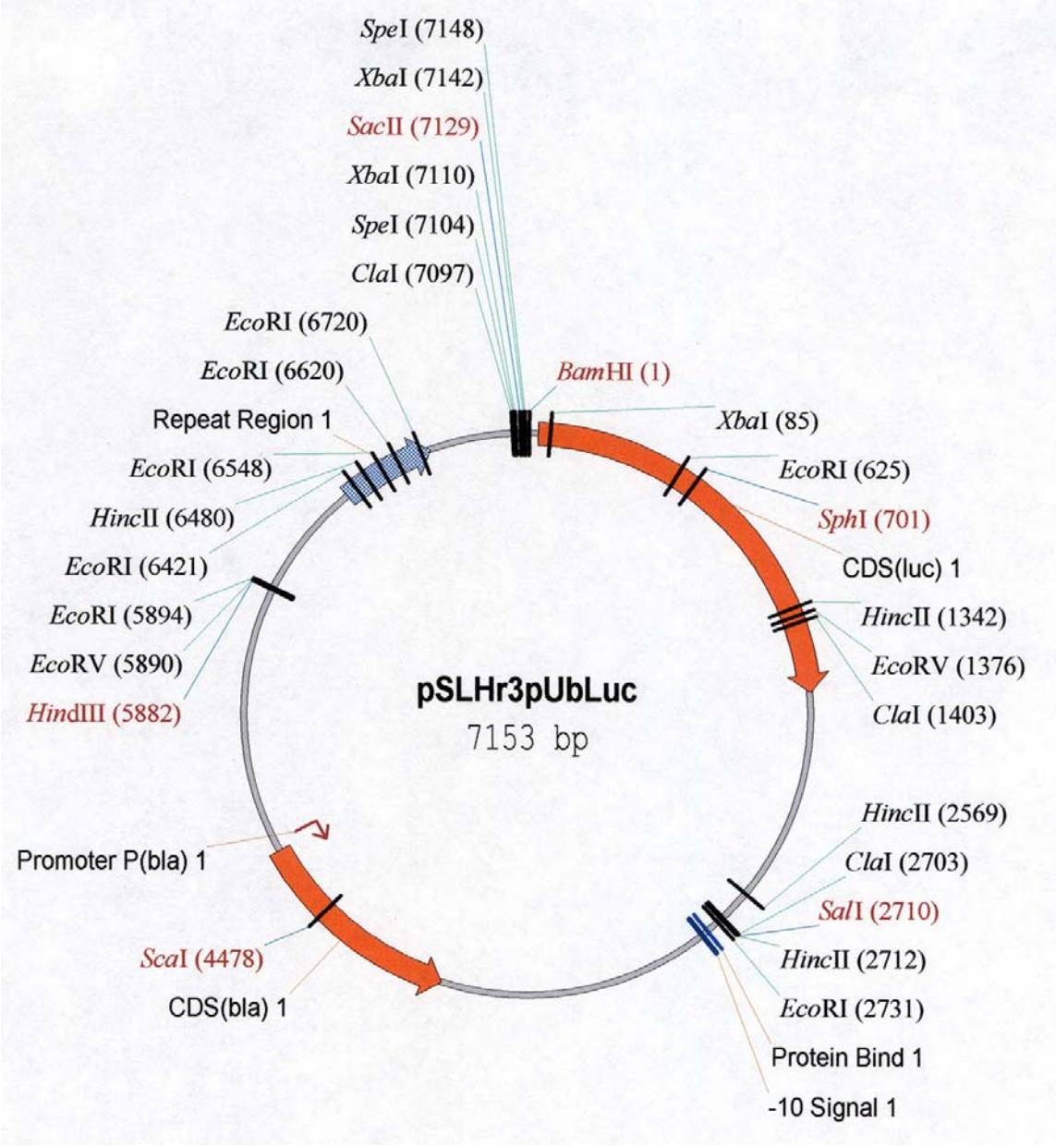
A-15


pSLHr3ACT5cLuc
9626 bp
EcoRI (9249)
EcoRI (9149)
Repeat Region 1
EcoRI (9077)
HincII (9009)
EcoRI (8950)
BamHI (8429)
SpeI (8423)
XbaI (8417)
SacII (8404)
SacI (8390)
XhoI (8383)
ScaI (8378)
SpeI (8373)
HindIII (8345)
-40M13(17-mer)
M13 FORW PRIM BIND SITE
Promoter P(bla) 1
ScaI (6941)
CDS(bla) 1
-10 Signal 1
Protein Bind 1
M13R
EcoRI (5194)
HincII (5175)
SalI (5173)
ClaI (5166)
HincII (5032)
BglII (89)
SmaI (384)
HincII (589)
SphI (1062)
HincII (1141)
Act5C Promoter
ClaI (1679)
SpeI (1929)
Misc Feature 2
BamHI (2454)
XbaI (2548)
LUC5'SEQ
EcoRI (3088)
SphI (3164)
CDS(luc) 1
HincII (3805)
EcoRV (3839)
ClaI (3866)

