University of Alberta

P element dependent gene silencing in *Drosophila melanogaster*

by

Alireza Sameny

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Molecular Biology and Genetics

Department of Biological Sciences
Faculty of Science

Edmonton, Alberta
Fall 2008

Your file Votre référence
ISBN: 978-0-494-46417-5
Our file Notre référence
ISBN: 978-0-494-46417-5

To my father for his kindness and all that he has taught me

To my beloved wife for her patience and support

And to my son for the joy that he has brought to my life

Abstract

P element dependent gene silencing in *Drosophila melanogaster*

This thesis examines how *P* elements silence genes in *Drosophila melanogaster* using the host's heterochromatin proteins. This has been modeled using the transgene *P{lacW}ci*^{DPlac}, a *white*$^{+}$ (*w*$^{+}$) *P* element construct inserted upstream of the *ci* gene on chromosome 4. In the absence of other *P* elements, this transgene produces a uniform red eye phenotype [white^{+} (w^{+})] in *w*$^{-}$ flies, but the presence of some *P* elements such as *P[SalI]* and *KP* changes the phenotype to a variegated eye color due to random silencing of the *w*$^{+}$ transgene expression in different ommatidial cells in the eye. This phenomenon, called *P* element dependent silencing, is dependent on host heterochromatin proteins such as SU(VAR)2-5 and SU(VAR)3-7. These heterochromatin proteins are dose dependent in their effects. As part of this study, I showed that by increasing the number of the *KP* elements in the genome the silencing of the transgene gets stronger. This means that the effects of *P* elements on PDS are also dose dependent.

To examine the role of the P element product in gene silencing, I induced and recovered 22 mutations in *P[SalI]89D* (*P-Sal* hereafter) that affected the process of PDS. These mutations were found in regions previously reported to be required for transposition. However, there were no mutations in the zinc finger domain, which is critical for DNA binding of P element proteins. Moreover, I found two new mutation hot spots in the *P* element DNA sequence. These sequences do not encode any known protein domains.

I also examined the *P-Sal* mutants for their effect on other *P* element constructs that are sensitive to the presence of *P-Sal.* I found that the *P{lacW}ci*^{E1}, *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, and *P{hsp26-pt-T}39C-12* constructs, which are inserted upstream of the *ci*, are similar in their response to my *P-Sal* mutants, while *P{hsp26-pt-T}39C-12*, which is located more distally on chromosome 4, is different both in phenotype and sensitivity.

The phenotypic differences in my *P-Sal* mutations could have been due to differences in polypeptide expression. I tried to induce antibodies against the P-Sal protein. Preliminary characterization of this antibody showed it was not specific enough to identify the protein expression (or lack of expression) in the fly protein extract.

Acknowledgement

My gratitude to the following people cannot be adequately expressed:

Dr John Locke who guided me throughout this project and without his advice I could not accomplish it.

Dr John Bell and Dr Andrew Simmonds who helped me to have a better view of the process throughout the project. They also helped me to overcome technical problems. This project could not be done without their help.

Scot Hanna who helped me a lot to learn about *Drosophila* genetics. I can not thank him enough for all of the interesting discussions that we had with each other.

MBSU staff, especially Gary Ritzel, who were always ready to answer my questions about technical difficulties. Gary's thorough knowledge of molecular biology was always an asset for me.

List of Tables

List of Figures

List of Symbols and Abbreviations

+: wild type allele
aa: amino acids
Ago1: Argonaute 1 gene
Amp^R: *Ampicillin resistance* gene
ap^{Xa}: *Xasta* allele of *apterous* gene
aubergine: *aubergine* gene
BLAST: Basic Local Alignment Search Tool
bp: base pairs
BSA: bovine serum albumin
Bl: Bristle gene
°C: degrees Centigrade
Cam^R: *Chloramphenicol resistance* gene
ci: cubitus interruptus gene
cM: centi-Morgans
Cy: Curly gene
CyO: Curly of Oster chromosome
Δ2-3: $P(ry^+, \Delta 2\text{-}3)99B$
Df(3): chromosome 3 deficiency
DNA: deoxyribonucleic acid
dp: dumpy gene
e: ebony gene
$E1: P\{lacW\}ci^{E1}$
EMS: ethylmethane sulfonate
E(var): Enhancer of variegation gene
E(var)3-9: Enhancer of variegation 3-9 gene
ey: eyeless gene
E(z): enhancer of zeste gene
g: grams
GuHCl: Guanidine Hydrochloride
HP1: heterochromatin protein 1
hPEV: heterochromatic position effect variegation
hsp70: heat shock promoter 70
hsp26 heat shock promoter 26
homeless: *homeless* gene
Hu: Humeral allele of *Antennapedia* gene
Hu: Humeral phenotype
$In(1)w^{m4}$: *Inversion (1) white-mottled*
IPTG: Isopropyl β-D-1-thiogalactopyranoside
kb: kilobases
kD: kiloDaltons
X^2: *Chi square* test
KP(2): chromosome 2 with KP insert
KP(3): chromosome 3 with KP insert
M: molar
mg: milligrams
mini-white: a recombinant *w* gene produced by adding $w^+3.3(Xb, Sp)$ to $w^+0.8(Sp, H3)$
miRNA: microRNA
ml: milliliters
mM: millimolar
MPS: mutated *P-Sal*
mRNA: messenger RNA
µg: micrograms

µl: microliters
NS: Nova Scotia Chromosome
Pak3: Pak3 gene
PBS: phosphate buffered saline
Pci: P{lacW}ci^{DPlac}
~ *Pci*: floating *Pci* in the stock
PCR: polymerase chain reaction
^{32}P-dCTP: 2'-deoxycytidine 5'-triphosphate with radioactive phosphate (^{32}P)
PDS: *P* element dependent silencing
piwi: *piwi* gene
P-Sal: *P[SalI]89D*
*P-Sal**: *P-Sal* mutants
RIGS: repeat induced gene silencing
RNA: ribonucleic acid
RpS3a: Ribosomal protein S3A gene
RT: reverse transcription
RT-PCR: reverse transcription polymerase chain reaction
ry: rosy gene
Sb: Stubble gene
Sb: Stubble phenotype
siRNA: short interfering RNA
sn^{w}: *weak* allele of *singed* gene
Su(var): Suppressor of variegation gene
Su(var)205: Suppressor of variegation 205 gene
Su(var)3-7: Suppressor of variegation 3-7 gene
Su(var)3-9: Suppressor of variegation 3-9 gene
Tb: *Tubby* gene
Tb: Tubby phenotype
THAP: THanatos (death)-Associated Protein
TM6: modified third chromosome 6
TM6B: modified third chromosome 6B
Ubx: Ultrabithorax gene
UTR: untranslated region
vg $^{21-3}$: *21-3* allele of *vestigial* gene
w: white gene
w^{+}: uniform red eye color phenotype
w^{var}: variegated eye color phenotype
y: yellow gene

Chapter 1: Introduction

Transposable elements (transposons) are DNA sequences that have the ability to replicate and insert themselves into the host's genome. Such insertions can disrupt the function of a gene and consequently affect the phenotype of an organism. Therefore, it is to the host's benefit to control transposons. Remnants of old, silent transposons are frequently seen in eukaryotic genomes, and in *Drosophila melanogaster* >50 different transposons make up to 10% of the whole genome (Labrador and Corces, 1997). *P* elements are one of the best characterized transposons in the *Drosophila melanogaster* genome and, because they are newly introduced to the organism, they are still able to actively transpose themselves in certain genetic backgrounds (reviewed by Pinsker *et al*, 2001). Researchers have used this ability to mutagenize genes and to transfer the gene of interest into different locations of a genome. However, *P* elements are also a good model to study how a host, *Drosophila melanogaster*, develops a control mechanism against transposition.

In the current study I use *P* element dependent silencing (PDS) as a model system, to investigate how *P* elements control their own activity. In PDS, the presence of *P* elements in the genome silences expression of reporter genes of certain *P* element constructs. There are two known *P* element constructs that are sensitive to PDS and this sensitivity causes variegated expression of their w^+ marker gene and induces easily seen eye color variegation. Both constructs are inserted upstream of the *ci* gene on chromosome 4 in *Drosophila melanogaster* (Bushey, 2004). I examined PDS and showed an additive effect with its strength being related to the number of *P* elements present in the genome. In another set of experiments I mutagenized a single element, that showed a strong PDS effect, to identify various protein domains that are necessary for PDS induction. I also examined the effects of these mutations on other *P* element related phenotypes.

P element

A complete *P* element is 2.9 kb. This transposon as well as many of its deletion derivatives is present in wild populations of *Drosophila melanogaster* all over the world (reviewed by Pinsker *et al*, 2001). These elements can be classified as autonomous versus non-autonomous, based on their ability to encode an 87 kD transposase, an enzyme that is responsible for *P* element transposition. Autonomous elements are able to transpose themselves while non-autonomous elements need a source of transposase to be transposed (Rio *et al*, 1986). *P* elements

can also be classified based on their ability to repress transposition as repressor versus non-repressor. Repressor elements are divided to two main types. Type I repressor elements have at least 1956 bp of the complete *P* element and are able to encode a 66 kD repressor protein. Type II repressors are shorter (Gloor *et al*, 1993).

History

P elements were discovered as the cause of hybrid dysgenesis, which included high mutation rate, male recombination, and male sterility among progeny of crosses between males from wild populations and females from laboratory stocks that were gathered at the beginning of 20th century. Reciprocal crosses did not show this phenotype (Engels and Peterson, 1980). Flies gathered from wild environments had *P* elements while old laboratory stocks did not have *P* elements. This difference in *P* element presence indicates that *P* element was introduced to *Drosophila melanogaster* wild type populations in the early 20th century (reviewed by Pinsker *et al*, 2001).

P element effects can be divided into two classes. First, there are effects in the maternal germline (maternal effects), such as suppression of hybrid dysgenesis and suppression of *P[lacZ]* expression in germinal tissue. This class includes a phenotype called P cytotype which is the ability to suppress hybrid dysgenesis when the mother has the *P* element. Second, *P* elements affect phenotypes in somatic tissue (zygotic effects) such as suppression of expression of *P[lacZ]* inserts, $vg^{21\text{-}3}$, sn^{w}, and *P* element dependent silencing (PDS) (Robertson and Engels, 1989; Lemaitre and Coen, 1991; Lemaitre *et al*, 1993; Hodgetts and O'keefe 2001; Bushey and Locke, 2004). Since these effects are not dependent upon paternal or maternal inheritance of *P* elements, I will avoid using the term "P cytotype" for these effects to avoid any confusion. I will refer to these phenotypes as somatic *P* element phenotypes.

Hybrid dysgenesis and Different cytotypes

Hybrid dysgenesis is a complex of different symptoms in progeny of crosses between a male that has an autonomous *P* element and females that do not have it. These symptoms include high mutation rate, male recombination, and male sterility (Engels and Peterson, 1980). Hybrid dysgenesis does not happen when *P* elements are inherited maternally. The difference in reciprocal crosses ends up defining cytotypes. Cytotype refers to presence/ absence of cytoplasmic factor(s) that suppress transposition and hybrid dysgenesis by suppressing expression of transposase or inhibiting the transposition process (Sved, 1987).

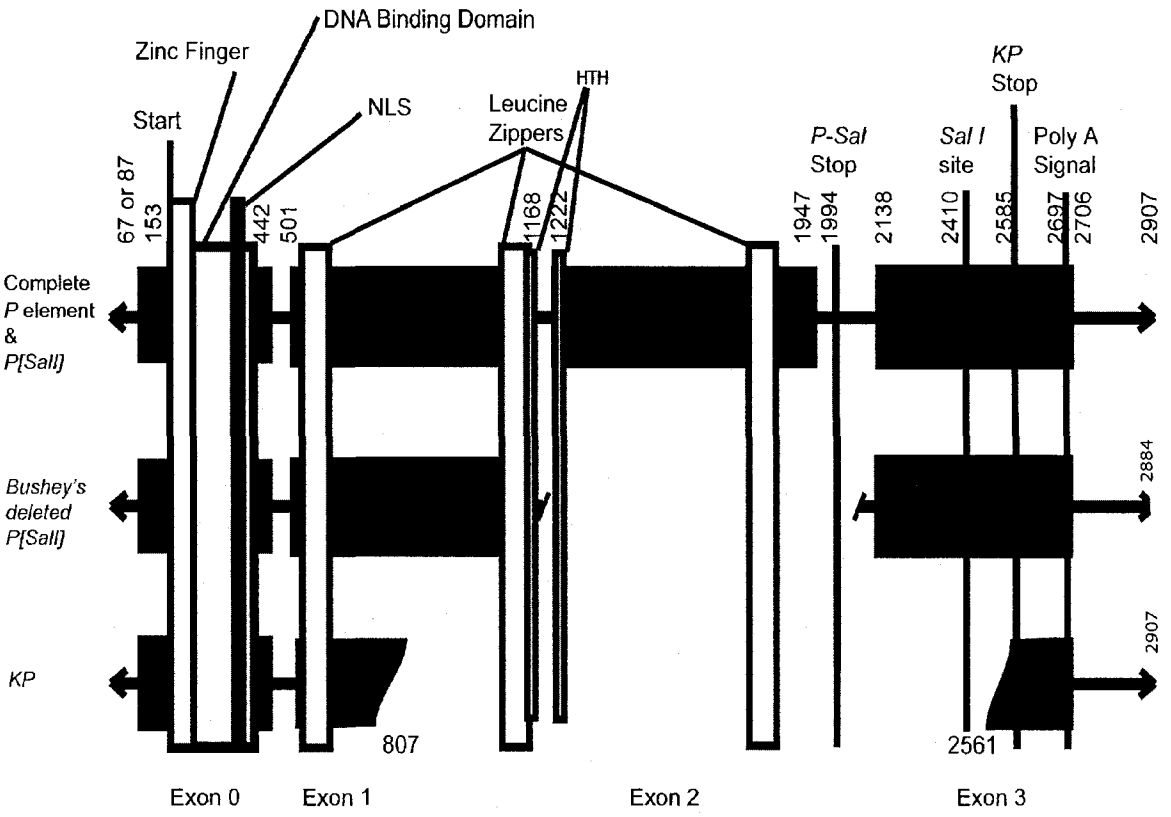


Figure 1-1) *P* element genes and their important domains.

The first element is a complete *P* element. The *P* element gene of *P[SalI]* is similar to this element except for a 4 bp insertion at the *Sal*I site. Bushey's deleted *P[SalI]* lacks exon 2 plus part of intron 1-2 and intron 2-3 (Bushey, 2004). *KP* is a naturally occurring *P* element that has a deletion of base pairs 808-2560.

The predicted structural domains of the polypeptide are shown along with the exon-intron boundaries (O'Hare and Rubin, 1983; Karess and Rubin, 1984; Rio, 1990). Hodgetts and O'keefe (2001) reported that transcription starts at base pair 67 instead of 87 (Karess and Rubin, 1984). Start: translation start codon, NLS: nuclear localization signal, HTH: helix-turn-helix domain, *P-Sal* Stop: translation stop codon for the type I repressor protein (P-Sal protein), *Sal*I site: the *Sal*I restriction site that is mutated in *P-Sal*, *KP* Stop: translation stop codon for the KP protein. Arrowheads indicate the presence of 31 inverted repeats at the ends (Except *P[SalI]* that has a 23 bp deletion at its 3' end). Numbers indicates location of each splicing site or domain on the complete *P* element sequence.

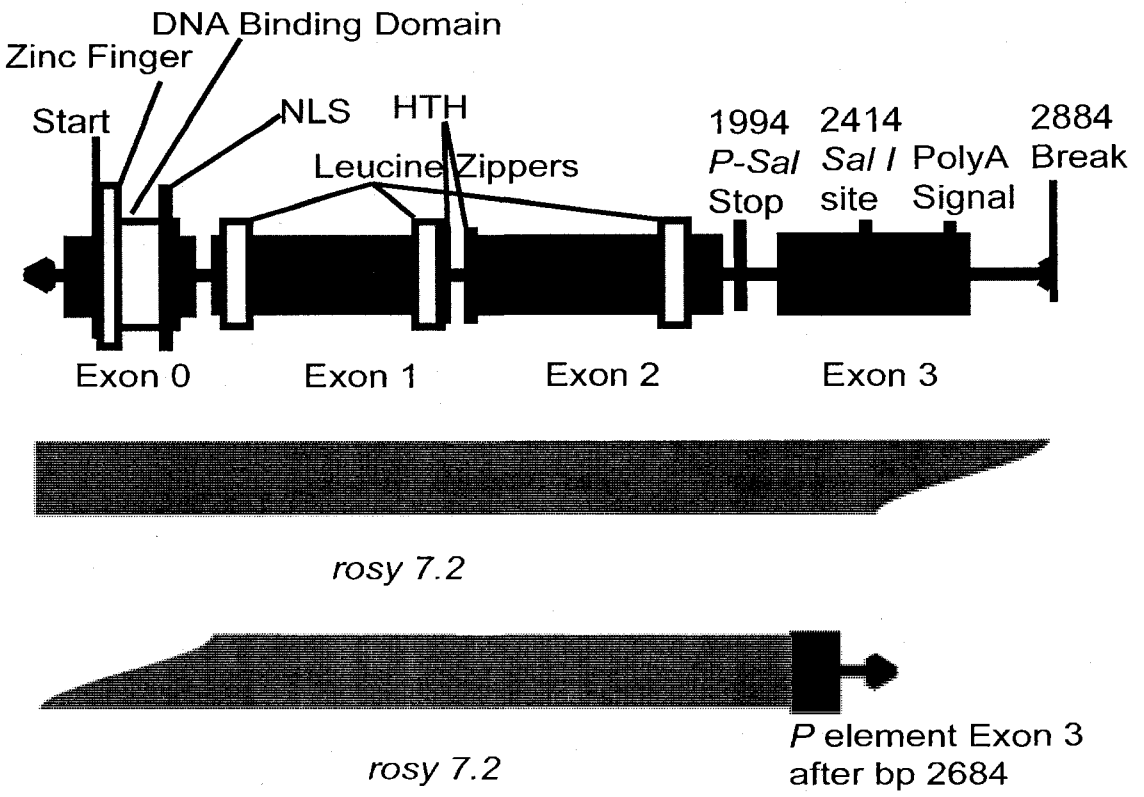


Figure 1-2) *P[SalI]* construct.

The mutated *Sal*I site is located at 2410-2415 and there is a 4 bp TCGA insertion between 2414 and 2415. In *P[SalI]*, the last 23 bp of *P* element sequence (base pairs 2884-2907) is replaced by a $ry^{7.2+}$ gene plus the last 224 bp of the complete *P* element sequence (*P* element base pairs 2684-2907). Therefore, the *P* element has the first 2884 bp of a complete *P* element (Karess & Rubin, 1984)

Start: translation start codon, NLS: nuclear localization signal, HTH: helix-turn-helix domain, *P-Sal* Stop: translation stop codon for P-Sal protein, *Sal*I site: the *Sal*I restriction site that is mutated in *P[SalI]*, *rosy 7.2*: $ry^{7.2+}$. Arrowheads represent 31 inverted repeats at the ends.

A *P[SalI]* insert at cytological location of *89B*, named *P-Sal*, has been used in this study.

Since *D. melanogaster* strains vary in their ability to induce and/or repress hybrid dysgenesis, they have been classified according to these abilities. P strains have 30-50 *P* elements per genome. These elements consist of approximately one third complete 2.9 kb *P* elements and two thirds deletion derivatives of *P* element (O'Hare *et al*, 1992). Progeny of a cross between a P strain male and M strain female show complete hybrid dysgenesis (dysgenic cross). P strain females suppress hybrid dysgenesis. Q strains are weaker versions of P strains that have a similar number of *P* elements, but induce a weaker version of hybrid dysgenesis when Q strain males were crossed to M strain females. Q strain females are able to suppress hybrid dysgenesis when crossed to P strain males. M' strains, which are also called pseudo-M strains, have some deletion derivatives of *P* elements and partially suppress hybrid dysgenesis in a dysgenic cross but are not able to induce it in a cross with M strain females (Reviewed by Rio, 1990). In contrast to Q strains, M' strain transposition suppression does not show a maternal effect and they carry 50-60 copies of *KP* elements per genome (Black *et al*, 1987). Finally M strains lack *P* elements and are not able to induce hybrid dysgenesis or suppress hybrid dysgenesis (Reviewed by Rio, 1990; reviewed by Castro and Carareto, 2004). There is no current M strain in wild *Drosophila melanogaster* populations. Most of the current populations in the wild are either Q or P strains, and therefore hybrid dysgenesis can't be seen in wild populations any more (reviewed by Castro and Carareto, 2004).

Models for P cytotype mechanism

Rio (1990) formalized four different models for the mechanism of transposition suppression. In the first model, a dimerization/ polymerization poisoning model, the repressor protein binds to transposase and deactivates it. In support of this model, Andrews and Gloor (1995) showed that point mutations in the leucine zipper of the KP protein, a classic type II repressor, decrease the potency of the mutated *KP* element (under control of an *actin 5C* proximal promoter in *P[wactKP]*) to suppress transposition induced by *P(ry(+), Δ2-3)99B* but they did not compare the mutant and parental flies for the amount of the KP protein. Since mutant and parental constructs were inserted in different genomic locations, the amount of expression of constructs and levels of the KP protein could be different in the two strains of flies and this can contribute to differences in phenotype. Moreover, Lee *et al* (1998) showed that a short version of the KP protein that has lost its dimerization domains is still able to suppress transposition *in vitro*. This is contradictory to the dimerization poisoning model.

The second model is based on the hypothesis that deletion derivatives of a *P* element can not encode transposase while they still have transposase binding sites. Therefore, any increase in

their number does not change the level of intra-nuclear transposase but increases the number of transposase binding sites. This leads to a lower number of transposase molecules per binding site and suppression of transposition due to lack of enough enzyme at each site. This model does not need any P repressor protein production. The observation that one *P[SalI]89D* element is enough to suppress transposition (Robertson and Engels, 1989) contradicts this hypothesis.

The third model involves competition between transposase and repressor protein in binding to DNA sites. The results of Lee *et al* (1998) support this model. They showed that the KP proteins that are mutated in the DNA binding domain are unable to suppress transposition *in vitro*. In contrary, Bushey (2004) showed that loss of exon 2 in *P[SalI]* (Figure 1-1 and Figure 1-2) is enough to eradicate the ability of *P[SalI]* to suppress PDS, although this deletion does not affect the *P[SalI]* DNA binding domain. Although these results contradict with the transposase-repressor protein competition model, Bushey did not test protein levels. Therefore, differences in protein production/ stability can contribute to the loss of phenotype as well. I was not able to find any data in the literature that show a P protein that is capable of binding DNA specifically is unable to suppress the transposition.

The last model is the suppression of *P* element gene expression due to binding of the P repressor protein to the *P* element promoter. It is known that the transposase binding site on *P* element DNA sequences overlaps part of the *P* promoter (Kaufman *et al*, 1989). The transposase DNA binding domain is the first 88 amino acids, which are also present in all the type I repressors and many of the type II repressors as well, thus these repressor proteins are able to bind to the same area of the *P* element (Gloor *et al*, 1993; Lee *et al*, 1998). Kaufman and Rio (1991) showed that transposase suppresses transcription from a *P* element promoter *in vitro*. *In vivo*, a P cytotype suppresses expression of a *P-lacZ* transgene in a promoter-dependent manner, but transposase encoded by *P(ry$^+$, Δ2-3)99B* does not have the same effect (Lemaitre and Coen, 1991). Moreover, there are some type II repressors, such as the *SP* element, that do not have this domain in their coding sequence but still are able to suppress some phenotypes that are related to transposition (Gloor *et al*, 1993). There are other *P* constructs such as clusters of *P[lacW]*s and telomeric *P[lacZ]* transgenes that are able to suppress transposition but they do not encode P proteins (Ronsseray *et al*, 1998; Ronsseray *et al*, 2001). Therefore, in these P cytotype systems P protein-DNA binding and even P protein production is not necessary.

In addition to Rio's models for P cytotype, involvement of heterochromatinization and RNAi are suggested to explain phenotype of telomeric *P* inserts. First, *Su(var)205* mutations can suppress telomeric effects (suppression of transposition by telomeric *P* inserts) while the regular

P strains are insensitive to *Su(var)205* (Haley *et al*, 2005). Second, mutations in *aubergine* but not *piwi* and *homeless*, which are involved in the RNAi pathway, also suppress telomeric effects but not regular P cytotype (Reiss *et al*, 2004; Simmons *et al*, 2007). There is also some evidence for siRNA in the control of *gypsy* retrotransposition hybrid dysgenesis (Sarot *et al*. 2004). However, the current data make it difficult to conclude that this is an inhibition of transcription. Role of RNAi, as a mechanism in targeting of heterochromatinization in *Drosophila*, still needs to be investigated (reviewed by Riddle and Elgin, 2008).

In summary, there is not a single model that can explain the results of all experiments that have been done in this field. It seems that a combination of models or a completely new model is needed to explain suppression of transposition of *P* elements in different cases.

P element proteins

Tissue specific alternative splicing

A complete *P* element is 2907 bp, includes 31 bp inverted repeats at ends, and a four exon gene (exons 0, 1, 2, and 3). In germ line cells, this gene encodes an 87 kD enzyme, transposase, which is responsible for the process of transposition when the embryo receives the *P* element paternally (Rio *et al*, 1986). In somatic tissues the third intron is not spliced out and the resulting premature stop codon at position 1992 leads to a 66 kD truncated protein, which acts as a repressor of transposition and is called the P repressor protein (Laski *et al*, 1986. Rio *et al*, 1986).

P(ry^{+}*, Δ2-3)99B* is a *P* element that lacks the last intron (intron 2-3) and is able to encode transposase in all tissues (Robertson *et al* 1988). It produces a higher level of transposase than a complete *P* element (Simmons *et al*, 2002) but its transposase activity can not be transferred to oocytes through a maternal effect. Transfer of transposase mRNA from mother to the oocyte is dependent on the presence of intron 2-3. A nine nucleotide motif (CTGTTTCTT), that is transcribed from base pairs 2089-2097 on DNA sequence, similar to sequences thought to be involved in the maternal transmission of *bicoid* and *nanos* RNAs has been found in this intron and most probably is responsible for transfer of transposase mRNA to oocytes (Simmons *et al*, 2002).

Type I repressors encode a 66 kD repressor protein (Laski *et al*, 1986; Gloor, 1993). These elements have all the *P* element DNA sequence through to the middle of last intron (at least 1956 bp). Type II repressors are shorter and their deletions may include most of the coding sequence of the first three exons. Although the mechanism of this activity can be different, both types share the ability to suppress transposition (Gloor *et al*, 1993).

Transposase protein and IRBP

Transposase is an 87 kD protein that is encoded by the complete *P* element in germ line cells (Rio *et al*, 1986). It binds near both ends of *P* element DNA sequences. At the 5' end it binds to base pairs 48-68 that overlap the promoter area (base pairs 58-103) and at the 3' end it attaches to base pairs 2858-2867. Both sites have a 10 bp AT-rich consensus (ATACACTTAA at the 5' and ATCCACTTAA at the 3' site), which is located 52 bp from the 5' end and 40 bp from the 3' end respectively. Transposase does not bind to any other *P* element sequence specifically (Kaufman *et al*, 1989). Kaufman *et al* (1989) also showed that in spite of specific binding to *P* element DNA sequence, transposase binding to genomic DNA is non-specific.

Weinert *et al* (2005) showed that P transposase mediated DNA cleavage happens in both G1 and G2 arrested cells, therefore, it does not need DNA replication. Transposase exists as a pre-formed tetramer (Tang *et al*, 2007). For DNA cleavage, transposase needs 138 bp at the 5' end and 216 bp at the 3' end of a *P* element (Beall and Rio, 1997). Before cleavage, a protein-DNA pre-synaptic complex forms at both ends of a *P* element (Beall and Rio, 1997; Tang *et al*, 2005). Then, in the presence of GTP or non-hydrolyzable GTP analogs these complexes proceed to synapse rapidly while the DNA cleavage happens slowly (Tang *et al*, 2007). During cleavage, transposase induces a 17 base 3' staggered single strand tail at each end of the *P* element that is protected from exonuclease activity by a stable protein complex including inverted repeat binding protein (IRBP) (Beall and Rio, 1997).

IRBP, a 66 kD *Drosophila* protein, attaches to the external 16 bp of the 31 bp inverted repeat and facilitates transposition of the *P* element (Rio and Rubin, 1988). IRBP has a mammalian homologue, the Ku70 subunit of Ku antigen (Beall *et al*. 1994). Mammalian Ku antigen, a heterodimer composed of 70 kD and 80 kD polypeptides, is the DNA-binding subunit of a DNA-dependent protein kinase. This kinase complex is involved in double strand break DNA damage repair, and VDJ recombination in antibody producing cells (Gottlieb and Jackson, 1993). Therefore, it is possible that IRBP is involved in the process of DNA repair after DNA cleavage by P transposase.

After P transposase mediated DNA cleavage, the resulting DNA double-strand break is repaired by two predominant mechanisms; homologous recombination which is restricted to the post-replicative G2 phase of the cell cycle, and non-homologous end joining (NHEJ) that may occur throughout the cell cycle. It has been shown that NHEJ happens even in the post-replicative stage of G2 arrested cells (Weinert *et al*, 2005).

Dose dependency of transposase and repressor proteins

Simmons *et al* (2002) studied synergistic effects between different sources of transposase by measuring repression of sn^w mutability and gonadal dysgenesis. They used a complete *P* element gene under the control of a *Drosophila hsp70* promoter that was inserted as a *hobo* transgene *H(hsp/CP)*. Not only they could not find any synergism between different *H(hsp/CP)* inserts, but also one of the *H(hsp/CP)* inserts showed an antagonistic effect against $P(ry^+, \Delta 2\text{-}3)99B$ (Simmons *et al* 2002).

Misra *et al* (1993) showed that one copy of the *P[ry,66K]* element, a type I repressor, is enough to suppress transposition in a position dependent manner, but they could not find a correlation between the amount of 66 kD repressor protein expression from a modified *P[ry,66K]* construct during oogenesis and the intensity of maternal effect on transposition repression. This can be explained by the fact that in the modified *P[ry,66K]* construct, part of last intron was deleted including the 9 nucleotide motif that is necessary for mRNA transport to the oocyte (Simmons *et al*, 2002). Lee *et al* (1998) showed that a type II repressor, *KP* element, suppresses transposition in a dose dependent manner *in vitro*.

***P* element homologues in eukaryotes and conservation studies**

Hammer *et al* (2005) tested genomes of 26 vertebrates for presence of *P* element homologous DNA sequences. Five tested rodents have only rudiments of this sequence while the rest of the tested genomes have the *P* homologous sequences.(Hammer *et al*, 2005). Hagemann and Pinsker (2001) reported a human homolog for the *Drosophila P* element. *Phsa* is a 19533 bp single copy gene on the long arm of chromosome 4 and has six exons and five introns. *Phsa* encodes a 903 aa polypeptide, which is found in a variety of tissues (Hammer *et al*, 2005). *Pgga*, the homologous chicken gene is located at the orthologous position of the long arm of human chromosome 4. *Phsa* and *Pgga* do not have any inverted repeats while the zebrafish homologue of *P* element gene, *Pdre*, does. This makes *Pdre* a possible candidate for transposition (Hammer *et al*, 2005). None of these proteins have the leucine zipper domain and only Phsa has a THAP domain (see below) (Hammer *et al*, 2005). Quesneville *et al*, (2005) reported that *Phsa* is the coding sequence of a larger gene that is called *THAP9*. The THAP9 protein has 21% similarity to *Drosophila melanogaster* P element transposase (Quesneville *et al*, 2005). The maximum divergence was observed between exon 3 of the *P* element and its human homologue (Hagemann and Pinsker, 2001).

P element protein domains

THAP domain: A conserved DNA binding domain and nuclear localization signal

Lee *et al* (1998) showed that the first 88 amino acids that constitute a C2HC zinc finger domain is enough for specific DNA binding of the KP protein and its suppression of transposition *in vitro*. Miller *et al* (1999) showed this N terminal DNA binding domain is conserved in 12 mobile and immobile *P* element related sequences in seven species of Diptera (Miller *et al*, 1999). In the set of different protein sequences that Miller *et al* (1999) published, I also found a conserved nuclear localization signal (KRRRL) near the C-terminus of the DNA binding domain.

This conserved N terminal zinc finger domain has been seen in more than 100 eukaryotic genes with different functions such as proliferation, apoptosis, cell cycle, chromosome segregation, chromatin modification, and transcriptional regulation. It is also called THAP [Thanatos (death, from the Greek god of death)-associated protein] domain after it was first found in an apoptotic related human protein, THAP1 (Roussigne *et al*, 2003; Clouaire *et al*, 2005). THAP domains have been found in mouse, rat, pig, cow, chicken, *Xenopus*, zebrafish, *Caenorhabditis elegans* and *Drosophila*, but are restricted to the animal kingdom, as there is no known or predicted proteins containing THAP domains in plants, yeast, fungi or bacteria (Roussigne *et al*, 2003). There are seven known or putative *Drosophila* THAP proteins. Three of them have a C2H2 zinc finger and are described in Flybase as putative transcription factors. One of them is already known as *DIP2*, a transcription factor. Another, *CG13894*, has homology to centromere protein B. Two of *Drosophila* THAP proteins have more than one THAP domain. *CG14860* has two domains and *CG10631* has 27 (Roussigne *et al*, 2003).

The THAP domain belongs to the zinc finger superfamily and has a C2CH consensus (Cys-Xaa_{2-4}-Cys-Xaa_{35-50}-Cys-Xaa_2-His) (Roussigne *et al*, 2003). Liew *et al* (2007) showed that the THAP domain from the "C-terminal binding protein" of *C. elegans* has a positively charged protein fold and binds to double strand DNA. This confirms the previous data about DNA binding activity of this domain. The THAP consensus, C2CH, is different from C2HC consensus in the *P* element (Cys-Xaa_3-Cys-Xaa_9-His-Xaa_3-Cys). Lee *et al* (1998) showed that mutation of the first two cysteines of the conserved C2HC signature, or deletion of the AVPTIF consensus box, abolishes site-specific DNA-binding activity of the *P* element transposase (Lee *et al*, 1998; Roussigne *et al*, 2003). The same results are true for THAP1, the human prototype of this protein group. Site-directed mutagenesis of single cysteine or histidine residues as well as the four other conserved residues (P, W, F, and P), which define the THAP consensus sequence, also abolished

DNA-binding ability (Clouaire *et al*, 2005). P transposase binds to a 10 bp AT-rich consensus $AT^{A}/_{C}CACTTAA$ while THAP1 recognizes an 11 bp sequence AG**TAAGGGCAA** and mutation in the T or GGCA motif abolishes this binding (Kaufman *et al*, 1989; Clouaire *et al*, 2005).

Based on evolutionary conservation of the THAP domain, I expected to find many mutations in this domain during my *P[SalI]89D* mutagenesis experiment.

Leucine zippers

Rio (1990) predicted three leucine zippers within three heptad hydrophobic repeats in the type I repressor polypeptide. These repeats are located at following amino acid positions:

- 101 (leucine), 108 (leucine), 115 (leucine), 122 (leucine): LLLL
- 283 (leucine), 290 (leucine), 297 (valine), 304 (leucine), 311 (leucine): LLVLL
- 497 (isoleucine), 504 (leucine), 511 (leucine), 518 (glutamine), 525 (leucine): ILLQL

Lee *et al* (1998) showed that the LLLL leucine zipper at position 101-122 is effective in dimerization of the KP protein but this dimerization is not necessary for the ability of the KP protein to suppress transposition *in vitro*. The KP protein does not have the other two leucine zippers and there is no data available about their function. However, a conservation study showed that these two zippers are conserved in 12 mobile and immobile *P* elements isolated from seven species of Diptera and six of these were from the Drosophilidae family (Nouaud and Anxolabéhère, 1997; Miller *et al*, 1999).

Helix turn helix

Rio (1990) also found a weak sequence homology between positions 308-327 and a bacterial helix turn helix DNA binding domain. This motif has been conserved in *P* elements of seven different species of Diptera (Nouaud and Anxolabéhère, 1997).

P element constructs that induce PDS

P[SalI]89D

A 10.3 kb ry^{+} bearing *P* construct and a classic type I repressor, *P[SalI]* has a 4 bp insertion at base pair 2410 within the last exon of the *P* element (exon 3), and is consequently unable to produce an active transposase in any tissue but is still able to encode the type I repressor protein (Karess and Rubin, 1984; Robertson and Engels, 1989) (Figure1-1 & Figure 1-2). Robertson and Engels (1989) studied *P[SalI]89D* and found that it can suppress transposition in both somatic and germ line tissues but it is different from a wild type repressor. While it is unable to suppress the transposition effect of many copies of the wild type *P* element ($\pi 2$ strain) measured by gonadal dysgenesis sterility, it suppresses the transposition effect of a single *P*(ry^{+}, *Δ2-3)99B* element strongly in both somatic tissue (measured by pupal lethality) and germ line

tissue (measured by gonadal dysgenic sterility). Moreover, suppression of the *P(ry$^+$, Δ2-3)99B* transposition effect does not show the reciprocal cross effect (*i.e. P[SalI]89D* is able to suppress transposition when it is transferred to embryo from either parent).

KP element

The *KP* element, a classic type II repressor, is a naturally occurring derivative of a complete *P* element that lacks base pairs 808-2506. This element has repressor effects on transposition (Black *et al*, 1987). Andrews and Gloor (1995) showed that the *KP* element (under control of an *actin 5C* proximal promoter in *P[wactKP]*) can suppress transposition ubiquitously both in the germ line (by Harwich P strain) and in the somatic tissues (by *P(ry+, Δ2-3)99B*).

The *KP* element is present in 50-60 copies per genome in all naturally occurring M' strains and its DNA sequence is identical in all of them (Black *et al*, 1987; Wook *et al*, 1996; Itoh *et al*, 2007). All *KP* elements have a T instead of an A at position 32 (Black *et al*, 1987). The only exception is a *KP* stock with G at position 32 that is reported by Itoh *et al* (1989), but the authors thought that this is most likely a second mutation that happened later during *KP* spreading in the world (Itoh *et al*, 1989). Although the A32T variation is reported in complete *P* elements, it is not the most common sequence for this location among them (O'Hare and Rubin, 1983). This indicates that all *Drosophila melanogaster* populations received the *KP* element from the same source (Itoh *et al*, 2007). Jackson *et al* (1988) showed that in laboratory stocks maintained for several generations in the presence of a source of transposase, the number of *KP* elements in the genome increased and the intensity of hybrid dysgenesis decreased. Moreover, comparison between current North American wild *Drosophila melanogaster* populations and samples that were gathered in 1980s showed that complete *P* and *KP* elements were the two major classes of *P* elements present in the genomes of current populations and the prevalence of the *KP* element has increased while the frequency of P and Q strains is still the same over these years (Itoh *et al*, 2007). This helped Itoh *et al* (2007) to hypothesize that *KP* spread in wild populations and the increase in *KP* genomic copy number is probably due to a transpositional advantage of *KP* elements rather than a positive selection for its role in *P* element regulation.

KP protein

A *KP* element is 1254 bp in length and encodes a 207 aa polypeptide. The first 199 amino acids are identical to the transposase and the type I repressor. The last 8 amino acids are encoded by an out of frame translation of exon 3. *KP* is not able to encode either transposase or type I repressor protein but it is still able to suppress hybrid dysgenesis and this suppression effect is

transmitted through both sexes, which is distinct from maternally transmitted P cytotype (Black *et al*, 1987).

The transposase protein binds to the transposase binding sites, but not to 31 bp inverted repeats or to 11 bp inverted repeats (Kaufman *et al*, 1989). Although the *KP* encoded protein has the same DNA binding domain as transposase and binds to transposase binding sites, it also binds to the 11 bp inverted repeats, as well as to the 31 bp inverted repeats when added in higher concentrations *in vitro* (Lee *et al*, 1996). In addition to a DNA binding domain, the KP protein has two dimerization domains, a leucine zipper (LLLL) which is shared with transposase at amino acid positions 101 through 122, and an unidentified domain in the last 69 amino acids at the C-terminus. While a DNA binding domain is necessary for suppression of transposition by the KP protein, dimerization is not necessary for suppression of transposition although it increases DNA binding affinity. In other words, the *KP* suppression effect on transposition is not through multimer poisoning (Lee *et al*, 1998). Andrews and Gloor (1995) showed that amino acid substitution in the first leucine of the KP protein leucine zipper decreases the ability of a *P[wactKP]* (that has a *KP* downstream of an *actin 5C* proximal promoter) to suppress *P(ry*$^{+}$*, Δ2-3)99B* induced transposition *in vivo*. Even changes in some of the other amino acids in leucine zippers (while all four key leucines are intact) affected this phenotype (Andrews and Gloor, 1995). Although these findings may indicate a leucine zipper function in transposition suppression, Andrew and Gloor did not compare the levels of the KP protein in the mutants versus parental flies. Therefore we do not know that this lack of transposition suppression is due to the mutant KP protein instability or its lack of dimerization.

Lee *et al* (1998) showed that the KP protein suppresses transposition at both DNA cleavage and strand transfer stages *in vitro*, but to see this effect the KP protein concentration needs to be 30 times more than transposase concentration. Simmons *et al* (2002) found that a *KP-hobo* transgene under control of a *Drosophila melanogaster hsp70* heat shock promoter *H(hsp/KP)* was able to suppress transposition from the modified *P* element in the *P(ry*$^{+}$*, Δ2-3)99B* transgene more effectively than that encoded by a complete *P* element in the *H(hsp/CP)2* transgene. They also did not find any maternal effect, or *KP* effect on gonadal dysgenesis. Furthermore, repression of *sn*^{w} mutability showed a zygotic pattern of inheritance which is in accordance with the findings of Black *et al* (1987). It has been shown that the KP protein suppression effect on transposition is dose dependent *in vitro* (Lee *et al*, 1998). This is in accordance with Jackson *et al* (1988) that suggested a dose dependent relation between the number of *KP* elements and the strength of suppression of hybrid dysgenesis. However, Simmons

et al (2002) could not strengthen the *H(hsp/KP)* suppression effect on gonadal dysgenesis or repression of sn^w mutability by adding a type I repressor of *H(hsp/CP)2*. Therefore, it seems that *KP* affects some systems in a dose-dependent way while in some others this is not the case.

P element constructs that are affected by PDS

P{lacW}, an enhancer trap construct, has a *lacZ* gene under the control of a *P* element promoter and a *mini-white* gene as a visible marker under control of the *white (w)* gene promoter (Figure 1-3) (Bier *et al*, 1989). In a w^- background this transgene normally produces a uniform red eye color if it is inserted in euchromatin (Figure 1-5). However, $P\{lacW\}ci^{DPlac}$ is an exception. $P\{lacW\}ci^{DPlac}$ (hereafter referred to as *Pci*) is a *P{lacW}* construct that is inserted 3 kb upstream of the first exon of the *ci* gene and 6 kb downstream of last exon of the *RpS3a* gene on chromosome 4 (Figure 1-4). In *Pci* the expression of the *mini-white* gene may be silenced by the presence of certain *P* elements, inducing a variegated eye phenotype (Figure 1-5). This variable silencing has been called P element dependent silencing (PDS) (Bushey and Locke, 2004). Insertion of a *gypsy* transposable element distal to *Pci* changes its phenotype from a uniform red eye to a variegated eye, and the presence of *P[SalI]89D* enhances this variegation. These *Pci* alleles are called $P\{lacW\}ci^{E1}$ (hereafter *E1*) and $P\{lacW\}ci^{E2}$ (Figure 1-4) (Bushey and Locke, 2004).

P[hsp26-pt-T] is another w^+ *P* construct with a *mini-white* gene under control of an $hsp70^+$ promoter. A 744 bp barley cDNA (from barley cDNA clone pcSIP1), under control of an *hsp26* promoter is cloned upstream of an *hsp70-white* reporter gene in an *A412 P* element vector (Figure 1-6). This is a unique sequence within *Drosophila* genome and helps tracking of the element. This construct produces a uniform dark red eye color in non-heat shocked conditions (Wallrath and Elgin, 1995). Different *P[hsp26-pt-T]* inserts on chromosome 4 of *Drosophila melanogaster* have been used in this study. $P\{hsp26\text{-}pt\text{-}T\}ci^{2\text{-}m1021.R}$, that is placed just 140 bp more proximal to the *ci* than the *Pci* (Bushey, 2004) (Figure 1-4), is variably silenced by *P* elements. *P{hsp26-pt-T}39C-12* is another *P{hsp26-pt-T}* insert on chromosome 4 (genomic location 102B) that induces a variegated eye color by itself and this variegation is inhibited by the presence of *P* elements (Sun *et al*, 2000; Bushey, 2004).

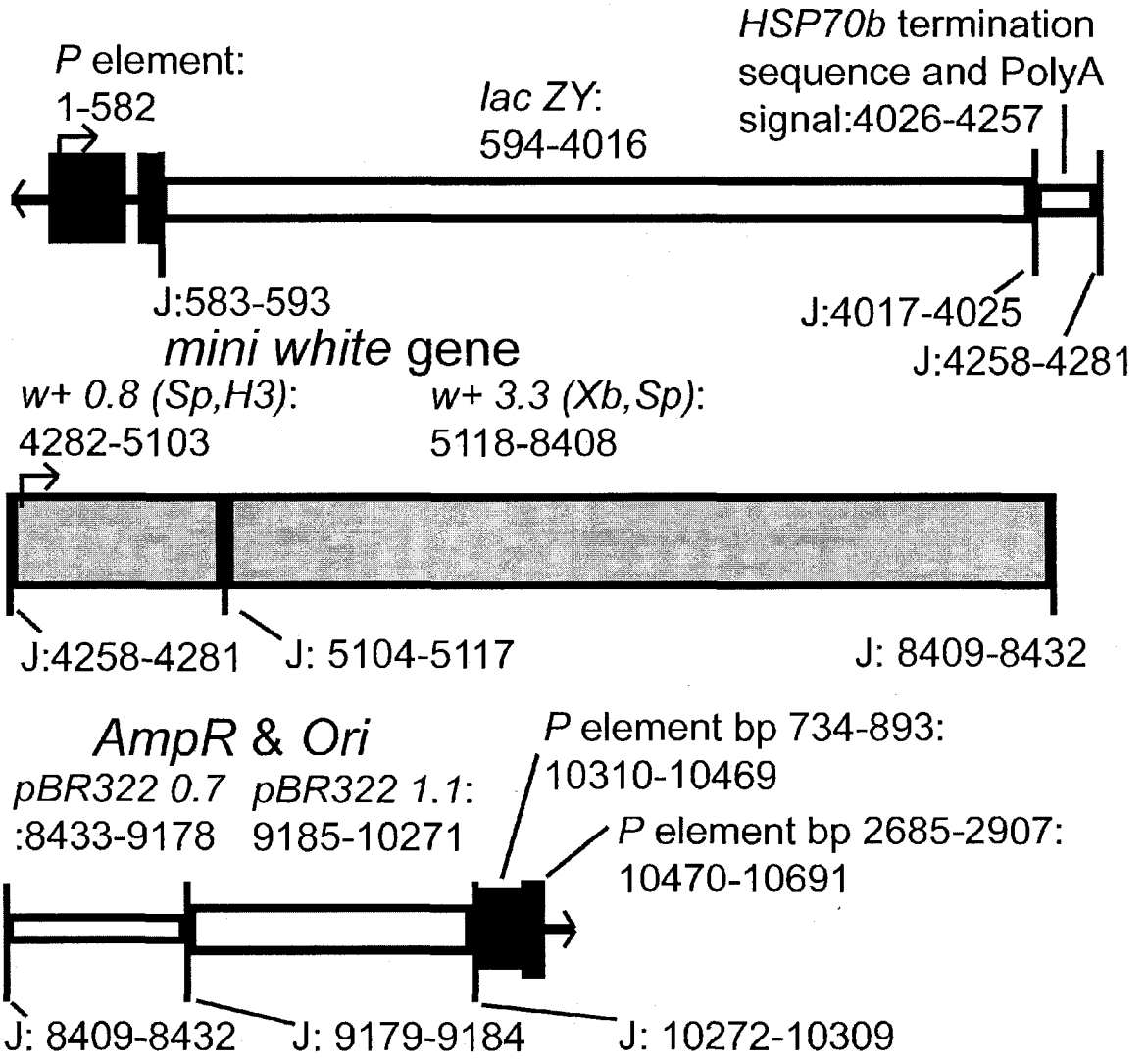


Figure 1-3) Diagram of the *P{lacW}* construct found in *P{lacW}ci* and its derivatives. There are *P* element sequences at both ends of the construct. Arrows shows the direction of transcription from two promoters present within the construct. Beginning and end locations of each segment are mentioned. J indicates short joint sequences. Arrowheads represent 31 bp inverted repeats (From FlyBase sequence FBtp0000204).

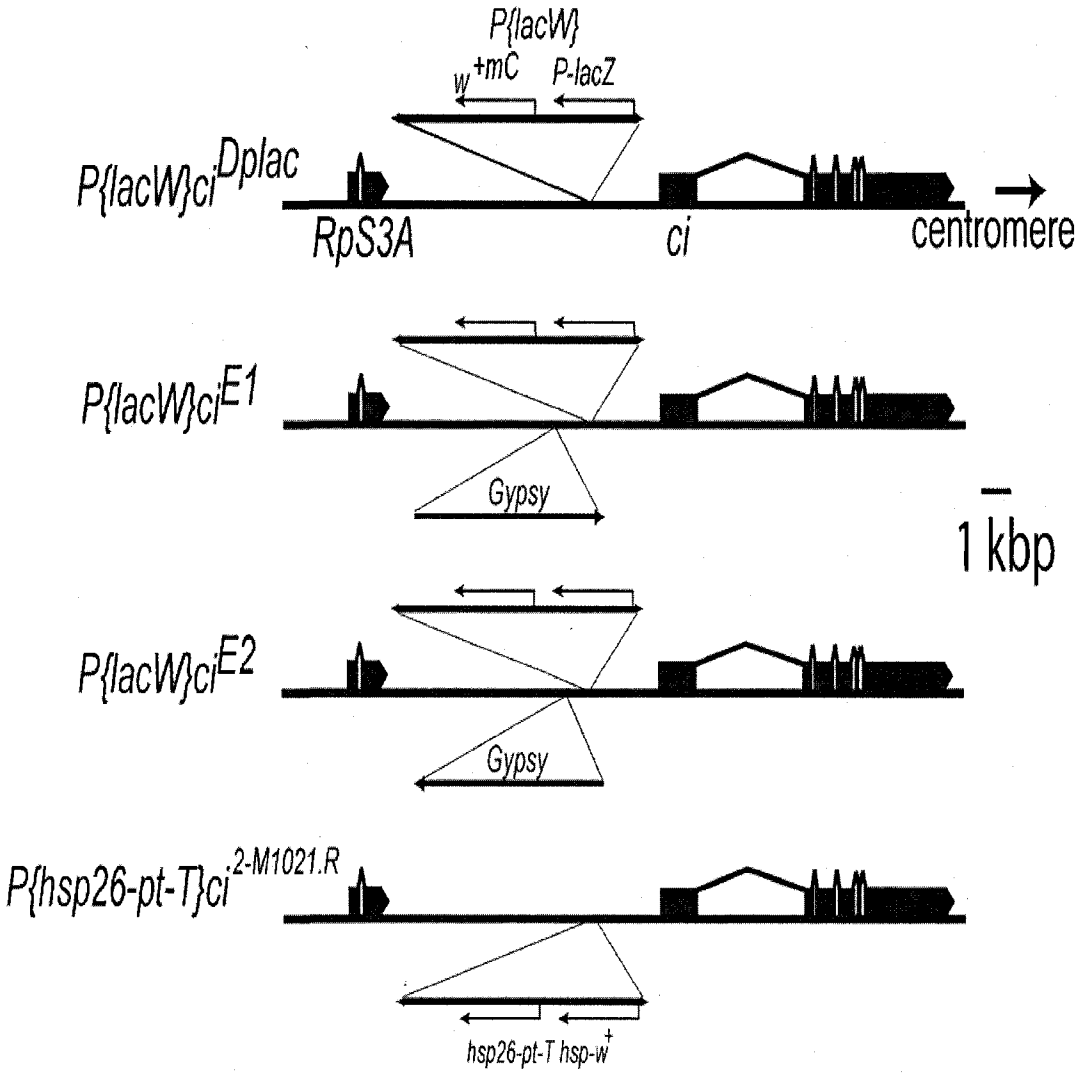


Figure 1-4) *Pci* and *P{lacW}ci^{E}* alleles.

P{lacW}ci^{E1} (*E1*) has a *gypsy* element inserted between a 4 bp duplication (TACA) starting at 81,294 (AE003854). *P{lacW}ci^{E2}* has a *gypsy* element inserted in between 4 bp duplications (TATA) starting at nucleotide 80,747 (AE003854). Arrows indicate the direction of transcription (from Bushey and Locke, 2004).