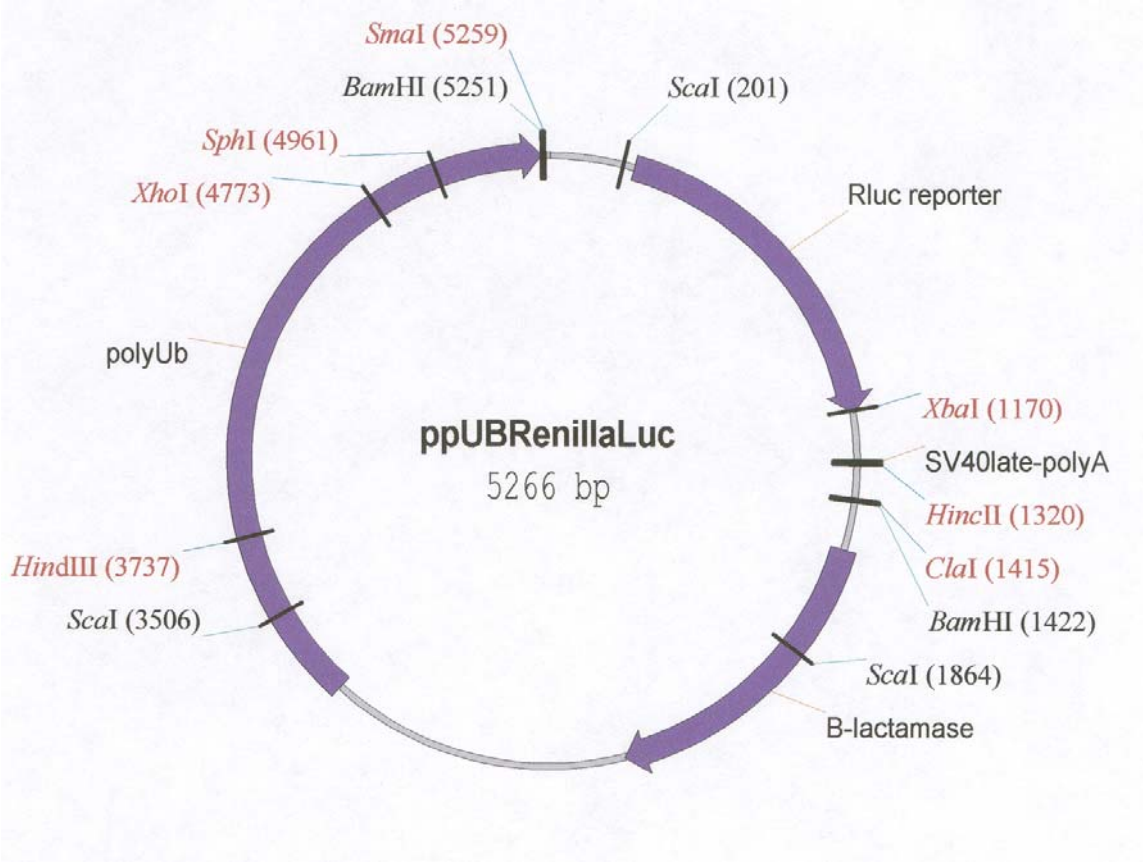
A-16


XbaI (7821)
SpeI (7815)
BamHI (7809)
SacII (1)
EcoRI (7288)
ClaI (20)
HincII (7233)
EcoRI (48)
EcoRI (7161)
SphI (190)
Repeat Region 1
IE1
EcoRI (7089)
HincII (620)
EcoRI (6989)
XbaI (663)
EcoRI (6609)
BglII (666)
EcoRV (6605)
SpeI (690)
HindIII (6597)
SmaI (703)
-40M13(17-mer)
BamHI (706)
M13 FORW PRIM BIND SITE
pSLHr3IE1Luc
7839 bp
XbaI (800)
LUC5'SEQ
Promoter P(bla) 1
EcoRI (1340)
SphI (1416)
ScaI (5193)
CDS(luc) 1
CDS(bla) 1
HincII (2057)
EcoRV (2091)
ClaI (2118)
SV-40 polyA
HincII (3284)
ClaI (3418)
SalI (3425)
HincII (3427)
EcoRI (3446)
M13R
Protein Bind 1
-10 Signal 1