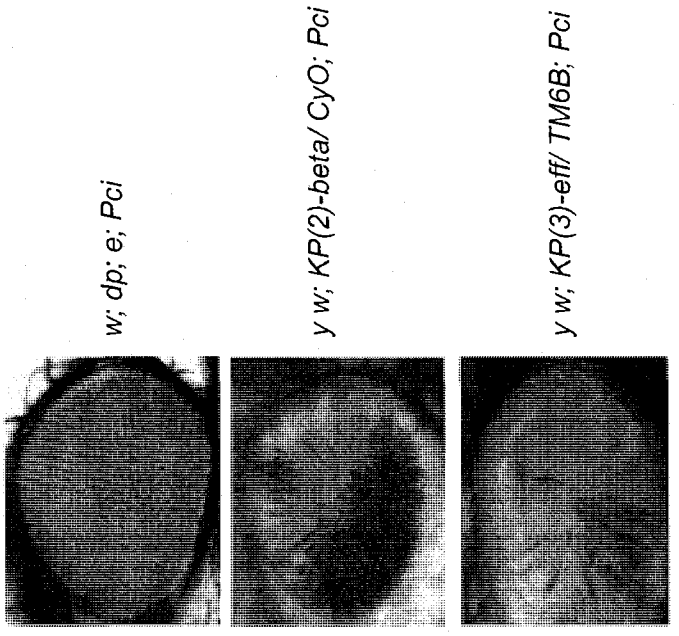


Figure 1-5) Range of PDS phenotype.

Pictures were taken of eyes of male flies from genotypes that were mentioned above each photograph. The first photograph is a *Pci* homozygote male fly without any other *P* element that shows uniform red eye color. Other photographs show different levels of silencing of *Pci* expression due to the presence of a *KP* insert in the genome.

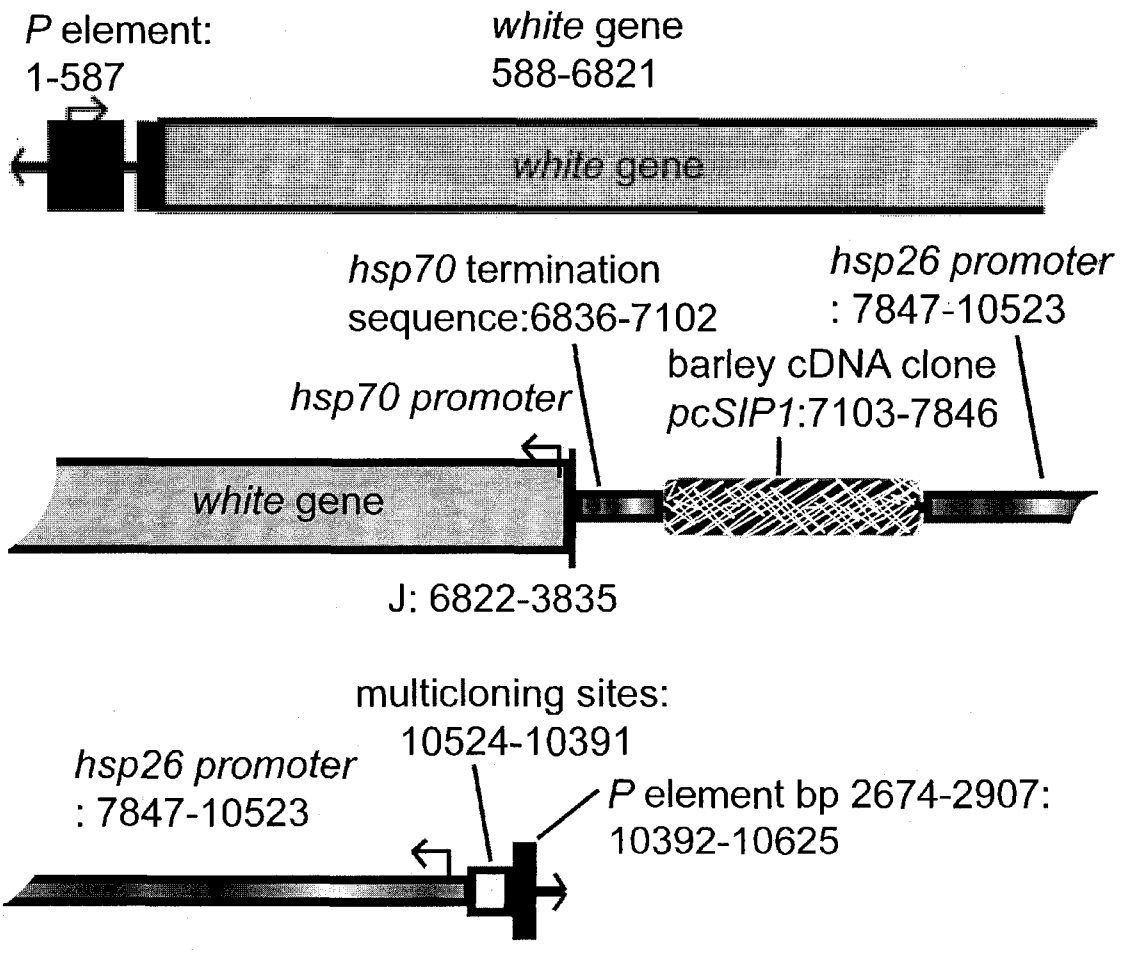


Figure 1-6) Diagram of the *P{hsp26-pt-T}* construct.

There are *P* element sequences at both ends of the construct. Arrows shows the direction of transcription from three promoters present. Note that the direction of transcription of both genes is opposite to direction of transcription of the *P* element promoter. Numbers indicate the position of first and last base pair of each segment in the whole construct. There is a barley cDNA sequence in the construct which is unique in *Drosophila* genome. This helps tracking of the element. J indicates short joint sequences. Arrowheads represent 31 bp inverted repeats (Personal communication with K. Haynes in Dr S. Elgin's lab).

Heterochromatin and hPEV

Heterochromatin (vs. euchromatin) is a cytological term. It describes the chromatin that remains condensed during interphase and replicates at late S phase of the cell cycle. It is more abundant at centromeres and telomeres of chromosomes. It forms about 30% of the *Drosophila melanogaster* genome. Other properties of heterochromatin are reduced meiotic recombination, low gene density, repressing euchromatic genes, and enrichment with highly repetitive DNA sequences (reviewed by Dimitri *et al*, 2005; Grewal and Elgin, 2007). The DNA component of heterochromatin is not unique. Instead, sequences within heterochromatin are often repetitive and consist of transposons and satellite sequences (Bartolome *et al*, 2002; Hoskins *et al*, 2002). Instead of DNA sequence, histone status is the determining factor here. Histone modifications are described as the histone code.

Heterochromatin is necessary for chromosome stability and cellular viability. In *Drosophila melanogaster* that are homozygous for *Su(var)205* mutations, complete lack of HP1 causes lack of chromosomal condensation during mitosis in the embryo (Kellum and Alberts 1995). Lack of HP1 also causes telomere fusion (Fanti *et al*, 1998; Savitsky *et al*, 2002).

Although heterochromatin is generally considered a gene-poor part of the genome, there are some genes, such as *rolled* (*rl*) and *light* (*lt*), residing within heterochromatin. These genes need heterochromatin for proper expression and will be silenced if they are transferred to euchromatin (Eberl *et al*. 1993; Weiler and Wakimoto, 1998). Mutations in heterochromatin maintenance genes such as *Su(var)2-5* decrease expression of heterochromatic genes while increasing expression of variegated *P{lacW}* inserts in heterochromatin (Lu *et al*, 2000). Interestingly, homologues of some *Drosophila melanogaster* heterochromatic genes can be euchromatic in other organisms (reviewed by Dimitri *et al*, 2005).

Heterochromatic position effect variegation (hPEV) is a random suppression of a euchromatic gene when it is transferred near heterochromatin by chromosomal rearrangement. The prototype of hPEV is $In(1)w^{m4}$, an inversion in the X chromosome of *Drosophila melanogaster* that moved the *white* gene to near the centromere and causes variegated eye color in flies. hPEV is a tool to search for modifiers of variegation (and heterochromatin). *Su(var)* mutations (suppressor of variegation) suppress variegation and increase the gene expression, therefore wild type alleles of these genes are involved in heterochromatin maintenance. Mutations in enhancer of variegation genes (*E(var)*) increase the variegation and gene suppression. Thus wild type alleles play a role in euchromatin maintenance (Wallrath and Elgin, 1995). So far, more

than 500 dominant suppressor and enhancer mutations of *In(1)w^{m4}* have been found (reviewed by Ebert *et al*, 2006) but the exact number of *Su(var)* and *E(var)* loci is not known yet.

Histone codes

In actively-transcribed euchromatin, histones are in a different state from non-transcribed chromatin. Histone 3 has methylation at H3K27 (lysine at position 27 (K27) at histone 3 (H3)) and histone 4 has H4K20, in addition to of H3K9 and H3K14 acetylation. Whereas the hallmark of heterochromatin at chromosome arms and telomeres is H3K9 mono- or di-methylation and tri-methylation at centromere heterochromatin (reviewed by Ebert *et al*, 2006). H3K9 is de-acetylated by HDAC1, and is methylated by SU(VAR)3-9. Other histone codes for heterochromatin are tri-methylation of H4K20 by SUV4-20 and mono-, di- or tri- methylation of H3K27 by E(Z). After H3K9 methylation by SU(VAR)3-9, a SU(VAR)3-9/HP1/ SU(VAR)3-7 complex forms in the area and establishes the heterochromatin (reviewed by Ebert *et al*, 2006). SU(VAR)3-9 methylates lysine 9 on histone 3 (H3K9) and this facilitates HP1 binding to this histone. HP1 recruits more SU(VAR)3-9 protein to the site and this feedback loop maintains the heterochromatic status and silences euchromatic gene expression (Nakayama *et al*, 2001). In areas of chromatin that have controlling effects, such as replication sites and topoisomerase recognition sites, histone codes are different from both euchromatin and heterochromatin. H3K4 and H3K36 methylation combined with H3S10 phosphorylation are more prevalent in these areas (reviewed by Ebert *et al*, 2006).

SU(VAR)3-9 protein is a very strong promoter of heterochromatin and dominates almost all other factors. Increase in SU(VAR)3-9 activity results in H3K9 hypermethylation, enhanced gene silencing and extensive chromatin compaction. It mainly methylates mono-methyl H3K9 to the di-methyl and tri-methyl state. In the absence of HP1, SU(VAR)3-9 binding to chromatin is non-specific and affects euchromatin as well. HP1 attaches to di-methyl H3K9 and recruits SU(VAR)3-9 and SU(VAR)3-7 to the location. HP1 also recruits SUV4-20 that tri-methylates H4K20, a marker of repressive heterochromatic domains (reviewed by Ebert *et al*, 2006). There are some defensive mechanisms against heterochromatin spreading to euchromatin. For example, phosphorylation of H3S10 by JIL-1 and acetylation of H3K9 prevents H3K9 methylation by SU(VAR)3-9 (reviewed by Ebert *et al*, 2006).

There is also some evidence from *Saccharomyces pombe* for role of RNAi in targeting of heterochromatin machinery toward specific genes and silencing of them. But it is not enough to make the same conclusion for *Drosophila* (reviewed by Riddle and Elgin, 2008).

Chromatin in chromosome 4

Chromosome 4 of *Drosophila melanogaster* has 4.5-5.2Mb of DNA and about 3-4Mb of it is simple satellite repeats. About 1.2 Mb of it forms the band pattern 101-102 on polytene chromosome 4. This area consists of a banded pattern of heterochromatin and euchromatin, and its heterochromatinization is different from centromeric and telomeric heterochromatin (Sun *et al*, 2000). It has been reported that HP1 protein, which is encoded by *Su(var)2-5* gene, attaches to chromosome 4 arm in a banded pattern (James *et al*, 1989). This is supported by the finding of interspersed euchromatin and heterochromatin blocks in chromosome 4 (Sun *et al*, 2000). Expression levels of different chromosome 4 *P{hsp26-pt-T}*s are not the same. They follow a clustered pattern. Insertions in some blocks of chromatin induce a uniform red eye color while in other blocks the phenotype is variegated. In variegated mutants, the accessibility of *P* construct DNA to restriction enzymes was lower than non-variegated ones (Sun *et al*, 2000). Moreover, the response to heat shock was lower in variegated mutants. Finally, the variegated phenotype in all of these mutants is suppressed by $Su(var)2\text{-}5^{02}$ and *Su(var)3-7* mutations. These data indicate heterochromatin involvement in variegated mutants. Although all of these mutants show similar response to $Su(var)2\text{-}5^{02}$ and *Su(var)3-7* mutants and do not show any response to *E(var)3-93E*, their sensitivity to $Su(var)2\text{-}1^{01}$ and *E(var)3-93D* mutations is varied. This may indicate differences in protein complexes of chromosome 4 heterochromatin blocks, but generally all of them show a pattern resembling pericentric heterochromatin (Sun *et al*, 2000). SU(VAR)3-9 is not active on the distal arm of chromosome 4 of *Drosophila melanogaster* (Schotta *et al*, 2002; Haynes *et al*, 2007). Recent data suggest that dSETDB1, another histone methyl-transferase, is responsible for tri-methylation on the distal arm of chromosome 4 (Tzeng *et al*, 2007). However, HP1 binding in the centromeric area of chromosome 4 is dependent on SU(VAR)3-9 for its H3K9 methylation (Haynes *et al*, 2007; Tzeng *et al*, 2007).

Transposons and heterochromatin

Transposons are more prevalent in heterochromatin than in euchromatin (Bartolome *et al*, 2002; Hoskins *et al*, 2002). This prevalence may have two reasons and both of them are about selection for suppression of transposition. First, cells may heterochromatinize transposons to control transposition. Second, the negative selection for cells that have euchromatic transposons that are more active ends up with the prevalence of cells with heterochromatic transposons. In other words, transposition to heterochromatin will prevent transposons from further transposition and the cell would be more fit than the cells with euchromatic transposons.

PDS

PDS is a local position effect. In two different *P* constructs inserted upstream of the *ci* gene, *Pci* and *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, expression of the marker genes (*mini-white* and *white* gene respectively) are affected by the presence of *P* elements independent of the marker gene promoter (Bushey and Locke, 2004). This makes PDS distinct from previously described effects of *P* element repressors, which act only on the *P* element promoter (Lemaitre and Coen, 1991; Kaufman and Rio, 1991). Additionally, *P{lacW}* inserts at other genomic locations are not sensitive to these *P* elements, which also argue against promoter silencing. Bushey and Locke (2004) showed that mutations in *Su(var)205* and *Su(var)3-7*, two hPEV suppressors, suppress the PDS effect, while mutations in *Su(var)3-9*, another suppressor of hPEV, do not. Putting all of this together, it is very likely that PDS is also heterochromatin dependent as RIGS (Repeat-Induced Gene variegation) is (see below). PDS seems different from RIGS in that *Pci* is not variegated by itself and PDS induces variegation while RIGS is already variegated and the presence of the *P* element enhances the phenotype. It is possible that *Pci* is also under an undetectable influence of heterochromatin even in the absence of other *P* elements and PDS is uncovering this influence. The fact that an increase in *Su(var)3-7*$^{+}$ copy number silences *Pci* even in the absence of any *P* element supports this hypothesis (Bushey and Locke, 2004).

There are other supporting data for the presence of gene silencing chromatin upstream of the *ci* gene. First, regulation of the *ci* gene, an important part of appropriate anterior-posterior formation in the embryo and imaginal discs, shows an interesting pattern. Normally the *ci* gene is only expressed in anterior compartments. Amorphic mutations are recessive lethal but do not have any phenotype in heterozygous flies. Gain-of-function mutations that express the *ci* gene in posterior compartments as well as in anterior compartments induce an interruption in the L4 wing vein in homozygous flies. Despite their gain of function nature, these are recessive mutations. The presence of one normal copy of the *ci* gene is enough to suppress expression from these neomorphic mutations but when there is a chromosomal rearrangement that transfers the mutated gene to a new location, it acts as a dominant mutation (Locke and Tartof, 1994). The proposed model is based on transvection. One normal regulatory sequence is enough to suppress expression from both copies of the *ci* gene through a chromosome pairing mechanism. Therefore, the neomorph mutations behave as recessive mutations. However, chromosome translocations prevent chromosome pairing and the neomorph mutation expresses and shows a dominant pattern (Henikoff, 1997; Henikoff and Comai, 1998).

Second, a suppressing effect of distal regulatory sequences of the *ci* gene can affect the *RpS3a* gene. Normally *Pci*, which is inserted in a distal regulatory sequence of the *ci* gene, does not affect the phenotype of either the *ci* or the *RpS3a* genes. This is also true for $P\{hsp26\text{-}pt\text{-}T\}ci^{2\text{-}m1921.R}$. However, there is an imprecise excision of *Pci* that affects *RpS3a* expression and induces a Minute phenotype without any change in the *RpS3a* gene sequence. It is interesting that the presence of *P* elements suppresses this Minute phenotype back to normal (Bushey, 2004).

The third piece of evidence for the presence of heterochromatin at distal regulatory sequences of the *ci* gene comes from $P\{lacW\}ci^{E1}$ (*E1*) and $P\{lacW\}ci^{E2}$ (Figure1-4). These *Pci* alleles have a *gypsy* transposable element distal to *Pci* and their eye phenotypes are variegated. The presence of *P[SalI]89D* enhances this variegation (Bushey and Locke, 2004). The *gypsy* inserts in *E1* and *E2* have the ability to silence *Pci* both in *cis* and *trans*, *i.e.* both $P\{lacW\}ci^{E1}$ and $P\{lacW\}ci^{E2}$ are able to silence a *Pci* element on the homologue chromosome even after complete deletion of the *mini-white* gene from $P\{lacW\}ci^{E1}$ and $P\{lacW\}ci^{E2}$. The intensity of suppression is dependent on the level of homology between deletion derivatives of $P\{lacW\}ci^{E1}$ or $P\{lacW\}ci^{E2}$, and *Pci* on the homologue chromosome (Huang, 2005). This is similar to the suppressing effect of telomeric insert *NA-P(1A)* on euchromatic *P* elements (Marin *et al*, 2000) and reminds us about the *ci* transvection effect (Locke and Tartof, 1994).

Other modifiers of hPEV affect PDS as well. Extra Y chromosomes, a suppressor of hPEV, suppress PDS weakly, but do not have any effect on *Pci* alone. High rearing temperature also suppresses PDS, but at the same time increases expression of *Pci* even in the absence of PDS. This is in accordance with the temperature effect on hPEV and supports the idea that there is a heterochromatic effect at *Pci* (Bushey, 2004). However, direct recruitment of heterochromatinization proteins by P repressor protein is unlikely. If it was true, P repressor protein would be able to silence *P* constructs at other locations, but it doesn't. Moreover *P* elements do not affect hPEV (Bushey, 2004). Bushey (2004) has also tested RNAi involvement in PDS. He tested *piwi* and *Argonaute-1* against *Pci* and RIGS without finding any effect. This is in contrast with the RNAi effect on telomeric *P* elements (Reiss *et al*, 2004; Simmons *et al*, 2007). Lack of effect of *Argonaute-1* argues against the involvement of miRNA in PDS (Bushey, 2004; Okamura *et al*, 2004). This is in contrast with *Argonaute-1*silencing the effect of the mutant phenotype of $vg^{21\text{-}3}$, which is a *P* elements insertion at the *vg* locus (Anderson, 2008).

PDS and hPEV

The variegated eye color phenotype of PDS is similar to that of heterochromatic position effect variegation (hPEV) seen in $In(1)w^{m4}$ (reviewed by Reuter and Spierer, 1992). Mutations in

at least two genes that suppress hPEV, *Su(var)2-5* and *Su(var)3-7*, are also known to suppress PDS (Bushey and Locke, 2004). Suppressing effects of other modifiers of hPEV (extra Y chromosome, and high rearing temperature) on PDS plus the presence of heterochromatin tendency upstream of the *ci* gene (as mentioned before) support a role for heterochromatin in PDS (Bushey, 2004). Three other facts support this connection. First, *Drosophila melanogaster* chromosome 4 has interspersed euchromatic and heterochromatic domains (Sun *et al*, 2000), so sensitivity to *Su(var)* mutations is not unexpected. Second, since PDS is independent of promoter (Bushey, 2004) it is likely due to higher order chromatin structure and not promoter specific silencing of gene expression. Third, increasing *Su(var)3-7*$^{+}$ dosage, in the absence of *P* elements, causes variegation in expression of the *mini-white* gene at *Pci* (Bushey and Locke, 2004), suggesting that SU(VAR)3-7 protein is involved in determination of expression of *Pci*. However, PDS and hPEV are not identical processes because *Su(var)3-9* mutations do not affect PDS but do strongly suppress variegation of *In(1)w*^{m4} (Bushey and Locke, 2004). Again this is not unexpected considering the data rejecting a role for SU(VAR)3-9 in heterochromatinization of the distal arm of chromosome 4 (Schotta *et al*, 2002; Haynes *et al*, 2007). Loss of function mutations in *Su(var)* loci cause suppression of silencing (less variegation) implying their wild type function is to silence genes (make heterochromatin). So, the induction of variegation of *Pci* by *P* elements is phenotypically similar to effect of a wild type *Su(var)* allele on hPEV.

Other somatic *P* element dependent phenotypes

In addition to their effect on transposition, the presence or absence of *P* elements can induce phenotypic changes in somatic tissues of some *P* element induced mutants and *P* constructs. One of these phenotypes, PDS, is subject of this study. This is a brief review of some other somatic *P* element dependent phenotypes.

vg $^{21\text{-}3}$

vg $^{21\text{-}3}$ has a non-autonomous *P* element insertion and the expression of the *P* element leads to aberrant expression of *vg* and suppression of normal *vg* expression that causes a vestigial wing phenotype. The presence of other *P* elements suppresses the expression of the *P* insert in *vg* and reverses the phenotype to wild type (Williams *et al*, 1988, Hodgetts and O'keefe, 2001; Anderson, 2008).

sn$^{\mathbf{w}}$

sn^{w} has two non-autonomous *P* elements inserted in its upstream region that produce a weak singed bristle phenotype. In the presence of *P(ry*$^{+}$*, Δ2-3)99B* which can produce transposase in somatic tissue, one of the inserts is randomly excised causing a change in the

phenotype from weak singed to either wild type or a strongly singed phenotype giving a mosaic phenotype on the adult fly. Some *P* elements can suppress this transposase activity and the fly shows a uniformed singed phenotype (Robertson and Engels, 1989).

Transposition of *P{lacW}* by *P(ry$^+$, Δ2-3)99B*

In a *w*$^-$ background most *P{lacW}* inserts induce a uniformly red eye. *P(ry$^+$, Δ2-3)99B* can mobilize *P{lacW}* and produce a variegated eye color due to excision/insertion mosaism in the ommatidia. Suppression of this variegation can be used as a test for suppression of transposition by *P* elements (Misra and Rio, 1990).

Repeat-Induced Gene Silencing (RIGS)

RIGS involves tandem repeats of at least 3-4 *P{lacW}* inserts in a euchromatic area of the genome. The repeats produce a variegated eye color while a single or double insert in the same area produces only a uniform red eye color. Increase in the number of the repeats enhances this variegation. The RIGS phenotype is similar to hPEV and this similarity is not limited to appearance. RIGS is susceptible to dosage change in *Su(var)205* (Dorer and Henikoff, 1994). It has been shown that HP1, the protein that is encoded by *Su(var)205*, binds to *P{lacW}* tandem repeats (Fanti *et al*, 1998). Moreover, distance between tandem repeats and heterochromatin affects RIGS intensity (Dorer and Henikoff, 1997). However, hPEV and RIGS are different in their sensitivity to temperature, while rearing flies at 25°C suppresses hPEV it enhances RIGS (Josse *et al*, 2002). This temperature effect is similar to what has been reported before for M-P cytotype (Ronsseray *et al*, 1984).

Josse *et al* (2002) showed that the presence of *P[SalI]89D* or *KP(D)* (that has four *KP* elements (Bushey, 2004)) enhances the RIGS phenotype without showing any maternal effect. However, elements that do not suppress transposition such as *P[XhoI]* and second chromosome of *Birmingham 2* (that has approximately 17 defective *P* elements), do not affect RIGS either (Engels *et al*, 1987; Robertson and Engels, 1989; Josse *et al*, 2002). Interestingly, single *P{lacW}* inserts in centromeric heterochromatin of chromosome 2 and 3 that have a variegated phenotype are not susceptible to the presence of the *P[SalI]89D* (Josse *et al*, 2002).

Telomeric effect

A *Drosophila melanogaster* telomere has two different types of chromatin: a euchromatic terminal retrotransposon area and a heterochromatic subterminal telomere-associated sequence (TAS) (Biessmann *et al*, 2005). Telomeric *P[lacZ]* inserts on the X chromosome tip, region 1A (*WG-1103* and *WG-1152)* are able to suppress expression of *P[lacZ]* inserts within euchromatin. This effect needs some sequence homology between these two inserts (Ronsseray *et al*, 1998;

Ronsseray *et al*, 2003). Although these elements can not affect hybrid dysgenesis by themselves, they can suppress it through a combination effect, which in turn can be suppressed by *Su(var)205*. This combination effect happens when the embryo receives a telomeric insert from the mother and other repressor *P* elements from the father. If the embryo receives both elements from the father, this effect is absent. Centromeric *P[lacZ]* inserts do not show a combination effect (Ronsseray *et al*, 1998).

Complete *P* elements that are inserted into heterochromatic telomeric associated sequences (TAS) at the X chromosome telomere (site 1A) can suppress hybrid dysgenesis if they are transmitted from mother to embryo (Ronsseray *et al*, 2003). Marin *et al* (2000) reported another telomeric *P* element in 1A, *NA-P(1A)*, that has an 871 bp deletion at its 5' end including the *P* promoter. It can not encode P repressor protein but it is able to strongly suppress dysgenic sterility in crosses with P strain Harwich-2 males and *P* transposition induced by $P(ry^{+}, \Delta 2\text{-}3)99B$. Its suppressing effect on other *P* inserts is related to the degree of similarity in DNA sequence between them. While it does not suppress *P[lacZ]* (233 bp in common), it suppresses *PLH3 P[lacZ]* that has 1.8 kb in common with *P(1A)*. *Su(var)205* mutations strongly suppress *NA-P(1A)* effects while higher temperature enhances them (Marin *et al*, 2000). This temperature effect is similar to what has been reported before for M-P cytotype (Ronsseray *et al*, 1984).

If present in the female, defective telomeric *P* inserts, TP5 and TP6, can suppress hybrid dysgenesis but not transposition in somatic cells (Niemi *et al*, 2004; Simmons *et al*, 2007). This effect in embryos is independent of the presence of the telomeric elements but it is not completely penetrant (Simmons *et al*, 2007). The suppressing effect of TP5 and TP6 can be enhanced by the presence of defective *P* elements of an M' cytotype. Both the telomeric element and the M' elements have to be present in females for the repression enhancement to occur, but repression can transfer to embryos independently of telomeric inserts themselves. Since TP5 and TP6 do not encode any protein, their effect may happen through RNA intermediates (Simmons *et al*, 2007). Mutations in *aubergine* but not *piwi* and *homeless*, which are involved in RNAi pathway, also suppress telomeric effects but not the regular P strains (Reiss *et al*, 2004; Simmons *et al*, 2007). Moreover, *Su(var)205* mutations can suppress this effect while the regular P strains are insensitive to *Su(var)205* (Haley *et al*, 2005).

Overview of the *KP* experiment

Considering similarities between PDS and hPEV, it is more likely that they are both using a similar mechanism. Since many *Su(var)* genes modify hPEV in a dose-dependent manner, I tested a dose-dependent effect of *P* elements and its silencing effect on a *mini-white* transgene

during PDS. *Drosophila melanogaster* stocks containing just one *KP* element did not induce PDS, while a strain with at least four *KP* inserts could induce PDS (Bushey, 2004). This difference in ability to enhance PDS could be due position of insertion of the *KP* elements as it has been already shown that *P* element effects on P cytotype vary depending on its position of insertion. However, it also could be due to a quantitative difference in the number of *KP* inserts, or some combination of both. Although Simmons *et al* (2002) did not find any additive effect between *H(hsp/CP)* and *H(hsp/KP)* elements in their suppression effect on transposition *in vivo*, Lee *et al* (1998) showed that the KP protein suppresses transposition *in vitro* in a dose dependent manner. Further, Jackson *et al* (1988) showed that the intensity of P cytotype is directly related to the number of *KP* inserts. My hypothesis was that the *KP* effect on PDS will be dose-dependent as well. I show that one *KP* element is sufficient to induce PDS and that by combining multiple *KP* elements more silencing can be obtained thereby showing PDS is subject to a dose-dependent effect of the *KP* element.

Overview of the *P[SalI]89D* experiment

Loss of function mutations in *Su(var)* loci cause suppression of silencing (less variegation) in hPEV implying their wild type function is to silence genes. So, the induction of *Pci* variegation by *P* elements is phenotypically similar to the *Su(var)* effect on hPEV. It is interesting that a foreign protein induces the same phenotype and probably uses a similar mechanism as SU(VAR) proteins. Here, I investigated the required domains for a type I repressor protein to induce PDS. Bushey and Locke (2004) showed that the *P[SalI]89D* insert on chromosome 3, a classic type I repressor, is able to silence *mini-white* gene expression in *Pci* and *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$. It also suppresses variegation of expression of the *mini-white* gene in *P{hsp26-pt-T}39C-12*. The purpose of this study is to define the important domains in the type I repressor protein for its effect on PDS and some other *P* inserts on chromosome 4. My hypothesis was that most of the domains should be in a region that is common for both *P[SalI]* and *KP* elements, since both are able to induce PDS.

First, I found that *P[SalI] 89D* is actually inserted in the 5'-UTR of the *Pak3* gene, which is in region 89B. To prevent further confusion, I simply call this insert *P-Sal*. I mutagenized *P-Sal* and looked for mutations that severely reduce (hypomorph) or eliminate (amorph) the silencing of *Pci*. I called these mutants *P-Sal** collectively. I tested these mutants with *In(1)w*^{m4} to find any *Su(var)* effect and with parental *P-Sal* for any antimorphic effect. I also tested them for their effect on *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, *P{hsp26-pt-T}39C-12* and *E1*. Then, I compared *P-Sal** mutants with the available phylogenic data and tried to predict changes in secondary protein

structure using protein prediction programs. Finally, I tried to create an antibody against the type I repressor protein to test for the presence of this protein in *P-Sal** mutants. I injected rabbits with the denatured type I repressor protein, which was produced in a prokaryotic system, to induce production of a polyclonal antibody against this protein. Although the final rabbit sera can detect the produced repressor protein strongly, they do not have enough specificity to be used on a fly protein extract. Affinity purification of the rabbit sera could increase their specificity for the P repressor protein.

Chapter 2: Materials and Methods

D. melanogaster stocks

Drosophila stocks were grown at a constant temperature of 21°C on cornmeal medium in plastic vials and bottles. The *KP* element in this study was derived from a KP(D) stock (Rasmusson *et al*, 1993). The original stock, received from Dr. S. Ronsseray, was described as +, *Cy Bl vg/ NS;* +; +. The source of *P[SalI]89D* (*P-Sal*) in this study was *y w;* +; *Sb P-Sal/ TM6B Tb Hu;* + stock (Karess & Rubin, 1984). The source of transposase was $P(ry^{+}$, *Δ2-3)99B* (hereafter *Δ2-3*) (Robertson et al, 1988) in *w;* +; *Sb e Δ2-3/TM6 Ubx e;* +. Balancer stocks were *y w; CyO/* ap^{Xa}; +; + and *y w;* +; *TM6B Tb Hu /* ap^{Xa}; +.

Cytology of polytene chromosomes

To determine *KP(D)* chromosome (*Cy Bl vg* chromosome of KP(D) stock) structure, salivary glands of third instar larvae were dissected, then fixed in 45% acetic acid. Glands were squashed in orcein /acetic acid/ lactic acid to stain polytene chromosomes (Ashburner, 1989). Chromosomes were examined and photographed using a Zeiss Axiophot microscope with a 3.2 mega-pixel Nikon Coolpix 995 digital camera.

EMS mutagenesis

To mutagenize *P[SalI]89D*, I used EMS (ethylmethane sulfonate) following the EMS mutagenesis protocol of Ashburner (1989). Based on this protocol, I used a total of more than 4000 *y w;* +; *Sb P-Sal / TM6B Tb Hu;* + males. After 2-4 hours of starvation in empty bottles, batches of 50-200 starved males were exposed overnight to Whatman paper soaked in EMS solution of 1.6 µl EMS (Sigma) in 600 µl of 1% sucrose (w/v).

Genomic DNA Extraction

For PCR reactions, I extracted the DNA from a single fly of each mutant strain using the protocol of Gloor and Engels (1992). For Southern transfer and vectorette PCR, genomic DNA was extracted from batches of 80-100 flies by crushing frozen flies in homogenization buffer (200mM EDTA pH8.0, 100mM Tris pH7.5, 1% SDS) and incubation for 30 minutes at 60°C, followed by a standard phenol/ chloroform DNA extraction protocol (Sambrook, 1989).

PCR

For the selected *KP* mutants, I confirmed the presence of *KP* elements using primer PRPT to PCR-amplify the whole element in genomic DNA extracted from a single fly. PRPT (Table 2-1) is complementary to part of the 31 bp inverted repeats at both ends of transposable

P elements. PCR with PRPT primer produces a 2.9 kb product from the complete *P* element, a 1.1 kb product from *KP* elements, and none from the *Δ2-3*. To confirm the absence of *Δ2-3*, and any complete *P* element, in DNA extracted from single *KP* mutant flies, I used PCR amplification with 2033F and 3195R primers (Table 2-1). Together, these two primers amplify a 0.95 kb PCR product from *Δ2-3* element, and a 1.1 kb PCR product from a complete *P* element, but no product from *KP* elements. For each *P-Sal** mutant (*y w; +; Sb P-Sal*/ TM6B Tb Hu; +*), DNA from a single male was amplified to confirm the presence of *P-Sal*. This PCR reaction used primers 2033F and 3195R (Table 2-1) that produces a 1.1 kb PCR product from *P-Sal*.

Southern Transfers

Genomic DNA for each *KP* mutant was digested with *Bgl*II or *Eco*RV, separated on a 0.7% agarose gel, transferred to Gene Screen Plus nylon membrane (NEN Life Science Products) and probed according to the manufacturer's instructions. Since *Bgl*II and *Eco*RV restriction sites are not present in *KP*, each insert should correspond to one hybridized band on a Southern transfer. The DNA probe was derived from a PCR amplified *P* element sequence from plasmid *pP{π25.1}* using PRPT primers (Table 2-1). The gel purified DNA fragment (Qiaquick gel extraction kit - QIAGEN) was used as a template for random hexamer priming to make DNA probes by ^{32}P-dCTP incorporation (J. Sambrook, 1989).

Vectorette PCR

The vectorette PCR method (Riley et al, 1990) was used to find the genomic location of the selected *KP* elements and the *P-Sal* inserts. This method permits the amplification of a restriction fragment that contains the *P* element end sequence. In summary, in this method the genomic DNA is digested with a restriction enzyme. The digestion products are ligated to a vectorette, a short unique DNA segment with known sequence. The ligation products are used for a PCR reaction with a specific primer for the vectorette and another primer that is specific for the gene sequence under investigation, in this case the *P* element end (Table 2-1). The amplified PCR product is sequenced and genomic sequence adjacent to the vectorette and to the *P* element is identified and used to search a genomic data base (BLAST) to find the location of insertion of the *P* element.

Table 2-1) PCR primers.

Primer Name	Sequence	Hybridize to *KP*	Hybridize to *P*	Application
11rpt5	CGTAAGGGTTAATGTTTTCAAA	138-117	138-117	Vectorette PCR
11rpt3	TAAGGGTTAATCAACAATCATAT	1010-1032	2763-2785	Vectorette PCR
1242F	CCTGCAGATGACCATTTAAAG	---------	1242-1262	Sequencing of *P* element
2033F	CCGAATGGACCCGGATACTC	---------	1038-1057	Detection of complete *P* element and *Δ2-3* element
2116R	CTGGATAAGCAAAAGGATAATG	---------	2137-2116	Sequencing of *P* element
2376R	TGAAGTGCCTCCTGAATTGT	---------	1381-1362	Sequencing of *P* element
2632F	GCTCGCAACCTTATGGCAAG	---------	1626-1655	Sequencing of *P* element
2712B	TTGAAATGGGAGCCTTTTGGG	---------	1745-1725	Sequencing of *P* element
2888F	TTCGACCATCCCACTCCACT	---------	1893-1912	Sequencing of *P* element
3195R	GCCAGCTTTCAGAGTTGTCC	---------	2190-2181	Detection of complete *P* element and *Δ2-3* element
3269F	CGATGAGATGTTAAGCAATAT	---------	2274-2294	Sequencing of *P* element
330R	CCTCCTTTTAAATGTCTGACC	350-330	350-330	Sequencing of *P* element
3413B	TAAGTCCGCCGTGAGACACC	---------	2137-2418	Sequencing of *P* element
766R	ACAAAGTCGTACGACTGGGCAA	787-766	787-766	Sequencing of *P* element
DPE02	CACACGTCTTTTCCTCTCAACAAGC AAACG	102-73	102-73	Nested vectorette PCR
KP-F1	AAGGCTATACCAGTGGGAG	412-431	412-431	Sequencing of *P* element
KP-R1	GAAGATTTTGCGTAGAGACTC	610-630	610-630	Sequencing of *KP* element
KP-R2	ATCCGTATCTGCGTGTCCG	650-632	650-632	Detection of *P* cDNA
PLW1	TCCACTTAACGTATGCTTGC	1113-1094	2866-2847	Sequencing of *KP* element
PLW2	AGCTTACCGAAGTATACACT	40-59	40-59	Sequencing of *P* element
PRPT	TAACATAAGGTGGTCCCGTCG	11-31	11-31	Detection of complete *P* element or *KP* element
		1144-1124	2897-2877	
SK08	TGCAAGCATACGTTAAGT	1093-1110	2846-2863	Nested vectorette PCR
JS1	ATACATATGTTTCATTACGG	None		Amplify chromosome 4 in the absence of *Pci*
JS2	GCGTTTGTATGTATATTG			
Bub-c	CTCTCCCTTC TCGAATCGTA ACCGTTCGTA CGAGAATCGC TGTCCTCTCC TTC			Vectorette PCR Cassette
	GAGAGGGAAG AGAGCAGGCA AGGAATGGAA GCTGTCTGTC GCAGGAGAGG AAG			
Bub-pp	CGAATCGTAACCGTTCGTACGAGAATCGCT			Vectorette PCR Cassette primer

For *KP* mutants, genomic DNA was digested with *Eco*RV, *Hpa*I, *Bsa*AI, or *Ssp*I. I used a cassette, Bub-c, and its primer, Bub-pp, (Munroe, 1996) for amplification (Table 2-1). To amplify the 5' end of *KP*, I used the Bub-pp and 11rpt5 primers (Table 2-1). I separated the amplified PCR product(s) by electrophoresis on a 0.7% agarose gel and extracted the band from the gel using Qiaquick gel extraction kit (Qiagen). The PCR products were re-amplified by nested PCR using primers Bub-pp and DPE02. The PCR products served as a template for sequencing by BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). To prime the sequencing reaction, I used primer DPE02. To amplify the 3' end of *KP* element, I followed an equivalent protocol except the first round of PCR amplification used 11rpt3 instead of 11rpt5 and the nested PCR and sequencing reactions used primer SK08 (Table 2-1). To find the location of insertion of *P-Sal* I used *Ssp*I enzyme for digestion and I looked for the 5' end of this *P* element as mentioned above. The genomic sequences that were adjacent to *P* element sequences were used to search against the *Drosophila melanogaster* genomic sequence database at National Center for Biotechnology Information (NCBI) using the Basic Local Alignment Search Tool (BLAST) and determine the exact location of the *KP* inserts and *P-Sal*.

Sequencing

For each *KP* mutant, the whole *KP* element was PCR amplified using PRPT primer. The antisense strand of gel purified PCR product (Qiaquick gel extraction kit from Qiagen) sequenced by BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using PLW1 and KPR1 to prime the reaction (Table 2-1).

To find mutations in each *P-Sal** mutant, different parts of the element were amplified using different sets of primer pairs in regular PCR reactions (Table 2-2). PCR products were sequenced as mentioned above using appropriate primers (Table 2-2). In general, every mutant was sequenced for its antisense strand of *P-Sal* from base pair 100, about 50 bp upstream of the *P* element start codon at base pair 153, through base pair 2100 that is about 100 bp downstream of the *P-Sal* stop codon at base pair 1994. More details are provided in Table 4-8. The results were compared with the published *P* element sequence (O'Hare and Rubin, 1983) by Genetools software (Biotools). If a mutation was found, it was confirmed by sequencing of the sense strand.

Table 2-2) *P* element sequencing primers.

Target on *P* element	PCR Primer Pair	Primer for Sequencing Antisense Strand	Primer for Sequencing Sense Strand
30-310	PRPT / 330R	330R	PRPT
120-740	PLW2 / 766R	766R	PLW2 KP-F1
670-1330	KPF1 / 2376R	2376R	KP-F1 2033F
1250-1710	1242F / 2116R	2712B	1242F
1420-2070		2116R	2632F (1670-2070)
1910-2390	2888F / PLW1	3413B	2888F
2170-2830		PLW1	3269F (2300-2830)

Eye pigment measurement

The amount of eye pigment in various groups of progeny of *KP* additive crosses was measured using the acidified-methanol protocol described by Ephrussi and Herold (1944). Pigments were extracted from each sample, which contained heads of 15 five to nine day old flies in 200 µl of acidified-methanol. The A_{480} was measured with a spectrophotometer (Genova Life Science Analyzer). Three samples were prepared and measured for each group of progeny from each cross. To document pigment intensity and distribution in eyes of adult flies, images of fly eyes that were aged for 5-9 days were photographed with a 3.2 mega-pixel Nikon Coolpix 995 digital camera mounted on a Zeiss DRC stereomicroscope.

Protein prediction programs

I used the online protein structure prediction programs PROSITE motif search (Gattiker *et al*, 2002), NetPhos 2.0 (Blom *et al*, 1999), DomPred (Marsden, 2002), Porter protein predict (Pollastri and McLysaght, 2005), PSIPRED (Jones, 1999), SABLE-2 (Adamczak *et al*, 2005), and SIFT (Ng and Henikoff, 2001) to find the possible change in protein domains and secondary structure of the *P-Sal** mutant polypeptides.

Statistical Analysis

I compared the mean of eye pigment measurements of three samples for each progeny group of each *KP* additive cross to other progeny groups of the same cross by One Way ANOVA test with Bonferroni post hoc analysis on SPSS12.0 software (SPSS Inc). Bonferroni post hoc was used as a method to compare results of each progeny group with other groups of

that progeny *inter se*. To compare the frequency of different phenotypes among progeny of crosses between the *P-Sal** mutants and *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, *P{hsp26-pt-T}39C-12*, or *E1*, I used cross-tabulation descriptive analysis. To do this I used *Chi square* (X^2) tests, and when expected values were less than five, I used one-sided Fisher's exact test. To statistically analyze the results from crosses between the *P-Sal** mutants and parental *P-Sal* or *In(1)w*^{m4}, and crosses between *KP* mutants and *In(1)w*^{m4}, I used cross-tabulation X^2 again. If the expected value was less than five I used two-sided Fisher's exact test. To calculate these tests I used SPSS 12.0 computer software (SPSS Inc).

Chapter 3: *P* element induced silencing (PDS) that is induced by *KP* inserts is partially additive

Introduction

It is known that the expression of the *mini-white* gene at *Pci* is sensitive to *Su(var)3-7*$^{+}$ dosage, even in the absence of *P* elements (Bushey and Locke, 2004). I tested the dose-dependency of PDS on the number of the *P* elements that are present in the genome.

The original *Drosophila melanogaster* stocks containing just one *KP* element did not induce PDS, while the KP(D) stock with at least four *KP* inserts on a *Cy Bl vg* marked chromosome 2 (*KP(D)* chromosome hereafter) could induce PDS (Bushey, 2004). This difference in PDS enhancement could be explained by three models. First, the insertion location of a *KP* element could dictate the effect: a "golden location" model. It has been already shown that the extent of *P* element effects on P cytotype depends on its position of insertion. Second, it could also be due to a quantitative difference in the number of *KP* inserts: a "numbers" model. Third, it could also be a combination of both factors. PDS is not an all or none phenomenon. Thus, the extent of PDS should be able to vary by both a higher level of expressions from just one *KP* insert and/or by adding expression from several *KP* inserts (an additive effect).

In vivo, a *KP* construct under the control of a heat shock promoter, *H(hsp/KP),* did not add to the level of suppression of transposition that was induced by a complete *P* element construct with similar promoter, *H(hsp/CP)* (Simmons *et al*, 2002). However, *in vitro* the KP protein suppresses transposition in a dose dependent manner (Lee *et al*, 1998). Further, the intensity of P cytotype is directly related to the number of *KP* inserts (Jackson *et al*, 1988). Based on these data I predicted that the *KP* effect on PDS will be dose dependent as well. I tried to test this by separating different *KP* inserts on the *KP(D)* chromosome and studying each of them individually. Due to multiple inversions of *KP(D)* chromosome it was impossible to separate the *KP* inserts by recombination. However, I was able to transpose *KP* elements to new genomic locations and study single *KP* inserts. I showed that one *KP* insert is sufficient to induce PDS. This is in contrast with previous data and can be explained by different locations of insertion. Further, I showed that by combining multiple *KP* elements more silencing can be obtained, thereby showing PDS is subject to a dose-dependent effect of the *KP* element product.

Results

KP elements in KP(D) strain

This investigation began with the KP(D) stock (*+; Cy Bl vg/ NS ; +; +*) that was able to induce the PDS phenotype and the phenotype segregated with the Cy and Bl phenotypes (Daniel Bushey, 2004). Genomic Southern transfer analysis of *Eco*RV and *Bgl*II digested DNA showed that the *Cy Bl vg* marked chromosome had at least four *KP* inserts. Vectorette PCR followed by DNA sequence analysis showed that these *KP* inserts are in cytological regions 26A, 37C (two inserts) and 54B of chromosome 2 (L. Podemski, personal communication; Bushey, 2004).

To identify which element(s) were linked to the PDS phenotype I undertook a series of crosses to map the gene silencing factor on chromosome 2. To do this I crossed *+; Cy Bl vg/ NS ; +; +* males to *w; +; +; +* virgin females and selected Cy Bl males from progeny (*w; Cy Bl vg/ + ; +; +*). These males were crossed to *w; dp; e; +* virgin females. From progeny of this cross, virgin Cy Bl females (*w; Cy Bl vg/ dp ; +/ e; +*) were selected and crossed to *w; dp; e; Pci* males. The progeny were scored for any segregation between Cy (cytologic location 2L-6), dp (cytologic location 2L-13), Bl (cytologic location 2L-55), and the PDS phenotype. This genetic mapping of the PDS phenotype to the marker loci (*Cy*, *Bl*, and *dp*) failed to recover any crossing over along the left arm of *Cy Bl vg* chromosome in over 400 progeny. Many recombinants would be expected, given the locations of these marker loci. Although one of the *KP* inserts was located at right arm of chromosome 2, this experiment could not segregate PDS phenotype from left arm related marker loci (Cy, Bl, or vg) either.

To further examine the lack of recombination on this chromosome, I examined the chromosome cytologically. I crossed *+; Cy Bl vg/ NS ; +; +* males to *w; dp; e; Pci* virgin females. Cy Bl males from progeny (*w; Cy Bl vg/ dp ; +/ e; +/ Pci*) were selected and crossed to *w; dp; e; Pci* males. I examined the salivary gland polytene chromosomes from third instar larvae of this cross. A cytological examination of the *Cy Bl vg* chromosome (Figure 3-1) revealed three inversions. On the left arm of chromosome 2 there were two inversions (*In(2L) 22B-33E* and *In 35A-36C*), while on the right arm of this chromosome there was one inversion at *42A-58F*, which was similar to *In(2R)Cy* (*In 42A-58A*) (Ashburner and Lemeunier, 1976). While the presence of multiple inversions precluded further mapping study of the multiple *KP* inserts on *KP(D)* chromosome, it provided an excellent source of *KP* for producing single *KP* inserts transposed to non-rearranged chromosomes.

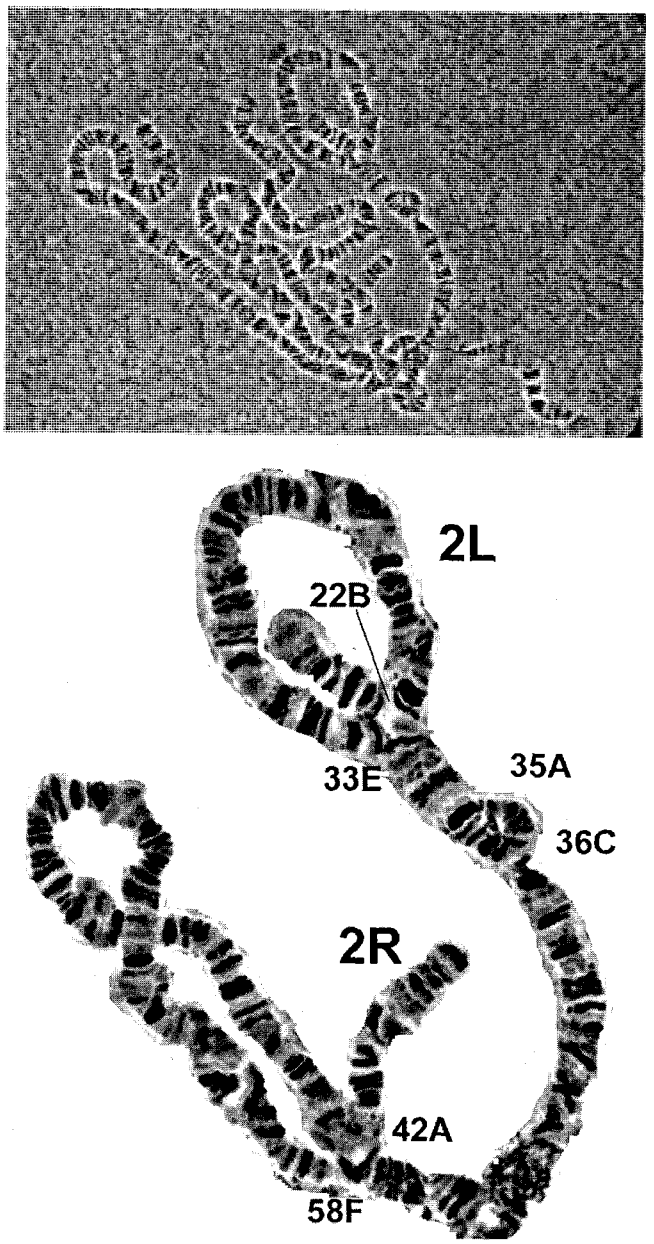


Figure 3-1) Polytene chromosomes from the Cy Bl vg stock.

Salivary glands of *+; Cy Bl vg/ +; +;* + third instar larvae were dissected, fixed and squashed in orcein /acetic acid/ lactic acid. The top panel is an overview of a set of chromosomes. The bottom panel shows multiple aberrations on chromosome 2. In 2L there are two inversions. One involves the region between cytological region 22B to 33E, while the other involves the region between 35A to 36C. On 2R there is one inversion involving the region between 42A and 58F.

Production of new *KP* insertion mutants causing PDS

To transpose *KP* elements from the *Cy Bl vg* chromosome, I crossed males from KP(D) (*+; Cy Bl vg/ NS; +; +*) to *w; +; Sb e Δ2-3/ TM6 Ubx; +* virgin females. Cy Sb progeny (*w; Cy Bl vg/ +; Sb e Δ2-3/ +; +)* were crossed to *w; dp; e; Pci* virgin females and progeny of this cross were screened for w^{var} Cy^{+} Sb^{+} flies; i.e. a PDS phenotype (variegated eyes) in *w; dp/ +; e/ +; Pci/ +* flies. The dominantly marked chromosomes, Cy and Sb, were selected against to remove the parental *KP(D)* and the *Δ2-3* chromosomes. Any recurrence of the PDS phenotype was attributed to a new mutation, such as a *KP* insert on a previously wild type chromosome. From 1938 Cy^{+} Sb^{+} (*w; dp/ +; e/ +; Pci/ +*) progeny 870 (45%) showed some PDS phenotype.

To show that only one *KP* insert is sufficient to induce PDS, I continued the experiment as follows. I selected 15 males that showed moderate PDS phenotype (30-70% of their eye ommatidia showed pigment). To avoid multiple *KP* inserts mutants, I selected against strongly variegated eye phenotypes. Theoretically there were 15 mutated second chromosomes and 15 mutated third chromosomes in the 15 selected mutant males. I backcrossed selected males to *w; dp; e; Pci* virgin females to segregate mutations on chromosome 2 from mutations on chromosome 3. Progeny showing PDS that were dp^{+} e (*w; dp/ KP(2); e; Pci/ +*) were considered mutated on chromosome 2 [*KP(2)*], and this chromosome 2 was recovered and maintained as a strain balanced with *CyO* (Lindsley and Zimm, 1992). Whereas, dp e^{+} (*w; dp; e/ KP(3); Pci/ +*) males showing PDS phenotype were considered mutated on chromosome 3 [*KP(3)*] and this chromosome 3 was recovered and balanced over *TM6B Tb Hu* (Craymer, 1984). Final stocked mutants were re-tested by crossing them to *w; dp; e; Pci* virgin females to confirm that the PDS phenotype segregated from the appropriate balancer chromosome. Although I selected against strong PDS phenotype to avoid multiple *KP* mutants, from the 15 selected mutants I recovered 25 mutant chromosomes (1 chromosome X, 12 chromosome 2, and 12 chromosome 3) that were able to induce a weak to moderate PDS with complete penetrance. As Southern transfer results showed later, most of these chromosomes had more than one *KP* insert. In other words, most of the 15 selected mutants had multiple *KP* inserts.