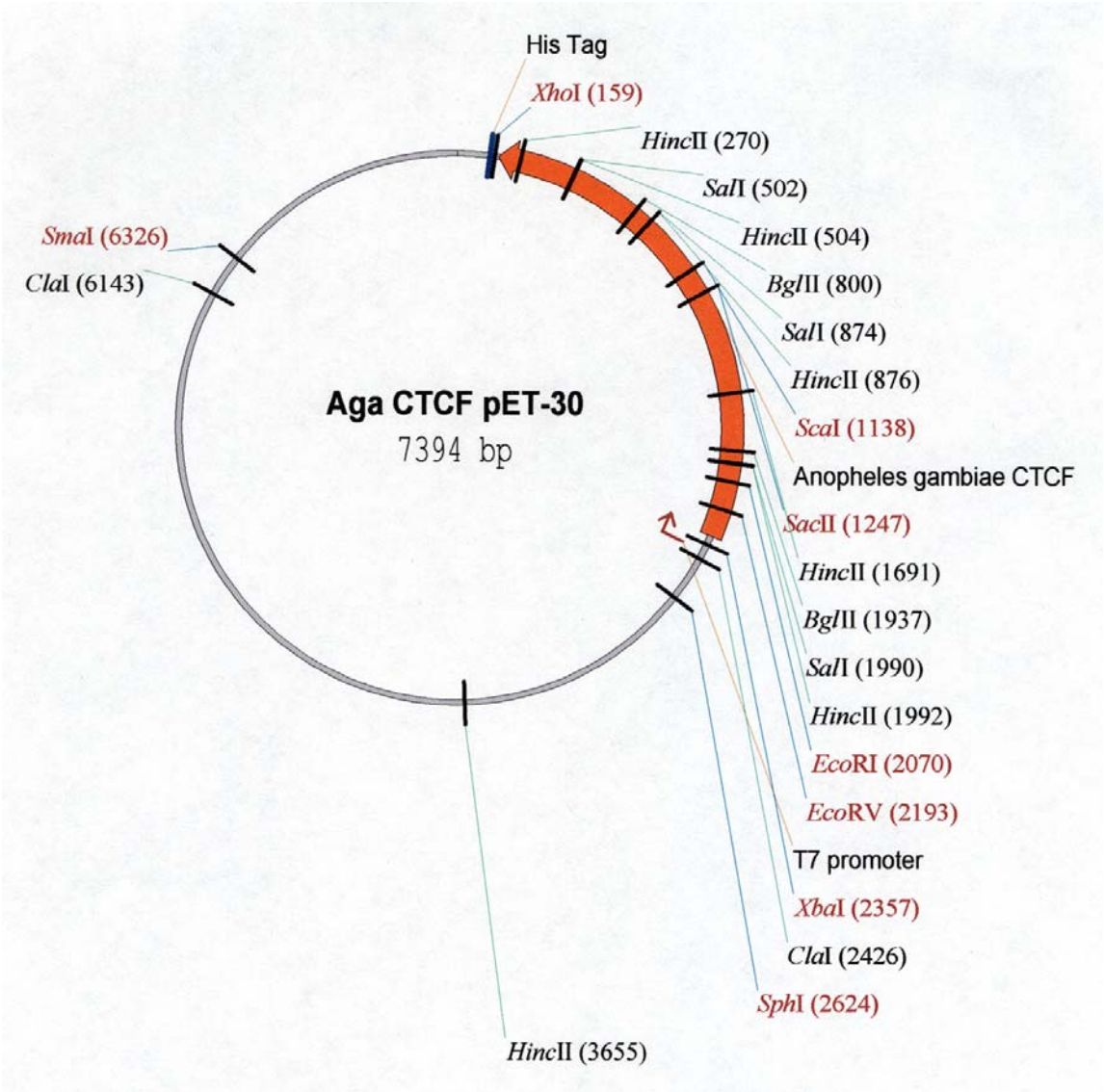
A-17


SpeI (7148)
XbaI (7142)
SacII (7129)
XbaI (7110)
SpeI (7104)
ClaI (7097)
EcoRI (6720)
EcoRI (6620)
Repeat Region 1
EcoRI (6548)
HincII (6480)
EcoRI (6421)
EcoRI (5894)
EcoRV (5890)
HindIII (5882)
BamHI (1)
XbaI (85)
EcoRI (625)
SphI (701)
CDS(luc) 1
HincII (1342)
EcoRV (1376)
ClaI (1403)
pSLHr3pUbLuc
7153 bp
HincII (2569)
ClaI (2703)
SalI (2710)
HincII (2712)
EcoRI (2731)
Protein Bind 1
-10 Signal 1
Promoter P(bla) 1
ScaI (4478)
CDS(bla) 1