Initial PCR examination of 12 chromosome 2 mutant (KP(2)) lines and 12 chromosome 3 mutant (KP(3)) lines showed a 1.1 kb PCR product from PRPT primers, thereby confirming the presence of *KP* elements (data for selected mutants is shown in Figure 3-2). The absence of any product in 2033F/3195R PCR confirmed absence of other *P* elements (Figure 3-2). Additional Southern Transfer probing of restriction digested genomic DNA of each line showed that of 24

mutants, five KP(2) mutants and two KP(3) mutants had only one band for both *Bgl*II and *Eco*RV digested DNA (Figure 3-3 & Table 3-1). I considered these mutants as single *KP* mutants and selected two from each chromosome for further work.

The presence of lines with a single *KP* element shows that one *KP* element is sufficient to induce PDS. However, they are different in their phenotype. This difference can be due to a difference in the location of the insertion [difference in the location of the insertion is studied by vectorette PCR (see below)]. This leads me to the third model that states the original *KP(D)* chromosome PDS strength is due to combination of both location and the number of *KP* inserts. The next question asks how different *KP* inserts affect each other's PDS effects. If PDS strength is dependent on *KP* dosage, *KP* elements can be antagonistic, partially additive, additive, or synergistic. The following experiment addresses this question. The *KP* element in all four lines was sequenced and found to be the same as the consensus *KP* sequence (Black *et al*, 1987). This *KP* consensus sequence differs from the consensus *P* element sequence in an A to T transversion at base pair 32, a previously known variation in *P* element sequence (O'Hare and Rubin, 1983). I used vectorette PCR (Riley et al, 1990) to determine the insertion location of these four novel *KP* insert strains. All of these mutants were inserted in the 5'-UTR of genes (Table 3-1). None of them are known to, or expected to, affect hPEV. I have labeled these *KP* inserts after the gene they inserted in, *KP(2)-lola, KP(2)-beta, KP(3)-eff*, and *KP(3)-CG9007*. Stocks that contain these inserts were called after them KP(2)-lola, KP(2)-beta, KP(3)-eff, and KP(3)-CG9007, respectively.

To use these stocks for the dose-dependent experiment, I had to make them homozygous for *Pci*. In each *KP(2)* mutant, multiple individual pairs of w^{var} Cy males and virgin females (*y w; KP(2) / CyO; +; ~ Pci*) were pair mated. From each pair that failed to show w or e phenotype in the progeny, I mated a single male and virgin female pair again and after one week of laying eggs, I gathered them to be tested for the presence of wild type chromosome 4 by JS1/JS2 PCR (Table 2-1). JS1 and JS2 hybridize to genomic DNA sequence just outside of *Pci* location of insertion. The expectation is that during the PCR reaction, they amplify a 627 bp product from a wild type chromosome 4, while no product from a *Pci* chromosome. Lack of the product in a PCR reaction of DNA extracted from a single fly means that the fly was homozygous for *Pci* (data not shown). I followed a similar cross scheme for KP(3) mutants (*y w; +; KP(3)/ TM6B Tb Hu, ~Pci*) but selected for lack of w and dp phenotypes in the progeny. From this experiment I was able to produce (*y w; KP(2) / CyO; +; Pci*) and (*y w; +; KP(3)/ TM6B Tb Hu, Pci*) stocks that I used for my dose-dependent experiment (Figure 3-4).

Table 3-1) Single *KP* inserts.

Mutant Name	Genomic Location (vectorette PCR)	Interrupted gene (*KP* orientation regarding direction of transcription of the interrupted gene)	Band size in Southern transfer (kb)		Homozygote viability
			*Bgl*II digestion	*Eco*RV digestion	
KP(2)-lola	2R: 47A11	5'-UTR of *lola* (parallel)	4	14	lethal
KP(2)-beta	2R: 56D7	5'-UTR of *beta Tub 56D* (opposite)	8	5	lethal
KP(3)-eff	3R: 88D2	5'-UTR of *eff* (opposite)	7	16	lethal
KP(3)-CG9007	3L: 70C11	5'-UTR of *CG9007* (opposite)	16	11	viable

Foot note: In all of the mutants, *KP* element is inserted in the 5'-UTR of a gene. None of these genes are known to affect hPEV.

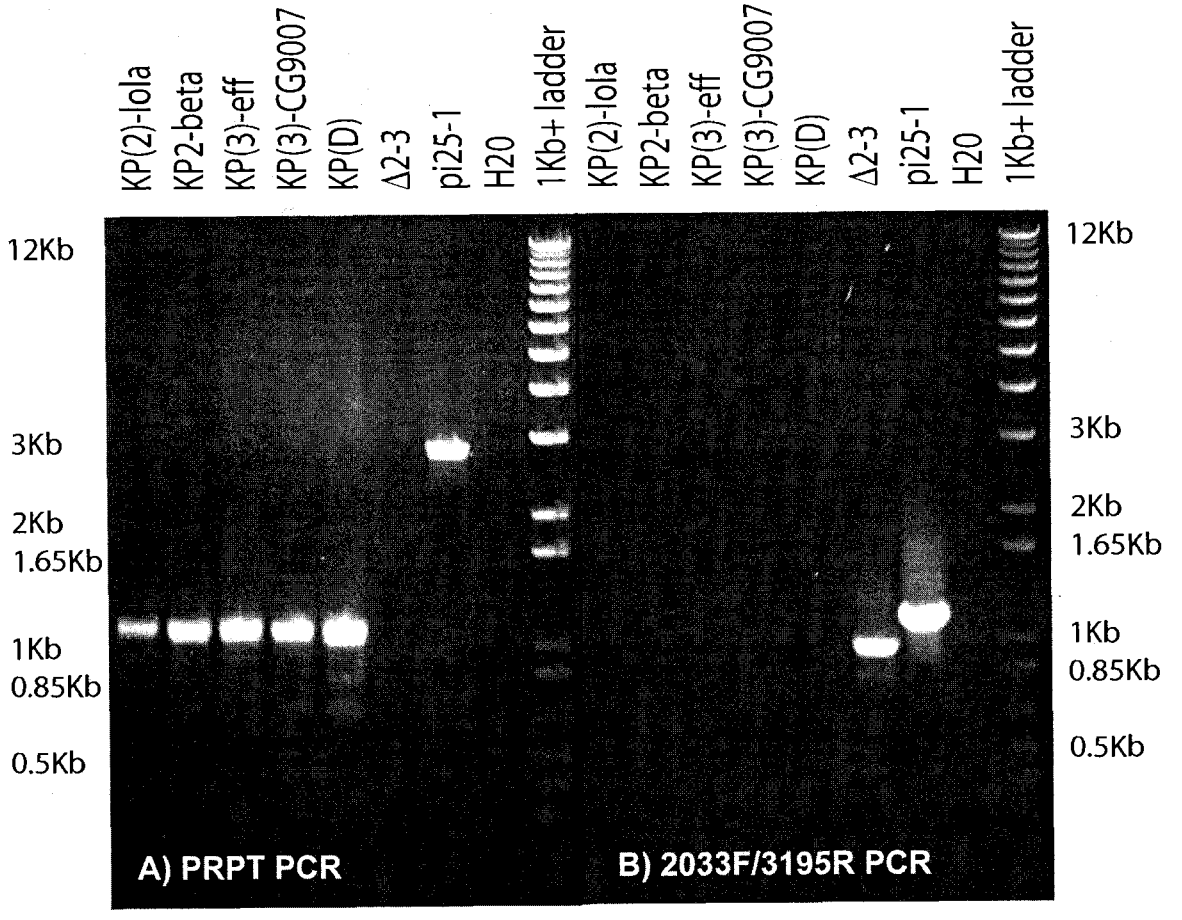


Figure 3-2) Results of PCR amplifications of selected single KP mutants.

Panel A represents results of PRPT PCR reaction that would amplify a 1.1 kb product from a *KP* and a 2.9 kb from a complete *P* element, but no product from *Δ2-3*. Panel B represents results of a 2033F/3195R PCR reaction that would amplify a 1.1 kb product from a complete *P* element and a 0.95 kb product from *Δ2-3*. The 2033F/3195R PCR reaction would not amplify any product from a *KP* element.

KP(2)-lola, KP(2)-beta, KP(3)-eff, and KP(3)-CG9007 are single *KP* lines. Since all of the tested mutants showed the 1.1 kb PCR product from PRPT PCR reaction and no product from 2033F/3195R PCR reaction, I concluded these mutants have only a *KP* elements and no other *P* element.

KP(D) stock is the source of *KP* element in this study (+, *Cy Bl vg/ NS;* +; +), Δ2-3 stock is the source of *Δ2-3* that was used in this study (*w;* +; *Sb e Δ2-3/TM6 Ubx e;* +), and *pi25-1* is complete *P* element cloned in a plasmid.

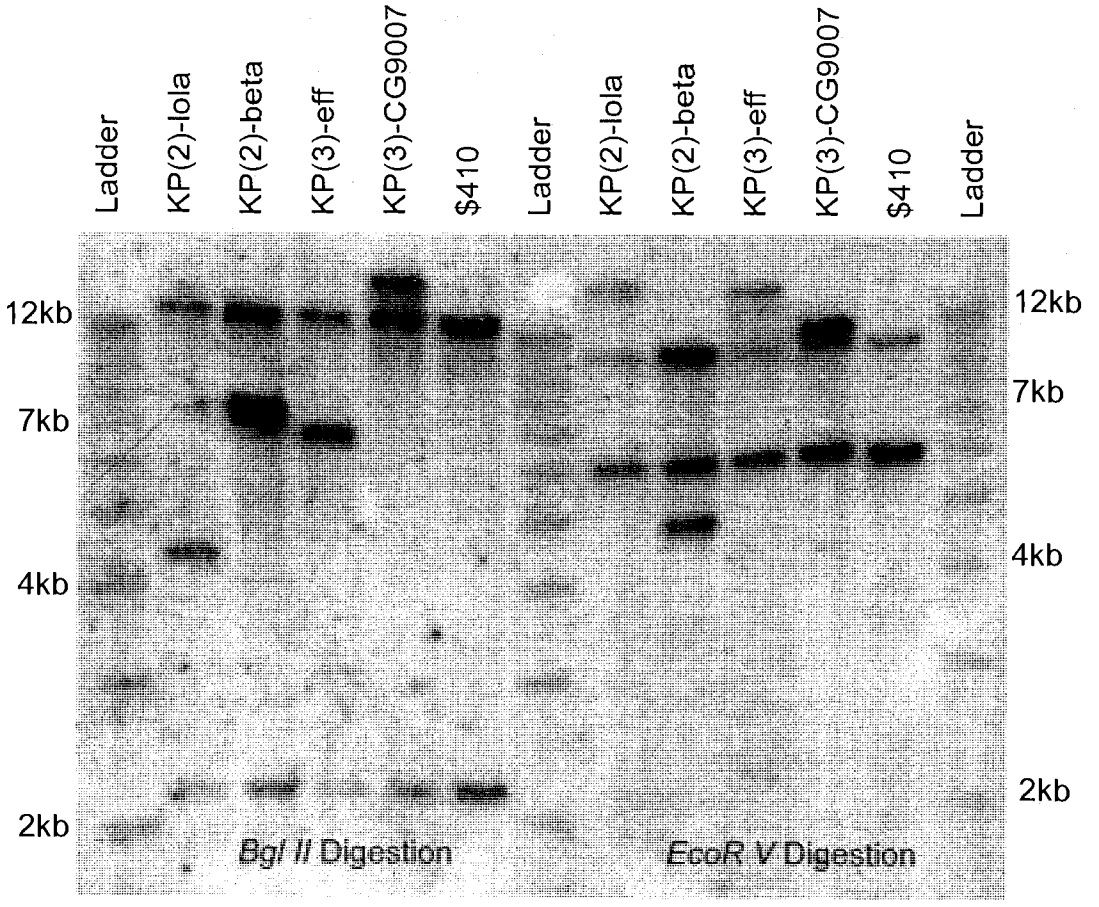


Figure 3-3) Southern blot of single *KP* mutants probed with *P* element DNA.

Genomic DNA was extracted, digested by *Bgl*II or *Eco*RV restriction enzymes, electrophoresed, and transferred to a nylon membrane before probing with random hexamer primed ^{32}P-dCTP DNA probe for the whole length of the complete *P* element. The X-ray film has been exposed to hybridized membrane for 24 hr. Ladder is Invitrogen 1kb^{+} ladder.

KP(2)-lola and KP(2)-beta are mutants with a *KP* insert on chromosome 2, KP(3)-eff and KP(3)-CG9007 are mutants with *KP* insert on chromosome 3. $410 is a *Pci* homozygous stock that is used as a control for the presence of *Pci*. The common 2.5 kb and 13 kb bands on *Bgl*II and 6 kb and 11 kb bands on *Eco*RV digestions are characteristic of *Pci*. In addition to *Pci* specific bands, KP(2)-lola, KP(2)-beta, KP(3)-eff and KP(3)-CG9007 each have only one band in both digestions. Therefore, it is more likely that they have only one *KP* insert.

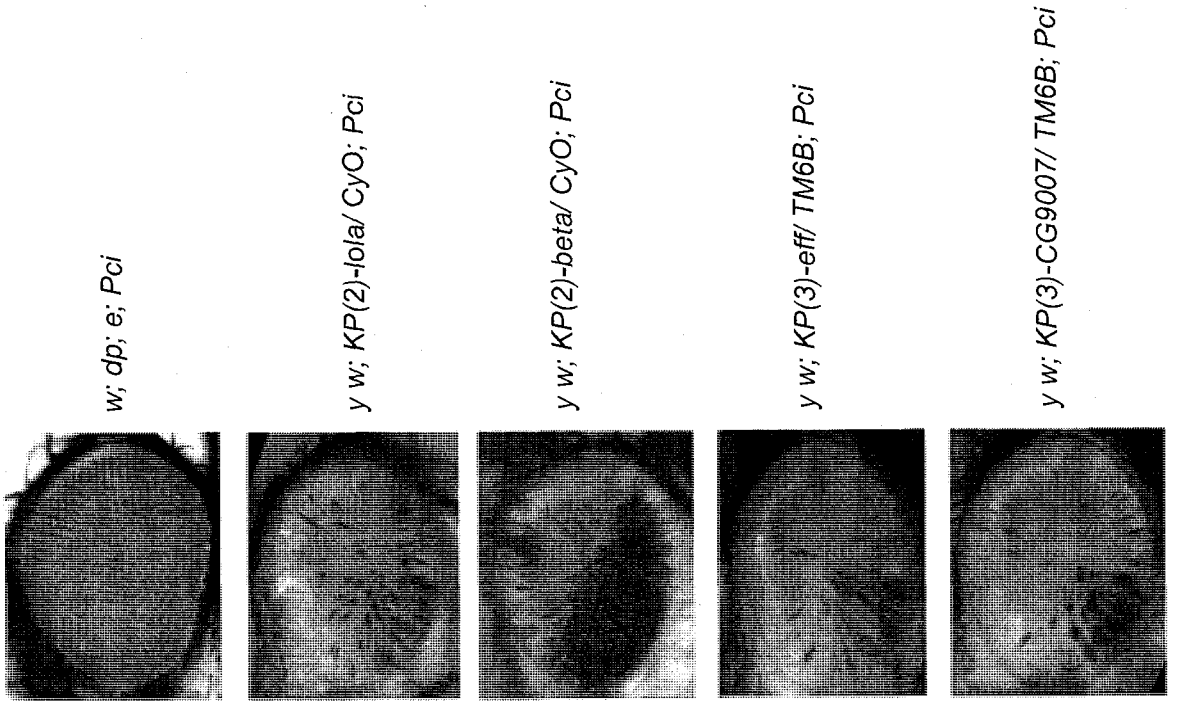


Figure 3-4) Eye photos of male flies from single *KP* stocks showing the extent of variegation induced by each insertion on *Pci* expression.

Each image is of a typical male of each *KP* mutant stock that is homozygous for *Pci* as mentioned in their genotype above the picture. The left panel is a *Pci* homozygous male fly without any *KP* insert that shows uniform red eye color.

KP inserts do not enhance *In(1)w^{m4}*, a classic hPEV allele

To test the possibility that the *KP* insertions caused an *E(var)* mutation that might be responsible for the original variegated eye color phenotype, I selected w$^-$ male KP(2) mutants (*y w; KP(2)/ CyO; +; +*) and crossed them to *In(1)w^{m4}; +; +; +* virgin females. If there was any *E(var)* mutation in the *KP(2)* chromosome, the frequency of w^{var} flies would be higher among the Cy$^+$ progeny (*KP(2)* carriers) in comparison to Cy progeny (*CyO* carriers). If there was a *Su(var)* mutation it would be reverse. For each mutant, the progeny were visually sorted into w^{var} and nearly w$^+$ (suppressed phenotype) in Cy$^+$ and Cy groups. In the progeny of the cross for both KP(2) mutants, there was no significant difference in the frequency of hPEV phenotype among the Cy$^+$ progeny versus Cy progeny. The smallest *p-value* was 0.217 which is much higher than the threshold of significance (0.05) (Table 3-2-a). Thus, these mutations failed to modify *In(1)w^{m4}* and I inferred that their enhancement of PDS is not due to a *KP* insertion into an *E(var)* locus. Comparing KP(2)-lola versus KP(2)-beta did not found any statistical difference between them either (*p-value*=0.422, Table 3-2-b).

I selected w$^-$ male KP(3) mutants (*y w; +; KP(3)/ TM6B Tb Hu; +*) and crossed them to *In(1)w^{m4}; + ; + ; +* virgin females. For each mutant, the progeny were visually sorted into w^{var} and nearly w$^+$ (suppressed phenotype) in Hu$^+$ (KP(3) carriers) and Hu (*TM6B Tb Hu* carriers) groups. I could not find any significant difference between KP(3)-eff and *TM6B Tb Hu* in the frequency of enhanced *In(1)w^{m4}* phenotype (*Chi square* test, *p-value*=0.469, Table 3-2-a). Since the *TM6B Tb Hu* chromosome, the balancer in KP(3) stocks, had an *E(var)3-9* mutation (Weiler, 2002), it is probable that KP(3)-eff has the *E(var)* mutation as well. Moreover, the cross between KP(3)-CG9007 and *In(1)w^{m4}* did not produce any Hu progeny that could indicate a very strong meiotic drift (Table 3-2-a). Comparison between KP(3)-eff and KP(3)-CG9007 did not show any statistical significant difference between them (*Chi square* test, *p-value*=0.191, Table 3-2-b). It seems unlikely that both of these mutants have a new *E(var)* mutation, thus most probably this phenotype is due to a mutation that was present on the parental chromosome. Although this can introduce a confounding factor to the *KP* dose-dependent test, during analysis of data I treated it as a constant. That is, I always compared the data from progeny that shared this factor; therefore this factor should not affect the analysis.

I did not compare KP(2) and KP(3) mutants with each other. The reason was the difference in the stocking scheme that had introduced different genetic backgrounds to these mutants.

Table 3-2-a) Effect of *KP* mutants on hPEV. Comparison between the *KP* carriers and the balancer carriers in the progeny of a cross between *KP* mutants and $In(1)w^{m4}$.

Name	Phenotype Frequency				Comparison with balancer
	***KP* chromosome**		**Balancer**		
	suppressed	**w^{var}**	**suppressed**	**w^{var}**	***Chi square, p-value***
KP(2)-lola	33	6	31	11	0.233
KP(2)-beta	54	6	51	11	0.217
KP(3)-eff	34	54	4	10	0.469
KP(3)-CG9007	66	73	0	0	No Test (No balancer carrier in progeny)

Table 3-2-b) Effect of *KP* mutants on hPEV. Comparison between *KP* mutants in their effect on hPEV.

Mutants	Phenotype Frequency				*Chi square, p-value*
	KP(2)-lola progeny		KP(2)-beta progeny		
	suppressed	w^{var}	suppressed	w^{var}	
KP(2)-lola vs. KP(2)-beta	33	6	54	6	0.422
	KP(3)-eff progeny		KP(3)-CG9007 progeny		
	suppressed	w^{var}	suppressed	w^{var}	
KP(3)-eff vs. KP(3)-CG9007	34	54	66	73	0.191

Footnote: *KP* mutants were crossed to $In(1)w^{m4}$ to test for any *E(var)* or *Su(var)* mutation in the mutant. I used cross-tabulation descriptive analysis to compare numbers statistically (SPSS 12.0). To do this I used *Chi square* test.

KP inserts show a partial additive PDS effect

To observe a possible *KP* dose-dependent effect on PDS, I crossed male KP(2) mutants (*y w; KP(2)/ CyO; +; Pci*) *inter se* to female KP(3) mutants (*y w; +; KP(3)/ TM6B Tb Hu; Pci*) resulting four different crosses. The balancer chromosome's markers permitted me to sort both male and female progeny into four separate groups (Figure 3-5):

(1) flies that had only a *KP(3)* element (*y w; +/ CyO; KP(3)/+; Pci*),

(2) flies that had only a *KP(2)* element (*y w; KP(2)/ +; +/ TM6B Tb Hu; Pci*),

(3) flies that had both *KP(2)* and *KP(3)* (*y w; KP(2) / +; KP(3)/ +; Pci*), and

(4) flies that had neither *KP* element (*y w; +/ CyO; +/ TM6B Tb Hu; Pci*).

Due to the presence of an *E(var)3-9* mutation on the *TM6B Tb Hu* chromosome (Weiler, 2002) I did not evaluate group (2) (*y w; KP(2)/ +; +/ TM6B Tb Hu; Pci*) in my study. Figure 3-5 shows photographs of eyes from the four different progeny groups of a typical cross.

Pigment A_{480} values (Table 3-3-a) showed that in male progeny, the presence of both *KP* elements gave more silencing (*i.e.* fewer colored ommatidia) than *KP(3)* alone in three out of four above mentioned crosses (one way ANOVA *p-value*< 0.001, largest Bonferroni *p-value*= 0.003). The only exception was the cross between KP(2)-beta and KP(3)-CG9007. In this cross the difference between group (1) and group (3) of the progeny was not statistically significant. Difference between group (1) and group (4), and (3) and (4) of the progeny of all crosses were significant (one way ANOVA *p-value*< 0.001, largest Bonferroni Post Hoc *p-value*< 0.001). In other words, although in all crosses there are significant differences in eye pigment between the presence and absence of *KP* element, in one cross, the difference between one and two *KP* elements is not significant.

In female progeny of three crosses, the presence of both *KP* elements silences *Pci* more than *KP(3)* alone (one way ANOVA *p-value*< 0.001, largest Bonferroni *p-value*= 0.005). There was one exception, female progeny of crosses between KP(2)-beta and KP(3)-eff did not show a statistically significant difference in eye pigment between group (1) and (3). Again in all four crosses, eye pigment A_{480} differences between group (1) and group (4), and group (3) and (4) of the progeny was significant (one way ANOVA *p-value*< 0.001, largest Bonferroni Post Hoc *p-value*< 0.001). Here again, the presence of at least one *KP* element in the genome is enough to change eye pigment measurements significantly, but at least in one cross there is not significant difference between one and two *KP* elements.

With two exceptions, both male and female progeny of all four crosses showed significantly lower eye pigment A_{480} in progeny group (3) than progeny group (1). The

exceptions are male progeny of the cross between KP(2)-beta and KP(3)-CG9007 and female progeny of the cross between KP(2)-beta and KP(3)-eff. However, in general, the additive crosses data support the hypothesis that the strength of PDS is related to the number of effective *KP* elements present in the genome, although this relationship is not linear (A_{480} results, Figure 3-6). In other words, the A_{480} difference between progeny group (4) and progeny group (3) is smaller than the sum of differences between A_{480} in progeny group (4) and each of the progeny group (2) and progeny group (3) of the same cross.

In the 3-3 tables "mean" refers to mean of three different samples for each group of progeny, "Sdev" refers to standard deviation, and "se" refers to standard error. The progeny groups are:

(1) flies that had only a *KP(3)* element (*y w; +/ CyO; KP(3)/+; Pci*)

(2) flies that had only a *KP(2)* element (*y w; KP(2)/ +; +/ TM6B Tb Hu; Pci*)

(3) flies that had both *KP(2)* and *KP(3)* (*y w; KP(2)/ +; KP(3)/ +; Pci*)

(4) flies that had neither *KP* element (*y w; +/ CyO; +/ TM6B Tb Hu; Pci*)

Table 3-3-a) Absorbance (A_{480}) of eye pigments of different progeny groups of crosses between KP(2) and KP(3) mutants.

Progeny group	KP(2)-lola X KP(3)-eff					
	Female Progeny				Male Progeny	
	mean	Sdev	se	mean	Sdev	se
(1)	0.17	0.005774	0.003333	0.17	0.020817	0.012019
(2)	0.07	0.01	0.005774	0.07	0.011547	0.006667
(3)	0.08	1.32E-09	7.6E-10	0.04	0.01	0.005774
(4)	0.32	0.01	0.005774	0.33	0.007071	0.005

Progeny group	KP(2)-lola X KP(3)-CG9007					
	Female Progeny				Male Progeny	
	mean	Sdev	se	mean	Sdev	se
(1)	0.15	0.025166	0.01453	0.13	0.03	0.017321
(2)	0.12	0.005774	0.003333	0.10	0.005774	0.003333
(3)	0.09	0.015275	0.008819	0.03	0.005774	0.003333
(4)	0.34	0.01	0.005774	0.39	0.007071	0.004082

Progeny group	KP(2)-beta X KP(3)-eff					
	Female Progeny				Male Progeny	
	mean	Sdev	se	mean	Sdev	se
(1)	0.14	0.015275	0.008819	0.16	0.015275	0.008819
(2)	0.13	0.02	0.011547	0.15	0.01	0.005774
(3)	0.10	0.011547	0.006667	0.08	0.017321	0.01
(4)	0.29	0.034641	0.02	0.36	0.028284	0.01633

Progeny group	KP(2)-beta X KP(3)-CG9007					
	Female Progeny				Male Progeny	
	mean	Sdev	se	mean	Sdev	se
(1)	0.18	0.026458	0.015275	0.11	0.011547	0.006667
(2)	0.26	0.011547	0.006667	0.21	0.025166	0.01453
(3)	0.09	0.005774	0.003333	0.07	0.005774	0.003333
(4)	0.33	0.01	0.005774	0.41	0.007071	0.004082

Table 3-3-b) Statistical analysis of eye pigment absorbance (A_{480}) results for male progeny of additive crosses.