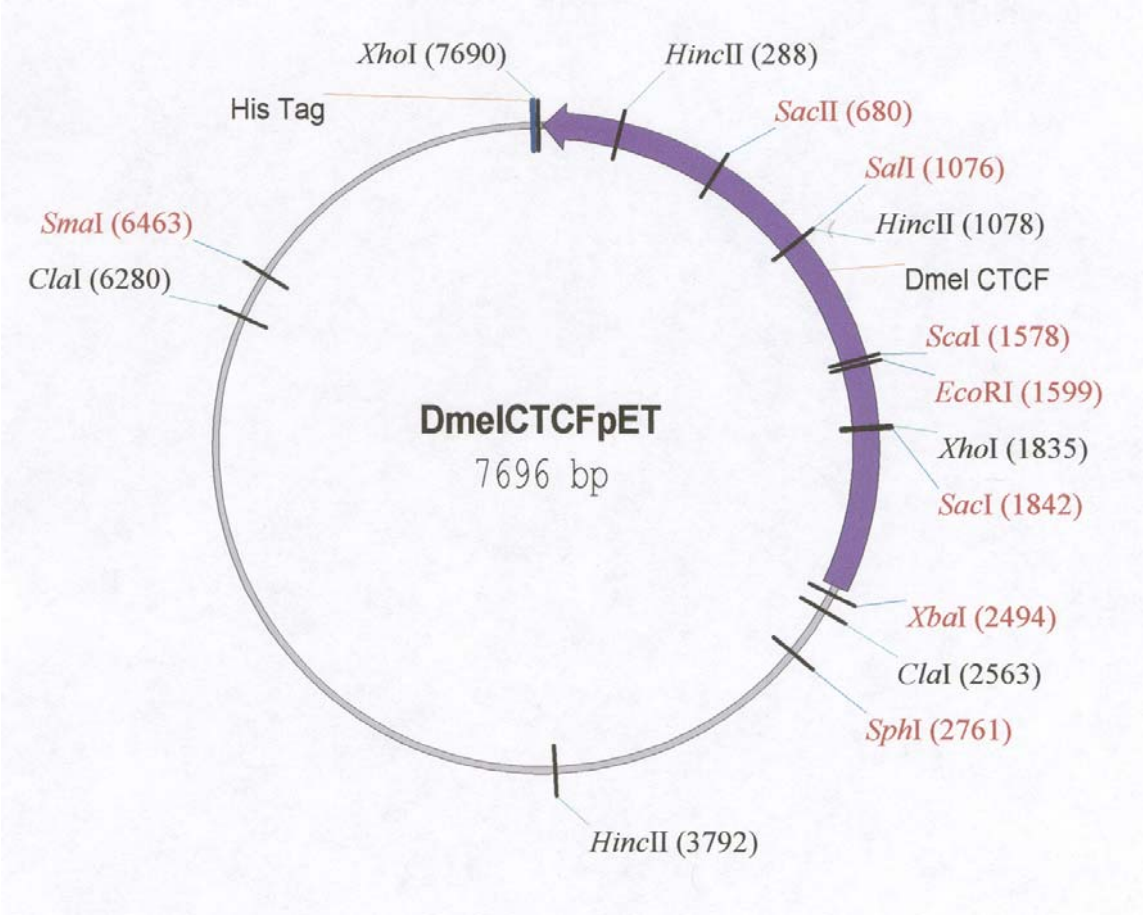
A-18


SmaI (5259)
BamHI (5251)
ScaI (201)
SphI (4961)
XhoI (4773)
Rluc reporter
polyUb
XbaI (1170)
ppUBRenillaLuc
5266 bp
SV40late-polyA
HincII (1320)
ClaI (1415)
HindIII (3737)
ScaI (3506)
BamHI (1422)
ScaI (1864)
B-lactamase

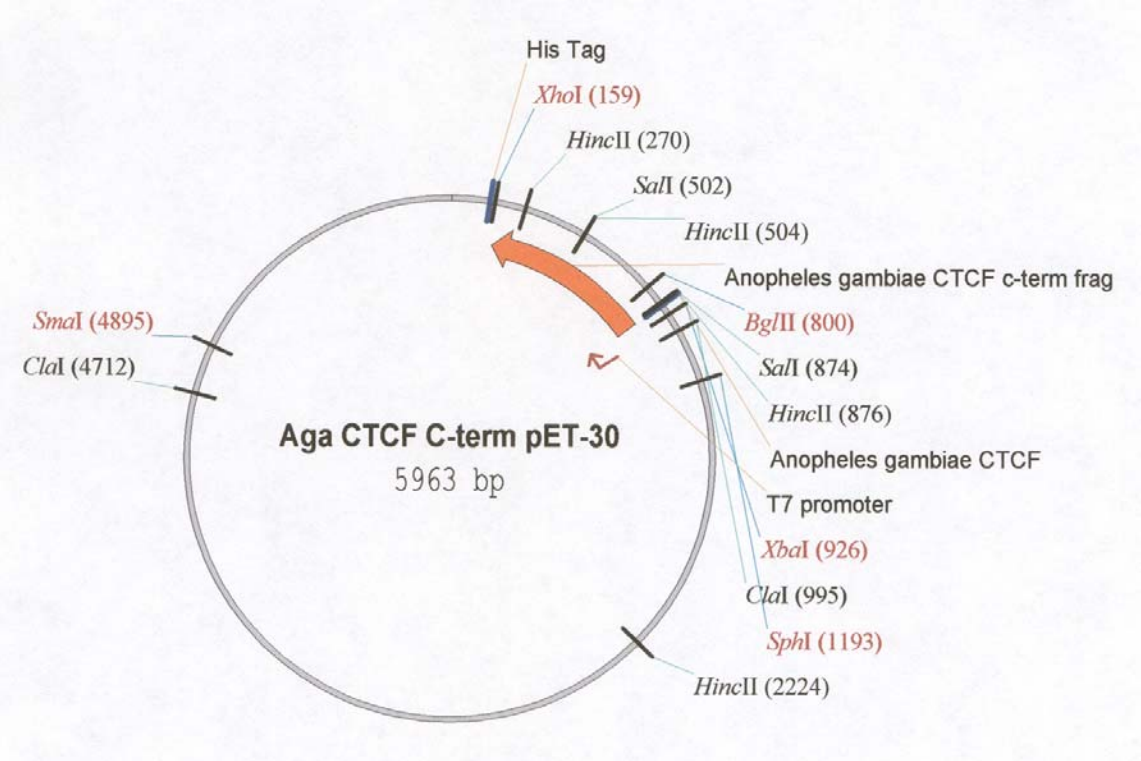
A-19


His Tag
XhoI (159)
HincII (270)
SalI (502)
HincII (504)
BglII (800)
SalI (874)
HincII (876)
ScaI (1138)
Anopheles gambiae CTCF
SacII (1247)
HincII (1691)
BglII (1937)
SalI (1990)
HincII (1992)
EcoRI (2070)
EcoRV (2193)
T7 promoter
XbaI (2357)
ClaI (2426)
SphI (2624)
HincII (3655)
SmaI (6326)
ClaI (6143)
Aga CTCF pET-30
7394 bp