KP(2)-lola X KP(3)-eff					
Progeny group	Male Progeny			Bonferroni post-hoc test, *p-value*	
	mean	Sdev	se	(1) vs. (2)	<0.001
(1)	0.17	0.020817	0.012019	(1) vs. (3)	<0.001
(2)	0.07	0.011547	0.006667	(1) vs. (4)	<0.001
(3)	0.04	0.01	0.005774	(2) vs. (3)	0.947
(4)	0.33	0.007071	0.005	(2) vs. (4)	<0.001
				(3) vs. (4)	<0.001

KP(2)-lola X KP(3)-CG9007					
Progeny group	Male Progeny			Bonferroni post-hoc test, *p-value*	
	mean	Sdev	se	(1) vs. (2)	0.336
(1)	0.13	0.03	0.017321	(1) vs. (3)	0.001
(2)	0.10	0.005774	0.003333	(1) vs. (4)	<0.001
(3)	0.03	0.005774	0.003333	(2) vs. (3)	0.020
(4)	0.39	0.007071	0.004082	(2) vs. (4)	<0.001
				(3) vs. (4)	<0.001

KP(2)-beta X KP(3)-eff					
Progeny group	Male Progeny			Bonferroni post-hoc test, *p-value*	
	mean	Sdev	se	(1) vs. (2)	1.000
(1)	0.16	0.015275	0.008819	(1) vs. (3)	0.003
(2)	0.15	0.01	0.005774	(1) vs. (4)	<0.001
(3)	0.08	0.017321	0.01	(2) vs. (3)	0.010
(4)	0.36	0.028284	0.01633	(2) vs. (4)	<0.001
				(3) vs. (4)	<0.001

KP(2)-beta X KP(3)-CG9007					
Progeny group	Male Progeny			Bonferroni post-hoc test, *p-value*	
	mean	Sdev	se	(1) vs. (2)	0.043
(1)	0.11	0.011547	0.006667	(1) vs. (3)	1.000
(2)	0.21	0.025166	0.01453	(1) vs. (4)	<0.001
(3)	0.07	0.005774	0.003333	(2) vs. (3)	0.007
(4)	0.41	0.007071	0.004082	(2) vs. (4)	0.002
				(3) vs. (4)	<0.001

Table 3-3-c) Statistical analysis of eye pigment absorbance (A_{480}) results for female progeny of additive crosses.

Progeny group	KP(2)-lola X KP(3)-eff				
	Female Progeny			Bonferroni post-hoc test, *p-value*	
	mean	Sdev	se	(1) vs. (2)	<0.001
(1)	0.17	0.005774	0.003333	(1) vs. (3)	<0.001
(2)	0.07	0.01	0.005774	(1) vs. (4)	<0.001
(3)	0.08	1.32E-09	7.6E-10	(2) vs. (3)	0.657
(4)	0.32	0.01	0.005774	(2) vs. (4)	<0.001
				(3) vs. (4)	<0.001

Progeny group	KP(2)-lola X KP(3)-CG9007				
	Female Progeny			Bonferroni post-hoc test, *p-value*	
	mean	Sdev	se	(1) vs. (2)	0.079
(1)	0.15	0.025166	0.01453	(1) vs. (3)	0.005
(2)	0.12	0.005774	0.003333	(1) vs. (4)	<0.001
(3)	0.09	0.015275	0.008819	(2) vs. (3)	0.468
(4)	0.34	0.01	0.005774	(2) vs. (4)	<0.001
				(3) vs. (4)	<0.001

Progeny group	KP(2)-beta X KP(3)-eff				
	Female Progeny			Bonferroni post-hoc test, *p-value*	
	mean	Sdev	se	(1) vs. (2)	1.000
(1)	0.14	0.015275	0.008819	(1) vs. (3)	0.251
(2)	0.13	0.02	0.011547	(1) vs. (4)	<0.001
(3)	0.10	0.011547	0.006667	(2) vs. (3)	0.630
(4)	0.29	0.034641	0.02	(2) vs. (4)	<0.001
				(3) vs. (4)	<0.001

Progeny group	KP(2)-beta X KP(3)-CG9007				
	Female Progeny			Bonferroni post-hoc test, *p-value*	
	mean	Sdev	se	(1) vs. (2)	0.001
(1)	0.18	0.026458	0.015275	(1) vs. (3)	<0.001
(2)	0.26	0.011547	0.006667	(1) vs. (4)	<0.001
(3)	0.09	0.005774	0.003333	(2) vs. (3)	<0.001
(4)	0.33	0.01	0.005774	(2) vs. (4)	0.001
				(3) vs. (4)	<0.001

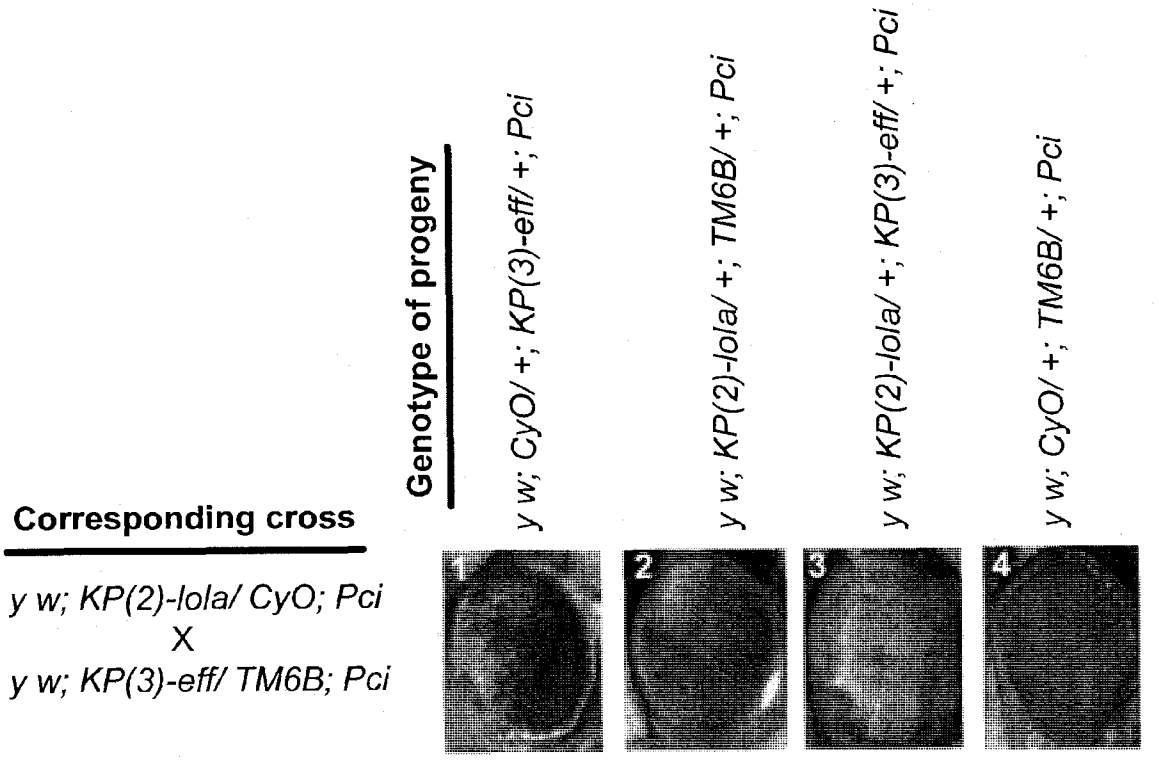


Figure 3-5) The extent of PDS is dependent on the number of *KP* present in the genome.

Eye pictures from male progeny of a cross between KP(2) and KP(3) stocks permit the examination of phenotype of flies with nil, one, and two copies of *KP* in the genome. The left column indicates the cross. The genotype for each fly is indicated above its picture. Number of each photograph indicates the corresponding progeny groups in the text.

(1) flies that had only a *KP(3)* element (*y w; +/ CyO; KP(3)/+; Pci*)

(2) flies that had only a *KP(2)* element (*y w; KP(2)/ +; +/ TM6B Tb Hu; Pci*)

(3) flies that had both *KP(2)* and *KP(3)* (*y w; KP(2)/ +; KP(3)/ +; Pci*)

(4) flies that had neither *KP* element (*y w; +/ CyO; +/ TM6B Tb Hu; Pci*)

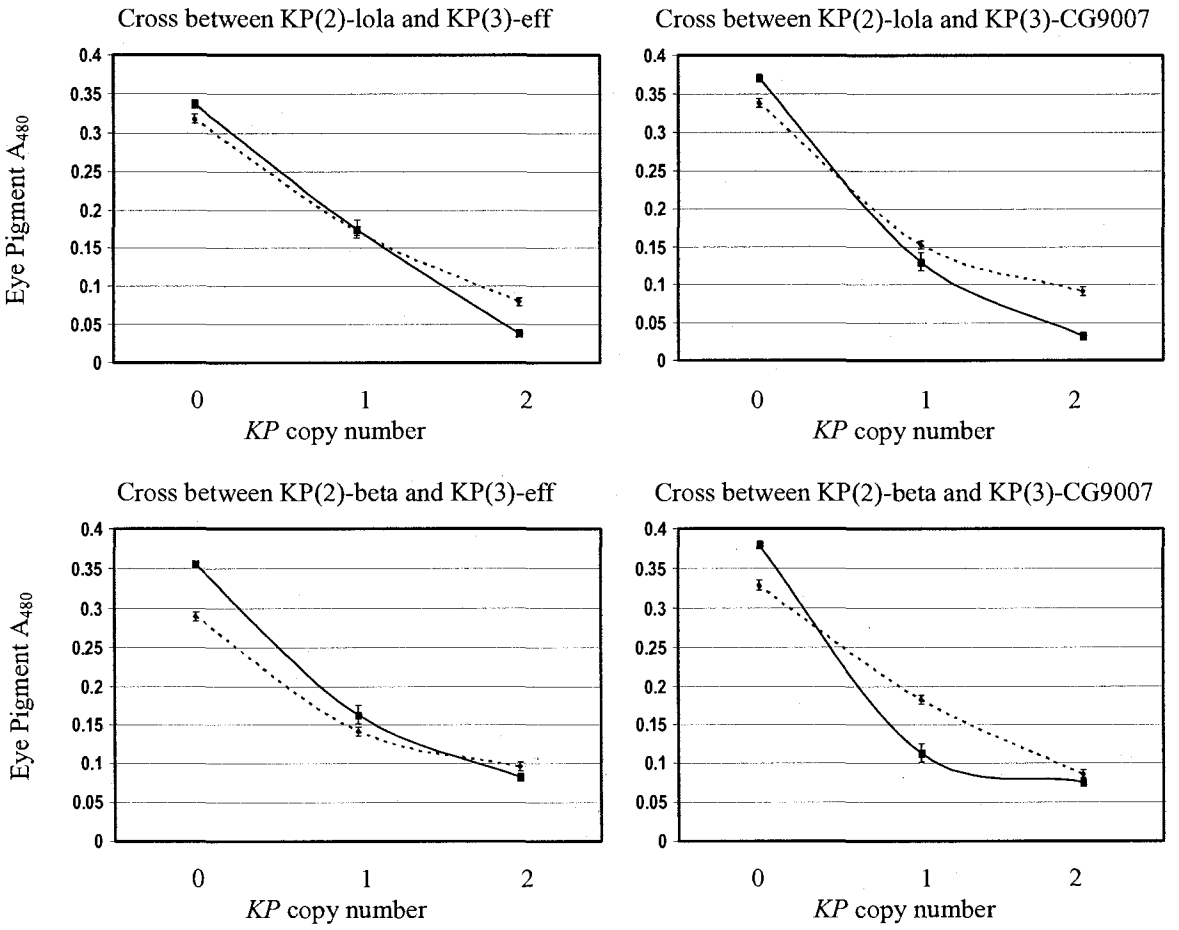


Figure 3-6-a) Eye pigment absorbance (A_{480}) for flies with 0, 1, 2 copies of *KP* elements.

Progeny of crosses between different *KP* mutants were gathered and classified based on sex and number of *KP* inserts in their genome. Amount of eye pigment in each class was measured in three samples containing 15 flies in each. The average for each class is plotted above. Error bars represent standard errors (N=3). Dashed line represents female progeny and continuous line represents male progeny for each cross. *KP* number changes from zero to two based on the number of *KP* inserts present in genome.

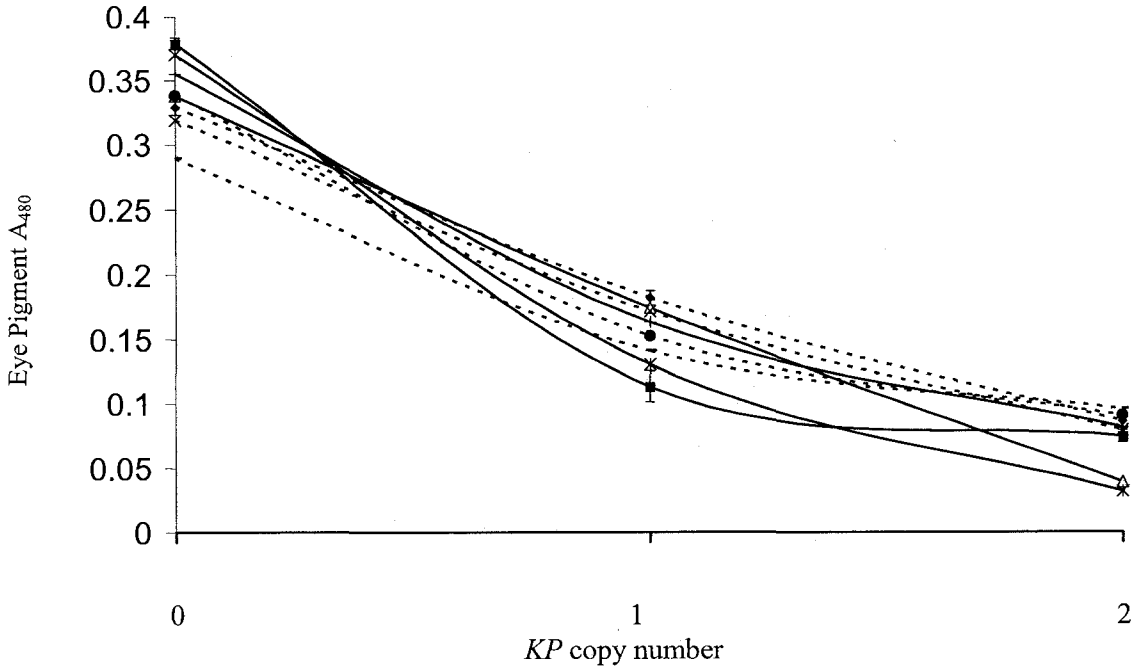


Figure 3-6-b) This is an overlay of both male and female progeny of all crosses compiled from Figure 3-6-a.

Chapter 4: Mutation analysis of *P[SalI]* for its functional domains in PDS

Introduction

P[SalI], a 10.3 kb ry^+ bearing *P* construct, is the classic example of a non-autonomous type I repressor (Figure 1-2). This *P* construct consists of a complete *P* element that has a 4 bp insertion in the *Sal*I site located in the middle of last exon. In *P[SalI]*, the last 23 bp of *P* element sequence is replaced by a construct consists of a $ry^{7.2+}$ gene plus the last 224 bp of complete *P* element sequence (Figure 1-2) (Karess & Rubin, 1984). Due to a frame-shift mutation in the middle of exon 3, the *P* element gene of the *P[SalI]* construct is capable of encoding a type I repressor protein but not the transposase (Robertson and Engels, 1989). Bushey and Locke (2004) showed that *P[SalI]89D* is able to silence mini-white gene expression in *Pci* and *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$. It also suppresses variegation of expression of the *mini-white* gene in *P{hsp26-pt-T}39C-12* (Bushey, 2004). My work has shown that *P[SalI]* is inserted in cytological region 89B, not 89D. Thus, to prevent further confusion hereafter I'll use *P-Sal* instead of *P[SalI]89D*. The purpose of this study is to identify important domains in the type I repressor protein for its effect on *Pci*. To do this I had two choices: *in vivo* mutagenesis or *in vitro* mutagenesis. The first method was preferred because it provides a chance to study both predicted and non-predicted domains whereas the *in vitro* mutagenesis would be limited as it would be targeted toward only the previously predicted domains. Also, *in vivo* mutagenesis using EMS essentially makes single base changes and the chromatin status of different mutants would remain the same, however, in the *in vitro* mutagenesis system the random insertion of mutated transgenes elements would affect their level of expression due to variation of the surrounding DNA sequence and its chromatin status.

After mutagenesis, each mutant was tested on *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ and *P{hsp26-pt-T}39C-12* to study effect of the mutations on these variegating w^+ genes as well.. Since both type I (e. g. *P-Sal*) and type II (e. g. *KP*) repressor proteins are able to silence expression of the *mini-white* gene in *Pci* and *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ (Bushey and Locke, 2004), my hypothesis was that the domains required for PDS should be restricted to the region which is common for both the type I repressor protein and the KP protein, *i.e.* before base pair 808 (Figure 1-1). To address this question I mutagenized a *Sb* marked chromosome containing *P-Sal*. I selected *P-Sal** mutants with reduced or no PDS phenotype. I then tested these mutants against *In(1)*w^{m4} to assay for any *Su(var)* effect. Furthermore, I tested these mutants against the parental *P-Sal* to test for any

antimorphic effect. I also tested them for their effects on *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, *P{hsp26-pt-T}39C-12* and *E1*, which are different *P* constructs inserted at different locations on chromosome 4 and are sensitive to the presence of *P-Sal* in the genome. In summary, I found 22 mutants, which I called MPS after "mutant *P-Sal*". Within the *P-Sal* coding sequence for 21 out of 22 mutants I found DNA sequence changes. Since there are two pairs with exactly the same mutations (MPS-5-1 and MPS-5-2, and MPS-11-1 and MPS- 11-2), out of original 22 mutants I found 19 different point mutations and one mutant without any recognized mutation.

Results

Exact location of the *P[SalI]89D* insert

To find the exact location of insertion of *P[SalI]89D* I amplified its 5' end using vectorette PCR (Riley *et al*, 1990). I sequenced this PCR product to find the genomic DNA sequence adjacent to *P-Sal* (chapter 2). This genomic sequence was searched against the *Drosophila melanogaster* genomic sequence database at the National Center for Biotechnology Information (NCBI) using Basic Local Alignment Search Tool (BLAST). I found that *P-Sal* is inserted between base pairs 12279162 and 12279161 on the right arm of chromosome 3. This is between base pairs 388 and 389 of a gene called *Pak3* at the cytogenetic map position 89B16 (Davies, 2001). This location is different from that originally reported for *P-Sal* at 89D that was done by in situ hybridization (Robertson and Engels, 1989). *P-Sal* is in the 5'-UTR region of the first exon of *Pak3* and *P-Sal* transcription is in the same direction as *Pak3*. This location and orientation suggests that *P-Sal* follows the *Pak3* expression pattern. Although I could not find any report about *Pak3*'s expression pattern, it is known that the PAK3 protein is involved in actin protein organization and the cytoskeleton (Kiger *et al*, 2003). Based on PAK3 protein function it is expected that *Pak3* is a widely expressed gene.

Genetic distance from *P-Sal* to *Sb*

The *P-Sal* chromosome that I used for my mutagenesis has a dominant *Sb* marker, which I used to follow the *P-Sal** mutants in the crossing scheme. Since I was going to select for amorphic *P-Sal** mutants, determining the linkage distance between *Sb* marker and the P-Sal was crucial. Robertson and Engels (1989) have already reported that *P-Sal* is within one map unit of *Sb*. To find the genetic distance between *P-Sal* and *Sb* I crossed *y w; +; Sb P-Sal / TM6B Tb Hu;* + virgin females to *w; dp; e; ci*^{1} *ey*^{R} males. I selected Sb virgin females (*y w/ w; +/ dp; Sb P-Sal / e; +/ ci*^{1} *ey*^{R}) in the progeny and crossed them to *w; dp; e; Pci* males. The progeny of this cross were examined for a PDS phenotype (as a marker for the presence of the *P-Sal* construct) and Sb phenotype. Seven out of 967 scored flies showed a crossover between these markers. Thus, the

genetic map distance between *P-Sal* and *Sb* is 0.7 cM, which is similar to the previously published value (Robertson and Engels, 1989).

***P-Sal* sequence**

I sequenced *P-Sal* in the parental stock (*y w; +; Sb P-Sal / TM6B Tb Hu; +*) from base pair 23 through base pair 2831 and compared the results with the published consensus *P* element sequence by Genetools (Biotools) (O'Hare and Rubin, 1983). There were two differences. *P-Sal* has a known variation at base pair 33. The 33A→T variation is a polymorphism previously found in autonomous *P* elements (O'Hare and Rubin, 1983). The second difference was also expected (Karess and Rubin, 1983). There was a TCGA insertion between base pairs 2414 and 2415 that disrupted the *Sal*I restriction site and introduced a *Pvu*I restriction site in its place.

Genetic Screen of *P-Sal* mutations for loss of the PDS phenotype

I used ethylmethane sulfonate (EMS) to mutagenize more than 4000 *y w; +; Sb P-Sal / TM6B Tb Hu; +* males and crossed them to *w; dp; e; Pci* virgin females in 517 vials. The progeny of these crosses were screened for loss of the PDS phenotype, which is recognized by the appearance of more pigment in the eyes of Sb progeny. After screening 14,565 Sb progeny I found 237 putative *P-Sal** mutants (1.6% of screened Sb progeny) with homogenous red eye phenotype *i.e.* lack of PDS. I back crossed each putative mutant individually to *w; dp; e; Pci*. Only 44 out of 237 putative *P-Sal** mutants (0.3% of original screened Sb progeny and 18.5% of putative mutants) retained the lack of PDS phenotype when transmitted to the next generation.

To reduce the possibility that the lack of PDS phenotype is due to a secondary site mutation on another chromosome, I replaced as many of the EMS mutagenized chromosomes as possible during the stocking process. To do this for each mutant, I selected a few dp Sb males with w^+ or very weak w^{var} (nearly w^+) eye color phenotype (*w; dp; mutant Sb P-Sal/ e; +/ Pci*) to stock the mutant. Therefore, chromosome X, both chromosome 2s, one of the chromosome 3s, and at least one of the chromosome 4s are coming from a non-mutagenized parent. For each mutant I crossed these males to *y w; +; TM6B Tb Hu/* ap^{Xa}; + virgin females and then selected w^+ Sb Hu Xa^+ males and virgin females (*w/ y w; dp/ +; mutant Sb P-Sal / TM6B Tb Hu; Pci/ +*) and crossed them to each other to make a stable stock. The progeny of this cross have received only two chromosomes from the mutagenized parent, the Y chromosome and one chromosome 3 (*Sb P-Sal*). For all 44 mutants the phenotype of the progeny of this stocking cross was stable and maintained a lack of PDS phenotype during observation for five generations. Each stocked mutant line tested positive for the presence of *P-Sal* by PCR using the KPF1/KPR2 primers (data not shown). This set of primers does not amplify *Pci* but will amplify *P-Sal*.

Differentiating mutations in *P-Sal* from mutations in second site modifiers on chromosome 3

The mutagenesis protocol should have recovered two classes of chromosome 3 mutations: those in *P-Sal* itself thereby inactivating the silencing, and those in second site modifiers that act to modify PDS. I used a genetic mapping technique to differentiate *P-Sal** mutants from second site modifiers. For each mutant line, I crossed *y w; +; mutant Sb* P-*Sal / TM6B Tb Hu;* + males to *w; dp; e;* + virgin females and selected *w; dp/ +; mutant Sb P-Sal/ e;* + virgin females in the progeny and backcrossed them to *w; dp; e; Pci* males. For each mutant, more than 130 progeny of this cross were scored for the presence of the PDS phenotype (Table 4-1). The presence of one or more PDS flies in the progeny means segregation between *P-Sal* and the PDS suppressor mutation. Reciprocally, the absence of any PDS phenotype among the progeny lets me conclude that the loss of the PDS phenotype maps near *P-Sal* and is most likely due to a mutation in *P-Sal* itself. With 130 progeny showing no PDS phenotype, the mutation would be mapped to less than 2.9 cM on either side of *P-Sal*. From the original 44 putative mutations, 22 mapped within this distance from *P-Sal* (*Chi square* test, *p-value*<0.05). These 22 mutant lines will hereafter be referred to collectively as *P-Sal** mutants. The remainder of the mutant lines are assumed to be mutated in various second site modifiers of PDS. These mutants were not examined further and may be the subject of another study about suppressor of PDS.

To make the calculation that 130+ progeny were needed, I first hypothesized that the mutation is not within "y" map units (cM) of *P-Sal*. Thus, if I score "x" flies in the progeny, I expected at least y/100 of them (x • y/100) to be recombinant, *i.e.* showing PDS phenotype. Since none of them showed PDS in my observation the *Chi square* table would be:

	Recombinant (PDS)	Parental (suppressed PDS)
Observed	0	x
expected	x•y/100	x•(100-y)/100

And the *Chi square* value would be x•y /(100-y). Critical *Chi square* value for df=1 and *p-value*=0.05 is 3.841, therefore if x•y/(100-y) = 3.9, I can reject the hypothesis, *i.e.* the mutation is within "y" map unit of *P-Sal*. In other words, if I score "x" number of progeny without any PDS among them, I can map the mutation within "y" map unit of *P-Sal* using the formula y=390/(x+3.9) thus if I score 130+ flies (x=130), the mutation will be within 2.9 cM from *P-Sal* with a *p-value* of less than 0.05 (Figure 4-1). Table 4-1 shows the maximum possible distance between any second-site mutation and *P-Sal* for each of 22 mapped mutants. All of these mutants

show a strong suppressed PDS phenotype, except one, MPS-19, which was only moderately suppressed (Table 4-1 & Figure 4-2).