A-20


XhoI (7690)
HincII (288)
His Tag
SacII (680)
SalI (1076)
SmaI (6463)
HincII (1078)
ClaI (6280)
DmeI CTCF
ScaI (1578)
EcoRI (1599)
DmelCTCFpET
7696 bp
XhoI (1835)
SacI (1842)
XbaI (2494)
ClaI (2563)
SphI (2761)
HincII (3792)

A-21


His Tag
XhoI (159)
HincII (270)
SalI (502)
HincII (504)
Anopheles gambiae CTCF c-term frag
BglII (800)
SalI (874)
HincII (876)
Anopheles gambiae CTCF
T7 promoter
XbaI (926)
ClaI (995)
SphI (1193)
HincII (2224)
SmaI (4895)
ClaI (4712)
Aga CTCF C-term pET-30
5963 bp

APPENDIX B

PUTATIVE, *IN SILICO* CTCF BINDING SITES

Location	Gene ID	Gene Name/Description	Gene Start	Gene End	CTCF site start	CTCF site end	CTCF BS sequence
X	ENSANGG00000006157.2	STAT2	1609333	1620712	1624289	1624302	CCCCTCCCCACCCC
X	ENSANGG00000017521.2	GPRnpy4	1865162	1866412	1867380	1867393	CCCCTCACCCCCCC
Xr	ENSANGG00000017521.2	GPRnpy4	1865162	1866412	1863193	1863180	CTCCTCCCCCCCCC
X	ENSANGG00000016966.2	Q9BIH5	2405694	2406041	2403313	2403326	CCCCTCACCCCCCC
X	ENSANGG00000010238.2	Q17010	3484517	3493483	3494647	3494660	CCCCTCCCCAACAC
X	ENSANGG00000018273.2	CTL5	7812685	7814127	7818994	7819007	CACCTCCCCAACCC
X	ENSANGG00000018253.2	GPRads	7834240	7835574	7836137	7836150	CACCTCCCCCCCCC
X	ENSANGG00000016765.2	GPRNNA19	11100509	11101428	11096265	11096278	CCCCTCCCCCCCCC
X	ENSANGG00000016765.2	GPRNNA19	11100509	11101428	11099549	11099562	CCCCTCCCCACCCC
X	ENSANGG00000009474.2	CEC2	12437472	12438056	12435464	12435477	CCCCTCCCCATCCC
X	ENSANGG00000001622.2	Q8WQN7/Cytochrome P450 CYP4G17	16619145	16621029	16621381	16621394	CTCCTCACCCACCA
X	ENSANGG00000010645.2	Q8MUQ2/Glutathione S-Transferase	16746632	16747761	16750958	16750971	CTCCTCCCCACCCC
Xr	ENSANGG00000014941.1	NFI	1050327	1053595	1054532	1054519	CTCCTCCCCCGCCC
Xr	ENSANGG00000014941.1	NFI	1050327	1053595	1054743	1054730	CCCCTCCCCCCGCC
Xr	ENSANGG00000005522.2	White gene	9874701	9882383	9872157	9872170	CCCCTCCCCCCCCC
Xr	ENSANGG00000007830.2	CASPS7	15364447	15364469	15364460	15364447	CCCCTCCCCACCCA
Xr	ENSANGG00000004098.2	CTLSE1	17503854	17514772	17501810	17501797	CACCTCCGCCACCA
Xr	ENSANGG00000015554.2	TOLL1A	190057424	19061018	19061600	19061587	CGCCTCCCCCCGCC
2L	ENSANGG00000020802.2	RL7A_ANOGA/60S Ribosomal Protein L7A	23486056	23486871	23491349	23491362	CCCCTCCGCCACCC
2L	ENSANGG00000022240.1	PGRPS2	29404371	29405108	29407898	29407911	CTCCTCCCCCCCCC
2L	ENSANGG00000018546.2	76810	31220036	31225567	31229078	31229091	CCCCTCACCCACAC
2L	ENSANGG00000011567.2	SUI1_ANOGA	31801749	31806393	31811163	31811176	CTCCTCCCCCCGAA
2L	ENSANGG00000019041.2	AAP78790/TMCA-like prot frag	47897833	47901140	47904556	47904569	CTCCTCCCCCCCCC
2L	ENSANGG00000019060.2	Ianb2/LAMININ GAMMA 1 PRECURSOR	47906096	47914547	47904557	47904570	CTCCTCCCCCCCCC
2L	ENSANGG00000019032.2	SRPN14	48515573	48517224	48514370	48514383	CGCCTCCCCAGGCC
2Lr	ENSANGG00000008181.2	GALE6	3261147	3262162	3258664	3258677	CGCCTCCCCCTCCC
2Lr	ENSANGG00000002705.2	GALE7	3262708	3263040	3258664	3258677	CGCCTCCCCCTCCC
2Lr	ENSANGG00000008306.2	GPRor31	6920607	6921966	6916411	6916424	CGCCTCCGCCGGAA
2Lr	ENSANGG00000017989.2	Q8I952 SMC3 PROTEIN	30474811	30478413	30478420	30478433	CACCTCCCCCGCCC
2Lr	ENSANGG00000017771.1	GNBPA1	37354962	37357227	37361331	37361344	CTCCTCCGCCGGAA
2Lr	ENSANGG00000019161.2	?	39322889	39324625	39325749	39325762	CTCCTCCCCACCCC
2Lr	ENSANGG00000005827.2	SRPN3	39325990	39327093	39325750	39325763	CTCCTCCCCACCCC
2Lr	ENSANGG00000019323.2	SRPN2	39327636	39330513	39325750	39325763	CTCCTCCCCACCCC
2Lr	ENSANGG00000019032.2	SRPN14	48515573	48517224	48519148	48519161	CCCCTCCCCCACAC
2Lr	ENSANGG00000018236.2	SRPN7	48521426	48522681	48519149	48519162	CCCCTCCCCCACAC

Location	Gene ID	Gene Name/Description	Gene Start	Gene End	CTCF site start	CTCF site end	CTCF BS sequence
2R	ENSANGG00000011456.2	O45048 Serine Protease	1275476	1276466	1271134	1275489	CCCCTCCGCCACCC
2R	ENSANGG00000008988.2	GPRopr	5668629	5670745	5667453	5667466	CCCCTCCGCAACAC
2R	ENSANGG00000009167.2	GPRmtn1	5680842	5682419	5683377	5683390	CACCTCCGCATCCA
2R	ENSANGG00000009190.2	GPRgnr3	5890398	5895766	5886432	5886445	CCCCTCCCCACCCC
2R	ENSANGG00000006494.2	Q8T5I0 F25C8.3 PROTEIN	6308197	6324138	6327714	6327727	CTCCTCCGCCTCCC
2R	ENSANGG00000014878.2	Q86LB8 MULTISUBSTRATE DEOXYRIBONUCLEOSIDE KINASE	6479830	6482286	6485171	6485184	CCCCTCCCCAGGAA
2R	ENSANGG00000015653.2	Q8T5K3 HYPOTHETICAL 34.1 KDA PROTEIN	6483108	6484745	6485171	6485184	CCCCTCCCCAGGAA
2R	ENSANGG00000022236.1	Q8T5K2 PUTATIVE V-ATPASE	6485798	6488932	6485172	6485184	CCCCTCCCCAGGAA
2R	ENSANGG00000021197.2	Q8T5J1 HYPOTHETICAL 29.2 KDA PROTEIN	6581887	6585340	6579845	6579858	CGCCTCCCCACCCC
2R	ENSANGG00000024946.1	Q8T5H5 PUTATIVE APYRASE/NUCLEOTIDASE	6634530	6636230	6639306	6639319	CTCCTCCCCAAGCA
2R	ENSANGG00000014879.2	Q8T5H4 PUTATIVE SODIUM CHANNEL	6641935	6644418	6639306	6639319	CTCCTCCCCAAGCA
2R	ENSANGG00000014879.2	Q8T5H4 PUTATIVE SODIUM CHANNEL	6641935	6644418	6645641	6645654	CCCCTCCCCCCCCC
2R	ENSANGG00000023897.1	Q86PT6 PUTATIVE ANTENNAL CARRIER PROTEIN A5	14551198	14552501	14550899	14550912	CTCCTCCCCCTCCC
2R	ENSANGG00000023897.1	Q86PT6 PUTATIVE ANTENNAL CARRIER PROTEIN A5	14551198	14552501	14550905	14550918	CCCCTCCCCCCCAC
2R	ENSANGG00000020751.2	PBD1 No description	20508177	20510100	20504339	20504352	CCCCTCCGCCAGCC
2Rr	ENSANGG00000020751.2	PBD1 No description	20508177	20510100	20513044	20513057	CCCCTCCCCCCCCC
2R	ENSANGG00000019506.2	GPRop8	21814019	21815244	21813455	21813468	CACCTCCCCACGCA
2Rr	ENSANGG00000019506.2	GPRop8	21814019	21815244	21818384	21818397	CTCCTCCGCACCAC
2R	ENSANGG00000019663.2	VATF_ANOGA VACUOLAR ATP SYNTHASE SUBUNIT F	21850799	21851844	21853447	21853460	CTCCTCCCCCGGCA
2R	ENSANGG00000018528.2	GPRor39	24850237	24851844	24849444	24849457	CCCCTCCCCCCCCC
2R	ENSANGG00000002327.2	IAP6	25143234	25161646	25139780	25139793	CCCCTCCCCACCCC
2R	ENSANGG00000014466.2	PPO1 PROPHENOLOXIDASE	28007271	28014859	28015680	28015693	CCCCTCCCCCTCCA
2R	ENSANGG00000013691.2	CYP6AA1 CYTOCHROME P450	28280942	28482563	28480090	28480103	CTCCTCCGCAACAA
2R	ENSANGG00000013691.2	CYP6AA1 CYTOCHROME P450	28280942	28482563	28483859	28483872	CACCTCCGCCCGCA
2R	ENSANGG00000016123.2	GPRmgl4	28945454	28948945	28953662	28953675	CACCTCCCCACCCC
2R	ENSANGG00000017835.2	Q9Y1K7 SERINE PROTEASE 14A	31605800	31609313	31603454	31603467	CACCTCCCCACCAC