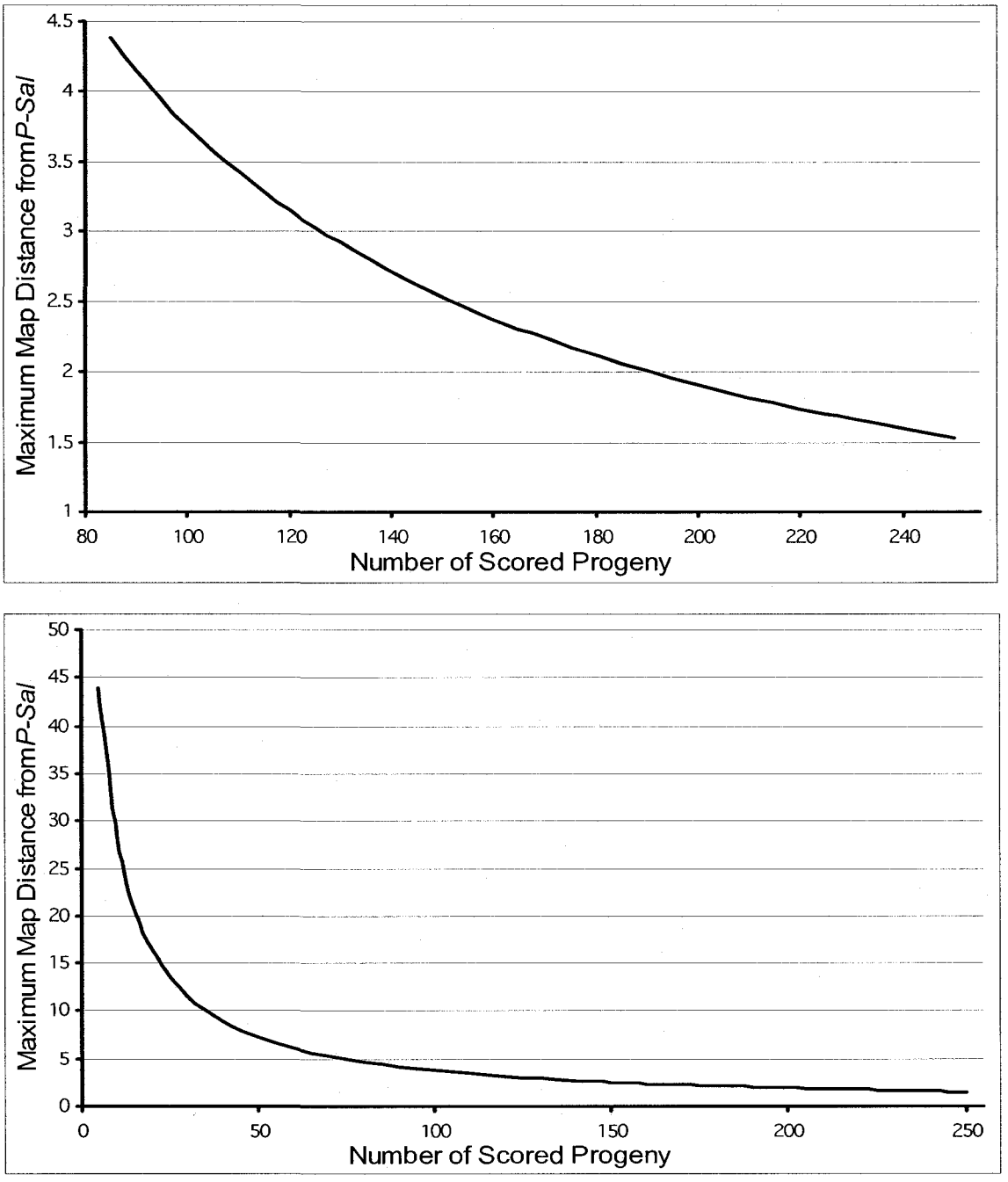


Figure 4-1) Calculated relationship between number of progeny and the maximal distance from *P-Sal* based on a *p-value* < 0.05 and the formula y=390/(x+3.9).

The line defines the number of progeny needed to conclude, with a 95% probability, that the suppressor of PDS mutation is within the plotted distance from the *P-Sal* insertion site. 130 non-recombinant progeny for any given mutant is enough to establish the linkage between the mutation and *P-Sal*. Lower panel: the large scale view.

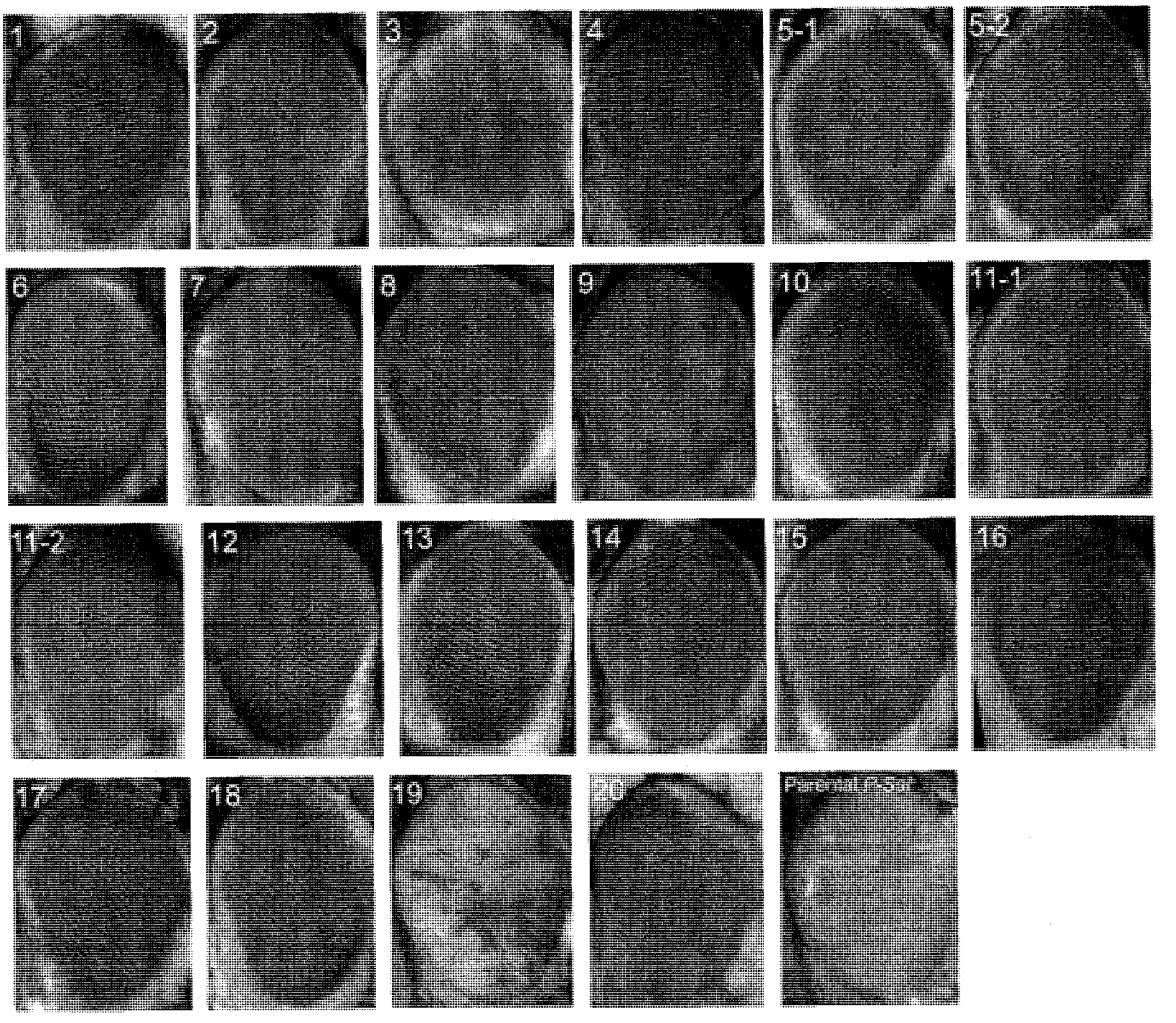


Figure 4-2) *P-Sal mutants.**

Eye photos of male *P-Sal** mutants (labeled 1-20) are compared with a parental *P-Sal* eye (lower right). The genotype was *y w; Sb P-Sal*/ TM6B Tb Hu; Pci*. For the parental *P-Sal* fly the genotype was *y w; Sb P-Sal/ TM6B Tb Hu; Pci/ +*. Numbers represent the numbered MPS mutant. While most mutations show a uniform red eye (PDS amorph), MPS-2, MPS-3, MPS-7, MPS-19, and MPS-20 show a PDS hypomorph phenotype.

MPS: Mutated *P-Sal*

Table 4-1) Summary of *P-Sal** mutants.

Mutant Name	Mutant Phenotype (A: amorph,, H-A: Strong hypomorph. H: hypomorph)	Number of scored flies	Mutation Maximum map distance from *P-Sal* insert (*p-value*<0.05)	DNA Change	Amino acid (AA) Change	AA Polarity Change (A: Acidic, B: Basic, P: Polar, NP: Nonpolar)	AA Water affinity (L: Hydrophil, B: Hydrophobe, S: neutral)	AA Tendency to Zndary Structure (A: Alpha Helix, B: Beta Sheet, T: Turn)	Phenotype pattern
MPS-1	A	413	0.92	238 G>A	29 Trp>**STOP**				
MPS-2	H-A	345	1.10	240 G>A	30 Glu>Lys	A>B	L>L	A>A	*39C-12*
MPS-3	H-A	218	1.73	283 G>A	44 Cys>Tyr	P>P	S>S	A>B	
MPS-4	A	430	0.89	346 G>A	65 Arg>Lys	B>B	L>L	A>A	
MPS-5-1	A	342	1.11	500 G>A	**SPLICING**				
MPS-5-2	A	366	1.04						
MPS-6	A	234	1.61	665 G>A	152 Gly>Glu	NP>A	S>L	T>A	Antimorph & *39C-12*
MPS-7	H-A	191	1.97	712 C>T	168 Leu>Phe	NP>NP	B>B	A>B	*39C-12*
MPS-8	A	294	1.29	794 G>A	195 Trp>**STOP**				*39C-12*
				1089A>T	**N/A**				
MPS-9	A	369	1.03	821 G>A	204 Gly>Glu	NP>A	S>L	T>A	
MPS-10	A	236	1.60	1097 T>A	296Leu>**STOP**				
MPS-11-1	A	199	1.89	1100 T>A	297 Val>Glu	NP>A	B>L	B>A	
MPS-11-2	A	278	1.36						
MPS-12	A	305	1.24	1106 C>T	299 Ala>Val	NP>NP	S>B	A>B	
MPS-13	A	288	1.31	1228 C>T	322 Thr>Ile	P>NP	S>B	B>B	Antimorph & *39C-12*
MPS-14	A	335	1.13	1352 A>T	363 Leu>Phe	NP>NP	B>B	A>B	
MPS-15	A	206	1.83	1473 G>A	404 Ala>Thr	NP>P	S>S	A>B	*39C-12*
MPS-16	A	131	2.85	1645 C>T	461 Pro>Leu	NP>NP	S>B	T>A	Antimorph & *39C-12*
MPS-17	A	325	1.17	1750 G>T	496 Gly>Val	NP>NP	S>B	T>B	Antimorph
MPS-18	A	211	1.79	1828 C>T	522 Thr>Ile	P>NP	S>B	B>B	
MPS-19	H	387	0.98	1963 A>T	567 Lys>Ile	B>NP	L>B	A>B	*In(1)w^{m4}, E1 & 39C-12*
MPS-20	H-A	1411	0.27	**None**					*39C-12*

Footnote: In Table 4-1 "Antimorph" refers to Antimorphic effect against *P-Sal, "E1"* refers to ability to enhance variegation of *E1*, "39C-12" refers to ability to suppress variegation of *P{hsp26-pt-T}39C-12,* and "In(1)w^{m4}" refers to ability to enhance variegation of *In(1)w^{m4}*. Also AA is an abbreviation for amino acid and MPS refers to Mutated *P-Sal.* The symbol ">" in "x>y" represents that x is changed to y.

No second site modifiers are near *P-Sal*: *Df(3)* experiment

The mutations that mapped to less than 2.9 cM from *P-Sal* might be due to a dominant second site modifier gene in this 5.8 cM region (*P-Sal: 59* $\pm$ 2.9 cM). *P-Sal* is located at cytological region 89B, which maps to genetic position 59 on chromosome 3 (Lindsley, 1992). Map unit 56 is equal to region 88D1, and map unit 62 maps to region 91A (Lindsley, 1992). Since all of the mutants were mapped within 2.9 cM of *P-Sal*, I used a series of *Df(3)*s spanning more than 3 cM on either side of *P-Sal* (Table 4-2).

The presence of dominant modifies can be tested using deficiencies (*Df(3)*s) missing parts of this region. If a second site modifier locus was present in this region, and its loss should dominantly suppress PDS in a heterozygous state (*w; dp; mutant Sb P-Sal / e; Pci/ +*), thus I should be able to produce a suppressed PDS phenotype with deletions of this region of chromosome 3, *i.e. w; +; P-Sal / Df(3); Pci/ +* should show a suppressed PDS phenotype.

I crossed males from different *Df(3)*s to *y w; +; P-Sal; Pci* virgin females and scored progeny for suppression of variegation. Since *w; +; P-Sal/ +; Pci/ +* flies show essentially a w^- phenotype, I would classify all w^+ and w^{var} flies in the progeny as suppressed PDS and would compare the frequency of w^{var} phenotype between each *Df(3)* (*y w; +; P-Sal / Df(3); Pci/ +*) and its balancer (*y w; +; P-Sal / chromosome 3 balancer of the stock; Pci/ +*). Tested *Df(3)* mutations showed no w^{var} progeny in more than 60 flies that were scored (Table 4-2). I found that neither the balancer nor the *Df(3)* were able to suppress PDS and thus there is not a locus within this region that produces a dominant Suppressor of PDS phenotype when has amorphic mutation.

Table 4-2) *Df(3)* set in the area ±3 cM from *P-Sal.*

Bloomington Stock Number	**Break Points**	**Number of screened flies**	**Genotype**
1534	87D1-2;88E5-6	97	*Tp(3;Y)ry506-85C/MKRS*
383	88E7-13;89A1	88	*Df(3R)ea, kni[ri-1] p[p]/TM3, Ser[1]*
756	88F9-89A1;89B9-10	98	*Df(3R)sbd105, p[p] Ubx[bx-1] sr[1] e[s]/TM3, Ser[1]*
1920	89B5;89C2-7	100	*C(1;Y)1, y[+]; Df(3R)sbd104/TM2, ry[*]*
1467	89B7-8;89E7	63	*Df(3R)P115, e[11]/TM1; Dp(3;1)P115/+*
4431	89E1-F4;91B1-B2	61	*Df(3R)DG2/TM2, red[1]*
1422	87E8-F1;93C	68	*Df(3R)MRS/Dp(3;3)MRS, M(3)76A[1] kar[1] ry[2] Sb[1]*

Effect of *P-Sal mutants on hPEV**

Bushey and Locke (2004) reported that some *Su(var)* mutants can affect both the hPEV and PDS systems. Thus, the 22 mutants I identified as *P-Sal**, which showed suppression of PDS, still had a possibility of being due to a *Su(var)* mutation (although there is no known *Su(var)* gene in the 3 cM distance from *P-Sal*). Therefore, I tested them against a classic hPEV variegating allele, $In(1)w^{m4}$.

For each *P-Sal** mutant, I crossed *y w; +; Sb P-Sal*/ TM6B Tb Hu; +* males to $In(1)w^{m4}$; *+; +; +* virgin females, and sorted the Sb progeny into variegated (w^{var}) and suppressed (nearly w^{+}). Frequency of these progeny classes were compared with the similar frequencies in progeny of cross between the parental *P-Sal* stock (*y w; +; Sb P-Sal/ TM6B Tb Hu; +*) and $In(1)w^{m4}$; *+; +; +*. In this experiment, both the control (parental *P-Sal*) and all *P-Sal** mutants were balanced over the same *TM6B Tb Hu* chromosome. Although this chromosome contains an *E(var)3-9* mutation (Weiler, 2002), its possible enhancing effect does not affect the comparison between control (parental *P-Sal*) and *P-Sal** mutants because it is present in all of them. In 21 out of the 22 *P-Sal** mutants, there was no statistically significant difference between the *P-Sal** mutant and the parental *P-Sal* chromosomes in the $In(1)w^{m4}$ phenotype modification (the smallest *p-value* was 0.094, Table 4-3). Only MPS-19 showed increased frequency of variegated phenotype in comparison to the parental *P-Sal* (*Chi square* test, *p-value* <0.001) indicating an hPEV enhancing effect and probably a second site *E(var)* mutation.

In addition to the above comparison, in the progeny of each cross, I also compared *Sb P-Sal** carriers versus *TM6B Tb Hu* chromosome carriers. I found in all crosses (except the cross involving MPS-19) the frequency of variegated flies is higher in Tb Hu flies than Sb flies (the largest *Chi square* test *p-value* was 0.029, Table 4-3), *i.e.* variegation was enhanced in *TM6B Tb Hu* carrier progeny in comparison with *Sb P-Sal** carrier progeny. This is probably due to the *E(var)3-9* mutation on *TM6B Tb Hu* chromosome (Weiler, 2002). MPS-19 was an exception again. In the progeny of the cross between the MPS-19 mutant and $In(1)w^{m4}$; *+; +; +*, there was no statistically significant difference between the frequency of variegated flies among Sb versus Tb Hu progeny. This supports my previous suggestion that there is an *E(var)* mutation on the *Sb MPS-19* chromosome.

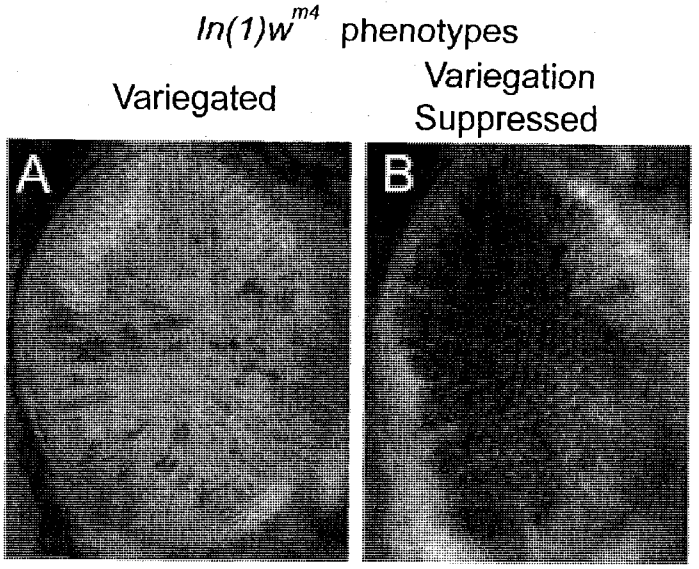


Figure 4-3) Phenotypes of *In(1)w^{m4}* stock, a classic hPEV allele.

This stock (*In(1)w^{m4}; +; +; +*) is dimorphic as shown in this figure. Majority of the progeny shows the variegated eye color (panel A) and the rest show suppressed variegation phenotype (panel B). *Su(var)* mutations increase the frequency of the suppressed phenotype (B) while *E(var)* mutations increase the frequency of the variegated phenotype (A). I crossed the *P-Sal** mutants (*y w; Sb P-Sal*/ TM6B Tb Hu; +*) to this stock and scored the progeny for the frequency of these two phenotypes (A and B) in Sb (mutated chromosome 3) and Tb Hu (*TM6B Tb Hu* balancer chromosome) background. I compared these frequencies in each mutant with the parental *P-Sal* stock.

Table 4-3) Effect of *P-Sal** mutants on hPEV.

Mutant Name	Phenotype Frequency				*Chi square* test, *p-value*	
	Sb		Tb Hu		Comparison with *TM6B Tb Hu* balancer	Comparison with parental *P-Sal*
	suppressed	w^{var}	suppressed	w^{var}		
P-Sal	36	2	2	29	<0.001	N/A
MPS-1	26	0	17	13	<0.001	0.21 f
MPS-2	35	1	18	5	0.029 f	1 f
MPS-3	12	0	0	5	<0.001 f	1 f
MPS-4	36	5	11	34	<0.001	0.434 f
MPS-5-1	25	1	5	17	<0.001	1 f
MPS-5-2	37	3	2	34	<0.001	1 f
MPS-6	23	0	0	21	<0.001	0.522 f
MPS-7	30	0	5	39	<0.001	0.5 f
MPS-8	37	0	6	11	<0.001 f	0.493 f
MPS-9	48	5	3	39	<0.001	0.695 f
MPS-10	33	0	9	22	<0.001	0.495 f
MPS-11-1	19	1	1	27	<0.001	1 f
MPS-11-2	35	8	15	28	<0.001	0.094 f
MPS-12	31	0	0	23	<0.001	0.498 f
MPS-13	45	3	19	22	<0.001	1 f
MPS-14	17	2	0	15	<0.001	0.594 f
MPS-15	35	5	1	43	<0.001	0.432 f
MPS-16	14	2	14	22	<0.001	0.573 f
MPS-17	18	2	0	24	<0.001	0.602 f
MPS-18	10	0	2	15	<0.001 f	1 f
MPS-19	0	39	5	36	0.055 f	<0.001
MPS-20	16	1	2	18	<0.001	1 f

Footnote: *P-Sal** mutants were crossed to $In(1)w^{m4}$ to test for the presence of *Su(var)* mutations. I used cross-tabulation descriptive analysis to compare numbers statistically (SPSS 12.0). To do this I used *Chi square* test, and when expected values were less than five, I used two-sided Fisher's exact test, as mentioned by "f" in the table. *P-values* less than 0.05 were considered significant.

Testing the *P-Sal** mutations for an antimorphic effect on the *P-Sal* phenotype

An antimorphic allele of *P-Sal** would suppress the PDS phenotype of a parental *P-Sal* allele and consequently result in more eye pigment present in the eye. To test this, I crossed *y w; +; Sb P-Sal*/ TM6B Tb Hu; +* males from each mutant to *y w, +; P-Sal; Pci* virgin females. In the progeny of each cross, I scored flies based on their eye color phenotype (w^+ vs. w^{var}) and their chromosome 3 marker phenotype (Sb vs. Tb Hu). Since *P-Sal* induces such a strong PDS (most of *y w; +; P-Sal/ +; Pci/ +* flies are essentially w^- in phenotype), I planned the assay so that the observation of any w^{var} phenotype in the progeny of this cross would be considered PDS suppression (Figure 4-4).

As expected, the control cross between parental *Sb P-Sal* (*y w; +; Sb P-Sal/ TM6B Tb Hu;+)* males to *y w, +; P-Sal; Pci* virgin females did not show suppression of PDS in the progeny. I then compared this outcome to the results of each cross between a *Sb P-Sal** mutant and *y w, +; P-Sal; Pci*. Any difference between *Sb P-Sal** mutants and parental *Sb P-Sal* on their effect on the *P-Sal* chromosome phenotype could be attributed to an antimorphic effect. Four mutants, MPS-6, MPS-13, MPS-16, and MPS-17 showed an increase in the frequency of a w^{var} phenotype amongst progeny of the above mentioned cross (*Chi square* test, *p-value* < 0.001). This indicates that these mutants are able to suppress PDS induced by *P-Sal*, *i.e.* have an antimorphic effect. The remaining *Sb P-Sal** mutants were similar to the parental *Sb P-Sal* in having no suppressing effect on PDS phenotype (Table 4-4).

I compared the frequency of w^{var} flies among the Sb progeny of each cross (carriers of *Sb P-Sal** chromosome) versus the frequency of w^{var} flies among the Tb Hu progeny (carriers of *TM6B Tb Hu* balancer chromosome). Again, MPS-6, MPS-13, MPS-16, and MPS-17, showed a higher frequency of a w^{var} phenotype among Sb progeny than among Tb Hu progeny (*Chi square* test, *p-value* < 0.001). This confirms the previous conclusion that these four mutated chromosomes suppress PDS induced by parental *P-Sal*. Thus, all four mutants clearly showed an antimorphic effect on PDS induction by *P-Sal* (Table 4-4).

It seems that, in addition to losing their silencing effect on *Pci*, MPS-6, MPS-13, MPS-16, and MPS-17 have gained an antimorphic function. This effect is most likely due to a mutation in the *P-Sal* locus. It also can be explained by a second site modifier mutation that does not suppress hPEV of $In(1)w^{m4}$, but is able to suppress PDS. If this is the case, this dominant mutation suppresses PDS both in *cis* (*P-Sal** mutant) and *trans* (the parental *P-Sal*). Based on

mapping results this mutation has to be located within 3 cM from *P-Sal*. This mutation also has to be a gain-of-function mutation, because as I showed in the *Df(3)* experiment there is no locus within 3 cM of *P-Sal* that its loss-of-function suppresses PDS.

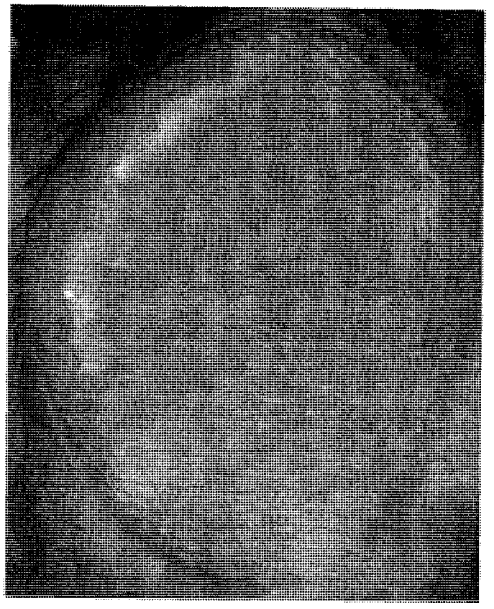


Figure 4-4) Typical phenotype showing silencing of *Pci* by *P-Sal* (*y w; +; P-Sal; Pci*). *P-Sal* typically produces an entirely white eye color in a *Pci* Background. I considered any red eye color as suppression of this effect.

Table 4-4) Testing the *P-Sal** mutants for antimorphic effect on parental *P-Sal.*

Mutant Name	Antimorphic Effect	Phenotype Frequency				*Chi square* test, *p-value*	
		Sb		Tb Hu			
		w^-	w^{var}	w^-	w^{var}	Comparison with *TM6b Tb Hu* balancer	Comparison with parental *P-Sal*
P-Sal		38	6	47	13	0.925	N/A
MPS-1		66	0	69	0	No test (no w^{var})	0.003 f
MPS-2		83	0	76	0	No test (no w^{var})	0.001 f
MPS-3		29	8	31	2	0.63 f	0.344
MPS-4		84	0	70	2	0.211 f	0.001 f
MPS-5-1		51	0	63	0	No test (no w^{var})	0.008 f
MPS-5-2		50	3	40	2	0.61 f	0.292 f
MPS-6	+	20	28	58	2	<0.001	<0.001
MPS-7		47	1	44	6	0.062 f	0.051 f
MPS-8		76	0	78	0	No test (no w^{var})	0.002 f
MPS-9		57	0	50	3	0.109 f	0.006 f
MPS-10		39	8	20	26	<0.001	0.655
MPS-11-1		74	4	72	7	0.36	0.166 f
MPS-11-2		59	0	66	0	No test (no w^{var})	0.005 f
MPS-12		77	0	82	0	No test (no w^{var})	0.002 f
MPS-13	+	25	24	56	4	<0.001	<0.001
MPS-14		39	7	46	15	0.749	0.293
MPS-15		79	0	107	0	No test (no w^{var})	0.002 f
MPS-16	+	4	28	24	10	<0.001	<0.001
MPS-17	+	14	22	41	6	<0.001	<0.001
MPS-18		75	4	89	3	0.416 f	0.165 f
MPS-19		59	0	67	10	0.003 f	0.005 f
MPS-20		39	0	30	6	0.01 f	0.019 f

Footnote: *P-Sal** mutants were crossed to parental *P-Sal* in the search for any suppressing effect on PDS induced by parental *P-Sal.* I used cross-tabulation descriptive analysis to compare numbers statistically (SPSS 12.0). To do this I used *Chi square* test, and when expected values were less than five, I used one-sided Fisher's exact test, as mentioned by "f" in the table. *P-values* less than 0.05 were considered significant.