Location	Gene ID	Gene Name/Description	Gene Start	Gene End	CTCF site start	CTCF site end	CTCF BS sequence
2R	ENSANGG00000012025.2	OBP17 ODORANT-BINDING PROTEIN G.15B.B	36735531	36736092	36734641	36734654	CCCCTCACCAGCCA
2R	ENSANGG00000011782.2	Q8WR27 HYPOTHETICAL 6.2 KDA PROTEIN	39819450	39820087	39814655	39814668	CCCCTCCCCATCCA
2Rr	ENSANGG00000009071.2	Q8T5I9 PUTATIVE NA-K-CL SYMPORTER	6186532	6193122	6194616	6194629	CCCCTCACCCCCAA
2Rr	ENSANGG00000015655.2	Q8T5I7 PUTATIVE TPR-CONTAINING PHOSPHOPROTEIN	6237596	6241453	6236424	6236437	CTCCTCCCCCACCA
2Rr	ENSANGG00000015671.2	Q8T5H3 HYPOTHETICAL PROTEIN	6644506	6645721	6650403	6650416	CCCCTCCCCCACCC
2Rr	ENSANGG00000019659.2	CLIPB17	7278265	7280367	7273874	7273887	CACCTCCCCAGCAA
2Rr	ENSANGG00000005982.2	GPROR8	12012990	12014364	12009441	12009454	CCCCTCCCCCCCCC
2Rr	ENSANGG00000015920.2	GPRvpr2	12906983	12909642	12904238	12904251	CACCTCCCCCACCC
2Rr	ENSANGG00000005919.2	AAP47144 RH-LIKE GLYCOPROTEIN	13892586	13895365	13899215	13899228	CCCCTCCCCACCCC
2Rr	ENSANGG00000010972.2	GPR5HT2a	17828578	17861840	17845507	17845520	CTCCTCCCCCACCC
2Rr	ENSANGG00000019167.2	CLIPB7	18407723	18410962	18403600	18403613	CTCCTCCCCCCCCC
2Rr	ENSANGG00000020872.2	CC42_ANOGA CDC42 HOMOLOG	21425187	21429628	21420412	21420425	CCCCTCCCCCCCCC
2Rr	ENSANGG00000008375.2	O77457 TRYPTOPHAN OXYGENASE	26234851	26238589	26231739	26231752	CCCCTCCCCCCCCC
2Rr	ENSANGG00000013732.2	GNBPB4	27635819	27637107	27637430	27637443	CTCCTCCCCCCCCC
2Rr	ENSANGG00000016173.2	GPRNNA3	28844791	28860913	28842819	28842832	CACCTCCCCCCGCC
2Rr	ENSANGG00000025143.1	CLIPB3	34682300	34685780	34689468	34689481	CCCCTCCCCCCCCC
2Rr	ENSANGG00000011053.2	CLIPB4; SERINE PROTEASE 14D	34686920	34688767	34689468	34689481	CCCCTCCCCCCCCC
2Rr	ENSANGG00000011095.2	CLIPB1; SERINE PROTEASE 14D2	34690716	34692394	34689468	34689481	CCCCTCCCCCCCCC
2Rr	ENSANGG00000022725.1	CLIPB6	34694291	34695024	34694291	34695024	CCCCTCCCCCCCCC
2Rr	ENSANGG00000000822.2	FABP_ANOGA	37222206	37227086	37218008	37218021	CTCCTCACCCGCAA
2Rr	ENSANGG00000020422.2	GPRor42	55538917	55540341	55542891	55542904	CTCCTCCCCCGGAC
2Rr	ENSANGG00000020422.2	GPRor42	55538917	55540341	55542941	55542964	CCCCTCCCCACCCA
2Rr	ENSANGG00000021070.2	GPRor26	56797729	56799189	56804096	56804109	CCCCTCCCCCTCCA
2Rr	ENSANGG00000010687.2	GPRor27	56801631	56803102	56804096	56804109	CCCCTCCCCCTCCA
2Rr	ENSANGG00000021023.2	GPRor56	56807102	56808927	56804096	56804109	CCCCTCCCCCTCCA
3L	ENSANGG00000015625.2	CASPS9	10454293	10455302	10459950	10459963	CTCCTCCCCACCAC
3L	ENSANGG00000022587.1	CASPS9	10463619	10464634	10459950	10459963	CTCCTCCCCACCAC
3L	ENSANGG00000022587.1	CASPS9	10463619	10464634	10469180	10469193	CTCCTCCCCACCAC
3L	ENSANGG00000024683.1	GPRNND1	15598369	15599280	15596927	15596940	CCCCTCCCCACCAA
3L	ENSANGG00000003778.2	GPR5HT1b	24312828	24335160	24310154	24310167	CCCCTCCCCCCCCC

Location	Gene ID	Gene Name/Description	Gene Start	Gene End	CTCF site start	CTCF site end	CTCF BS sequence
3L	ENSANGG00000020781.2	GPRor54	33022150	33023656	33026860	33026873	CTCCTCACCACCCA
3L	ENSANGG00000021236.2	GPRor64	35255007	35256433	3256980	3256993	CACCTCAGCCGCCA
3L	ENSANGG00000010870.2	GPRnpr4	37561676	37562691	37558148	37558161	CCCCTCCCCACCCA
3Lr	ENSANGG00000006278.2	TOLL10	17692277	17695780	17699166	17699179	CGCCTCCCCCCGAA
3Lr	ENSANGG00000008873.2	GPRfz1b	22120465	22121010	22116729	22116742	CGCCTCACCAGGAA
3Lr	ENSANGG00000020666.2	Q8WRX3 ODORANT BINDING PROTEIN 1	30074263	30075132	30078626	30078639	CCCCTCCCCCCCCC
3Lr	ENSANGG00000012893.2	Q9TW03 PUTATIVE APYRASE PRECURSOR	35018973	35021033	35014051	35014064	CCCCTCCGCCCCCC
3Lr	ENSANGG00000012715.2	SCRC1	35035447	35038293	35035048	35035061	CCCCTCACCAGCCC
3Lr	ENSANGG00000013240.2	AAP78792 TMCC-LIKE PROTEIN	35762268	35764652	35758384	35758397	CCCCTCCCCAGGCC
3R	ENSANGG00000007441.2	Q8I9N3 XANTHINE DEHYDROGENASE (FRAGMENT)	2793002	2805222	274311	274324	CGCCTCCGCCGGAC
3R	ENSANGG00000019362.2	CLIPD2	6442215	6443979	6447988	6448001	CTCCTCCGCCAGCC
3R	ENSANGG00000015841.2	Q9UB32 D7-RELATED 3 PROTEIN PRECURSOR	8561538	8562255	8566971	8566984	CGCCTCCGCATGAC
3R	ENSANGG00000015851.2	Q9UB30 D7-RELATED 1 PROTEIN PRECURSOR	8562777	8563560	8566971	8566984	CGCCTCCGCATGAC
3R	ENSANGG00000016684.2	IAP5	10567005	10567740	10568590	10568603	CGCCTCCGCAGCAC
3R	ENSANGG00000010559.2	GPRsmo	14073205	14076818	14079479	14079492	CCCCTCCCCCTCCC
3R	ENSANGG00000010766.2	Q9XZN6 PUTATIVE INFECTION RESPONSIVE SHORT PEPTIDE	14236771	14237628	14240366	14240379	CTCCTCCCCCAGCA
3R	ENSANGG00000014070.2	O61469 TU37B2	25789693	25790986	25794881	25794894	CGCCTCACCCCCAA
3R	ENSANGG00000014039.2	PSD3_ANOGA PROBABLE 26S PROTEASOME NON-ATPASE REGULATORY SUBUNIT 3 (26S PROTEASOME SUBUNIT S3) (DIPHENOL OXIDASE A2 COMPONENT) (DOX-A2)	25791318	25792808	25794881	25794894	CGCCTCACCCCCAA
3R	ENSANGG00000019014.2	SCRAC1	27750928	27772019	27750005	27750018	CACCTCACCAGCAC
3R	ENSANGG00000016394.2	GPRor21	32811715	32813100	32818022	32818035	CCCCTCCCCCTCCC
3R	ENSANGG00000009816.1	SCRBQ1	49974623	49976167	49973952	49973965	CACCTCCCCAACAC
3Rr	ENSANGG00000020665.2	Q86QN5 CYTOCHROME P450 CYP12F4	4318878	4320925	4317145	4317158	CCCCTCCCCCCCCC
3Rr	ENSANGG00000021908.1	Q86QN4 CYTOCHROME P450 CYP12F3	4321531	4323866	4317145	4317158	CCCCTCCCCCCCCC
3Rr	ENSANGG00000020684.2	OBP21 ODORANT-BINDING PROTEIN G.29A.A	10317255	10317835	10315426	10315439	CGCCTCACCATGCC

Location	Gene ID	Gene Name/Description	Gene Start	Gene End	CTCF site start	CTCF site end	CTCF BS sequence
3Rr	ENSANGG00000016684.2	IAP5	10567005	10567740	10565975	10565988	CGCCTCAGCAGGAA
3Rr	ENSANGG00000014695.2	CLIPD7	23268448	23269709	23272208	23272221	CACCTCACCAGCAC
3Rr	ENSANGG00000003995.2	RL5_ANOGA 60S RIBOSOMAL PROTEIN L5	24695660	24698102	24690985	24690998	CTCCTCCCCAACCC
3Rr	ENSANGG00000008612.2	REL1 IMMUNE FACTOR	35618914	35640927	35643379	35643392	CCCCTCCCCCTCCC
3Rr	ENSANGG00000012629.2	CASPS8	44732403	44733366	44729161	44729174	CTCCTCCCCAGGCC
3Rr	ENSANGG00000013326.2	CLIPB15 SERINE PROTEASE	44790480	44791914	447995611	44795624	CGCCTCACCAGCCA
UNKN	ENSANGG00000018728.2	Q868R5 GAG-LIKE PROTEIN	43987402	43998734	44002778	44002791	CTCCTCACCACCCC
UNKN	ENSANGG00000011193.2	Q818S7 ODORANT-BINDING PROTEIN G.15D	55771575	55772264	55775526	55775539	CACCTCCCCAACAA

VITA

Christine Elizabeth Gray received her Bachelor of Science degree in biological sciences and secondary education from DePaul University (Chicago) in 1986. She taught high school science for eleven years and studied theology for two years before beginning her graduate studies in genetics at Texas A&M University in September 1999. Over the course of her doctoral studies, Ms. Gray has taught general biology at Blinn College (Bryan, TX). Ms. Gray completed her studies and received her Doctor of Philosophy degree in August 2005.

Ms. Gray intends to continue teaching and to do research in the fields of vector biology and science education. Her research interests include identification and characterization of endogenous insulators in medically-important mosquitoes, development of the flour beetle as a research organism in undergraduate programs, strategies for improving undergraduate research experiences in small school settings and methods in science education at both the secondary and post-secondary levels.

Ms. Gray may be reached at 1908 Hollow Ridge Drive, Cedar Park, TX 78613. Her email address is c-gray@tamu.edu.