Effect of *P-Sal** mutants on *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$

P{hsp26-pt-T}ci$^{2\text{-}m1021.R}$ is a *P{hsp26-pt-T}* insert upstream of the *ci* gene. It is inserted 150 bp nearer to the *ci* than the *Pci* and its *mini-white* gene expression is also affected by the presence of *P-Sal* (Bushey, 2004). To test the effect of *P-Sal** mutants on *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, I crossed *y w; +; Sb P-Sal*/ TM6B Tb Hu; +* males from each mutant to *w; +; +; P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ virgin females. From the progeny of each cross I scored flies based on their eye color phenotype (w^+ vs. w^{var}) and their chromosome 3 marker phenotype (Sb vs. Tb Hu). I compared the frequency of w^+ and w^{var} in Sb progeny (*Sb P-Sal* carriers) with the frequency of w^+ and w^{var} in Tb Hu progeny (*TM6B Tb Hu* carriers) (Figure 4-5).

In the control cross between parental *y w; +; Sb P-Sal/ TM6B Tb Hu;+* males and *w; +; +; P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ virgin females, the frequency of variegated flies was significantly greater in Sb progeny than in Tb Hu progeny (*Chi square* test, *p-value*< 0.001) (Figure 4-5). This confirmed the previous report of a *P-Sal* silencing effect on *mini-white* gene expression in *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ (Bushey and Locke. 2004). However, when the *P-Sal** mutants were tested, none showed any w^{var} progeny like those seen in the *P-Sal* control (Table 4-5). In pair-wise comparisons to the results of progeny from the control cross, I found the control cross progeny had a significantly greater frequency of PDS phenotype (largest *Chi square* test *p-value*< 0.001, Table 4-5). Thus all the *P-Sal** mutants have lost their ability to silence the *mini-white* gene on both *Pci* and *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$.

For each cross, I also compared the frequency of w^{var} flies in the Sb progeny versus Tb Hu (balancer) progeny. This comparison did not find any statistically significant difference in any of 22 *P-Sal** (smallest *Chi square* test *p-value* was 0.173 which is larger than the significance threshold of 0.05, Table 4-5). This supports the conclusion that all *P-Sal** mutants have lost their ability to silence expression of *mini-white* gene in *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$.

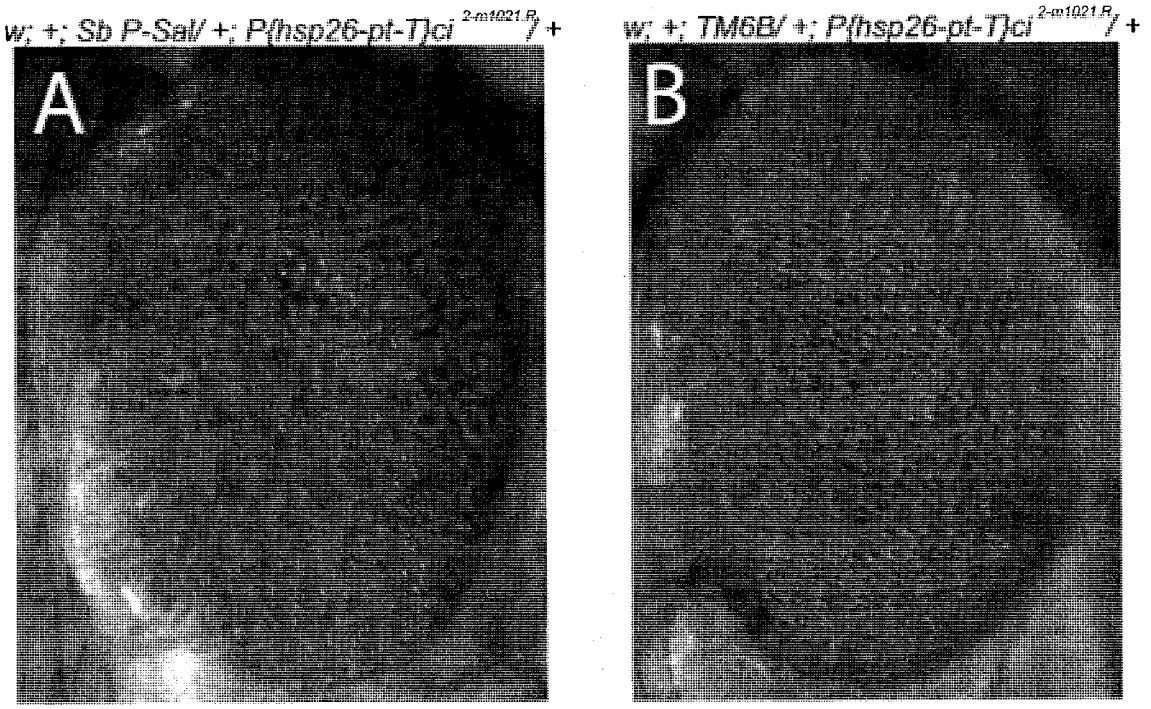


Figure 4-5) Partial silencing of the *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ transgene by *P-Sal*.
Panel A: *P-Sal* induces variegation of the *mini-white* gene expression in *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, as can be seen at posterior border of the eye (left side). Panel B: *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ in the absence of *P-Sal* (in *TM6B Tb Hu* carriers) produces a uniform red eye color. Genotypes of flies are mentioned above each panel.

Table 4-5) Effect of *P-Sal** mutants on *P{hsp26-pt-T}ci* $^{2\text{-}m1021.R}$ eye color.

Mutant Name	PDS Effect on *P{hsp26-pt-T}ci* $^{2\text{-}m1021.R}$	Phenotype Frequency				*Chi square* test, *p-value*	
		Sb		Tb Hu		Comparison with *TM6b Tb Hu* balancer	Comparison with parental *P-Sal*
		w^{var}	w^{+}	w^{var}	w^{+}		
P-Sal	+	13	7	0	37	<0.001 f	N/A
MPS-1	-	0	55	0	59	No test (no w^{var})	<0.001 f
MPS-2	-	0	29	0	32	No test (no w^{var})	<0.001
MPS-3	-	0	40	0	37	No test (no w^{var})	<0.001 f
MPS-4	-	0	12	0	14	No test (no w^{var})	<0.001 f
MPS-5-1	-	0	31	0	33	No test (no w^{var})	<0.001
MPS-5-2	-	0	41	0	38	No test (no w^{var})	<0.001 f
MPS-6	-	0	17	0	14	No test (no w^{var})	<0.001
MPS-7	-	0	40	0	26	No test (no w^{var})	<0.001 f
MPS-8	-	0	27	0	25	No test (no w^{var})	<0.001
MPS-9	-	0	38	0	42	No test (no w^{var})	<0.001 f
MPS-10	-	0	35	0	38	No test (no w^{var})	<0.001 f
MPS-11-1	-	0	72	0	69	No test (no w^{var})	<0.001 f
MPS-11-2	-	0	45	0	39	No test (no w^{var})	<0.001
MPS-12	-	0	28	0	30	No test (no w^{var})	<0.001 f
MPS-13	-	0	30	0	25	No test (no w^{var})	<0.001
MPS-14	-	0	63	0	60	No test (no w^{var})	<0.001 f
MPS-15	-	0	33	0	31	No test (no w^{var})	<0.001 f
MPS-16	-	0	18	0	30	No test (no w^{var})	<0.001
MPS-17	-	0	34	0	38	No test (no w^{var})	<0.001 f
MPS-18	-	0	59	0	88	No test (no w^{var})	<0.001 f
MPS-19	-	2	36	0	47	0.173 f	<0.001
MPS-20	-	0	39	0	33	No test (no w^{var})	<0.001 f

Footnote: *P{hsp26-pt-T}ci* $^{2\text{-}m1021.R}$ is a *P* construct in chromosome 4 that is sensitive to PDS induced by *P-Sal*. *P-Sal** mutants were crossed to *P{hsp26-pt-T}ci* $^{2\text{-}m1021.R}$ to test for their ability to induce PDS. I used cross-tabulation descriptive analysis to compare numbers statistically (SPSS 12.0). To do this I used *Chi square* test, and when expected values were less than five, I used one-sided Fisher's exact test, as mentioned by "f" in the table. *P-values* less than 0.05 were considered significant.

Effect of *P-Sal** mutants on *P{hsp26-pt-T}39C-12*

The *P{hsp26-pt-T}39C-12* insertion is a *P{hsp26-pt-T}* construct inserted at 102B and induces a variegated (w^{var}) eye color by itself (Sun et al, 2000). Bushey (2004) showed that, unexpectedly, *P-Sal* inhibits the silencing of *P{hsp26-pt-T}39C-12* to produce an essentially w^+ phenotype in *y w; +; P-Sal/ +; P{hsp26-pt-T}39C-12/ +* flies. To examine this effect of the *P-Sal** mutants, I crossed *y w; +; Sb P-Sal*/ TM6B Tb Hu; +* males from each mutant to *y w, +; +; P{hsp26-pt-T}39C-12* virgin females. I compared the frequency of w^+ vs. w^{var} progeny in the Sb versus Tb Hu (balancer) background as an indicator for the retention of *P-Sal* activity in the *P-Sal** mutants (Table 4-6, Figure 4-6).

In the control cross between parental *y w; +; Sb P-Sal/ TM6B Tb Hu; +* males and *y w, +; +; P{hsp26-pt-T}39C-12* virgin females, there was a significant increase in the frequency of the w^+ phenotype among Sb progeny versus Tb Hu (balancer) progeny (*Chi square* test, *p-value*<0.001). All the Sb progeny are w^+, while all the Tb Hu progeny are w^{var}. This increase of suppressed silencing was expected based on the results of Bushey (2004).

I also compared the frequency of a w^+ flies among the Sb progeny of each of the *P-Sal** mutants versus the Sb progeny from the control cross. Most of the mutants (except MPS-6, MPS-8, MPS-13, and MPS -19) had significantly fewer w^+ progeny than the parental *P-Sal* (largest *Chi square* test *p-value* was less than 0.001, Table 4-6). These four exceptions showed no significant difference from parental *P-Sal* in affecting *P{hsp26-pt-T}39C-12* (Table4-6), indicating that they retained the ability to suppress the variegation of this insert. In addition to these 4, there are five other *P-Sal** mutants (MPS-2, MPS-7, MPS-15, MPS-16, and MPS-20) that showed a significantly higher frequency of w^+ among the Sb versus Tb Hu (balancer) progeny of the same cross (largest *Chi square* test *p-value* was 0.026 which is smaller than threshold (0.05), Table 4-6). In other words, although these nine mutants have lost their ability to induce PDS (both *Pci* and *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$), they still have retained their ability to suppress variegation in *P{hsp26-pt-T}39C-12* as the parental *P-Sal* did.

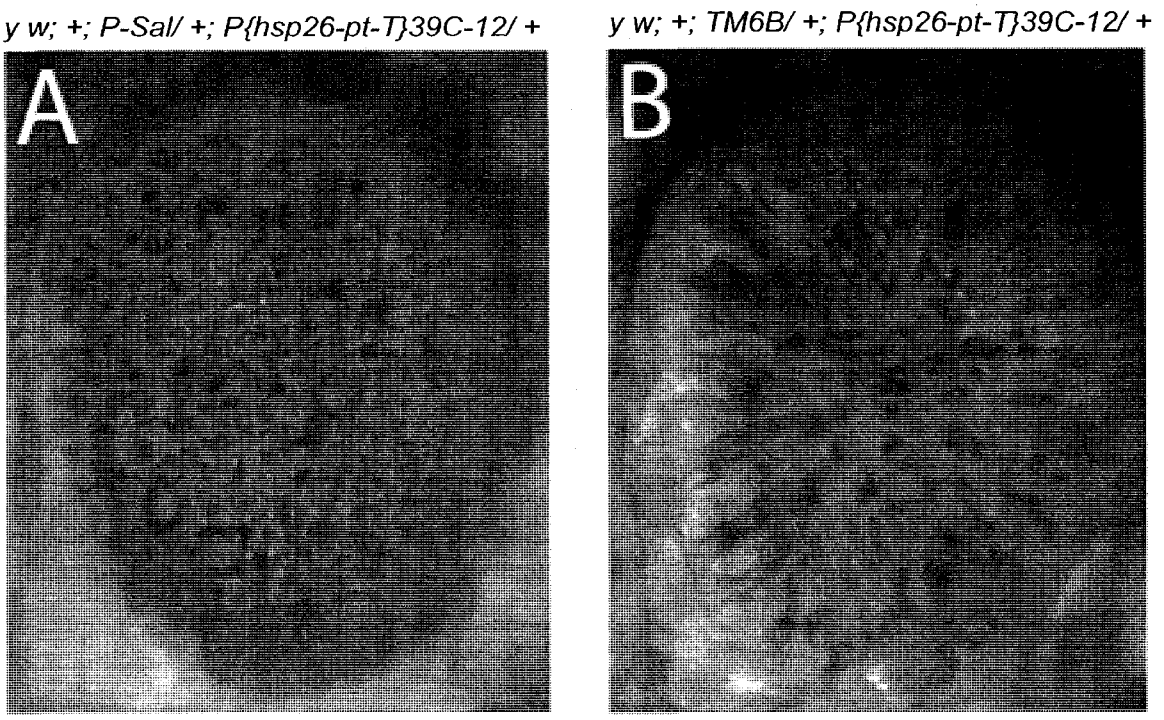


Figure 4-6) *P-Sal* suppresses variegation of *P{hsp26-pt-T}39C-12.*

Panel A shows *P-Sal* suppressing the variegation of *P{hsp26-pt-T}39C-12* (shown in Panel B). Genotypes of flies are mentioned above each panel.

Table 4-6) Effect of *P-Sal** mutants on *P{hsp26-pt-T}39C-12* eye color.

Mutant Name	Suppression of variegation of *P{hsp26-pt-T}39C-12*	Phenotype Frequency				*Chi square* test, *p-value*	
		Sb		Tb Hu		Comparison with *TM6b Tb Hu* balancer	Comparison with parental *P-Sal*
		w^{var}	w^{+}	w^{var}	w^{+}		
P-Sal	+	0	63	65	0	<0.001	N/A
MPS-1	-	84	1	61	1	0.667 f	<0.001
MPS-2	weak	42	28	72	0	<0.001	<0.001
MPS-3	-	86	1	50	1	0.58 f	<0.001
MPS-4	-	82	0	123	0	No test (no w^{var})	<0.001
MPS-5-1	-	38	1	154	0	0.202 f	<0.001
MPS-5-2	-	64	1	81	0	0.445 f	<0.001
MPS-6	+	1	64	73	1	<0.001	0.5 f
MPS-7	weak	36	32	88	0	<0.001	<0.001
MPS-8	+	6	104	103	0	<0.001	0.059 f
MPS-9	-	88	0	126	0	No test (no w^{var})	<0.001
MPS-10	-	82	0	85	0	No test (no w^{var})	<0.001
MPS-11-1	-	65	0	105	0	No test (no w^{var})	<0.001
MPS-11-2	-	71	0	93	0	No test (no w^{var})	<0.001
MPS-12	-	70	0	63	0	No test (no w^{var})	<0.001
MPS-13	+	0	112	97	0	<0.001	No test (no w^{var})
MPS-14	-	74	0	94	0	No test (no w^{var})	<0.001
MPS-15	weak	19	45	94	0	<0.001	<0.001
MPS-16	weak	48	3	119	0	0.026 f	<0.001
MPS-17	-	59	0	71	0	No test (no w^{var})	<0.001
MPS-18	-	84	0	73	0	No test (no w^{var})	<0.001
MPS-19	+	0	59	88	0	<0.001	No test (no w^{var})
MPS-20	weak	85	11	77	0	0.001 f	<0.001

Footnote: *P{hsp26-pt-T}39C-12* is a *P* construct that is sensitive to the presence of *P-Sal*. Its phenotype is w^{var} by itself and *P-Sal* suppresses its variegation. To test the ability of the *P-Sal** mutants to suppress this variegation I crossed them to *P{hsp26-pt-T}39C-12*. I used cross-tabulation descriptive analysis to compare numbers statistically (SPSS 12.0). To do this I used *Chi square* test, and when expected values were less than five, I used one-sided Fisher's exact test, as mentioned by "f" in the table. *P-values* less than 0.05 were considered significant.

Effect of *P-Sal** mutants on *P{lacW}ci^{E1}* (*E1*)

P{lacW}ci^{E1} (*E1*) is derived from *Pci* in that a *gypsy* transposon has been inserted upstream of the *P{lacW}* transgene insert (Bushey and Locke, 2004). *E1* induces a w^{var} eye color phenotype that differs from the w^{+} eye phenotype of *Pci*. The presence of other *P* elements will enhance the w^{var} phenotype (more silencing of the *mini-white* gene expression). For example, the presence of *P-Sal* with *E1* converts the w^{var} phenotype to an essentially w^{-} eye phenotype (Bushey and Locke, 2004). Since the *P-Sal** mutants lost the ability to induce variegation in PDS, I tested them for their ability to enhance variegation in *E1*. For each mutant, I crossed *y w; +; Sb P-Sal*/ TM6B Tb Hu;* + males to *y w, +; +; E1* virgin females and examined the progeny eyes for w^{-} or w^{var} color.

In the control cross between parental *y w; +; Sb P-Sal/ TM6B Tb Hu;* + males and *y w, +; +; E1* virgin females, the frequency of a w^{-} phenotype was significantly higher among Sb progeny than Tb Hu (balancer) progeny (*Chi square* test, *p-value*<0.001) (Figure 4-7). When each *P-Sal** mutant was crossed to *E1*, all except MPS-19 showed a w^{var} phenotype, and not the w^{-} phenotype of *P-Sal* (*Chi square* test, *p-value*< 0.001, Table 4-7), indicating the loss of any *P-Sal* function that enhances *E1* variegation.

For each cross, I also compared the frequency of w^{-} flies among the Sb progeny versus frequency of w^{-} flies among Tb Hu progeny. Twenty-one out of 22 *P-Sal** mutants, did not show any statistically significant difference in this comparison (Table 4-7). However, in the cross between MPS-19 and *y w, +; +; E1*, the frequency of w^{-} phenotype was significantly higher in Sb progeny than Tb Hu progeny (*Chi square* test, *p-value*<0.001). The MPS-19 modification result of the E1 experiment is consistent with my previous findings for MPS-19 being a hypomorphic allele and possibility containing an *E(var)* mutation.

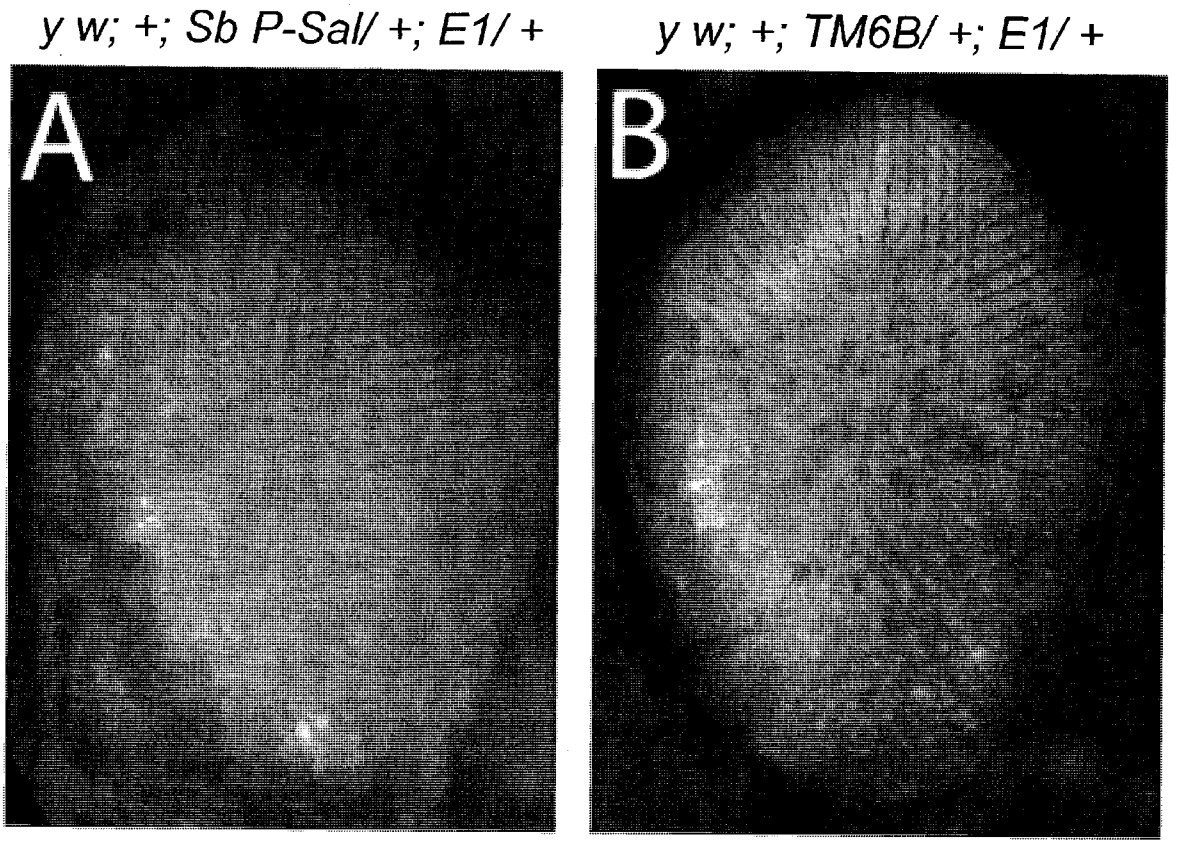


Figure 4-7) *P-Sal* enhances the variegation of *E1*.

The presence of *P-Sal* enhances the variegation to a white eye color (A), compared to the *TM6B Tb Hu* control, which has a typical variegated phenotype (B). Genotypes of flies are shown above each panel. *E1* is a *Pci* insert that has a *gypsy* element inserted at its upstream and induces a w^{var} eye color phenotype.

Table 4-7) Effect of *P-Sal** mutants on *E1*.

Mutant Name	PDS Effect on *E1*	Phenotype Frequency				*Chi square* test, *p-value*	
		Sb		Tb Hu		Comparison with *TM6b Tb Hu* balancer	Comparison with parental *P-Sal*
		w^{var}	w^{-}	w^{var}	w^{-}		
P-Sal	+	0	49	47	0	<0.001	N/A
MPS-1	-	25	0	18	0	No test (no w^{var})	<0.001
MPS-2	-	40	0	33	0	No test (no w^{var})	<0.001
MPS-3	-	43	0	36	0	No test (no w^{var})	<0.001
MPS-4	-	56	0	48	0	No test (no w^{var})	<0.001
MPS-5-1	-	66	0	58	0	No test (no w^{var})	<0.001
MPS-5-2	-	34	0	27	0	No test (no w^{var})	<0.001
MPS-6	-	37	0	23	0	No test (no w^{var})	<0.001
MPS-7	-	38	0	30	0	No test (no w^{var})	<0.001
MPS-8	-	53	0	47	0	No test (no w^{var})	<0.001
MPS-9	-	38	0	42	0	No test (no w^{var})	<0.001
MPS-10	-	73	0	42	0	No test (no w^{var})	<0.001
MPS-11-1	-	29	0	32	0	No test (no w^{var})	<0.001
MPS-11-2	-	74	0	75	0	No test (no w^{var})	<0.001
MPS-12	-	51	0	53	0	No test (no w^{var})	<0.001
MPS-13	-	89	0	58	0	No test (no w^{var})	<0.001
MPS-14	-	43	0	38	0	No test (no w^{var})	<0.001
MPS-15	-	35	0	20	0	No test (no w^{var})	<0.001
MPS-16	-	47	0	45	0	No test (no w^{var})	<0.001
MPS-17	-	38	0	35	0	No test (no w^{var})	<0.001
MPS-18	-	44	0	35	0	No test (no w^{var})	<0.001
MPS-19	+	0	56	59	0	<0.001	No test (no w^{var})
MPS-20	-	35	0	29	0	No test (no w^{var})	<0.001

Footnote: *E1* is a *Pci* construct with a *gypsy* transposon at its upstream. *E1* induces a w^{var} eye color phenotype that is sensitive to PDS. *P-Sal** mutants were tested for their ability to induce PDS in *E1*. I used cross-tabulation descriptive analysis to compare numbers statistically (SPSS 12.0). To do this I used *Chi square* test. *P-values* less than 0.05 were considered significant.

Sequencing the mutant *P-Sal* alleles

All *P-Sal** mutants were sequenced for the whole coding sequence of the type I *P* repressor in addition to some of the adjacent sequences. This is from about 50 bp upstream of the start codon (*i.e.* around base pair 100; start codon is at base pair 153) through about 100 bp downstream of the stop codon (*i.e.* base pair 2100; *P-Sal* stop codon is at base pair 1994) (Table 4-8). Five mutants (MPS-1, MPS-2, MPS-8, MPS-19, and MPS-20) were also sequenced from 23 to 100 and three mutants (MPS-1, MPS-8, and MPS-20) were also sequenced from base pair 2100 through 2830 on the standard *P* element sequence. The results are summarized in Table 4-8. Most of the *P-Sal** mutants (20 out of 22) have single base substitutions. MPS-8 has two changes and MPS-20 has none (Table 4-8 & Figure 4-8). MPS-8 has a nonsense mutation at position 794 (at codon 195) and a missense mutation at 1089. MPS-5-1 and MPS-5-2 have the same change: a mutation at base pair 500, which is the acceptor splicing site of intron 0-1. Both MPS-11-1 and MPS-11-2 have the same mutation at base pair 1100. There were three nonsense mutations; MPS-1 at base pair 238 (at codon 29), MPS-8 at base pair 794 (at codon 195), and MPS-10 at base pair 1097 (at codon 296). There are four mutations in exon 0 of *P* element gene, which encodes the DNA binding domain and nuclear localization signal. All four mutations are in the DNA binding domain and one of them, MPS-4, is in the nuclear localization signal of the domain (Figure 4-8). Two mutations affect the acceptor splicing site at the beginning of exon 1. Despite being an expected target, none of the mutations altered the LLLL leucine zipper domain at the beginning of exon 1. I found three mutations, MPS-6, MPS-7, and MPS-8, clustered in a 130 bp area in the middle of exon 1 that is present in *KP* element, but there is no known domain in this area. Moreover, near the end of exon 1 there is a hot spot where four mutations, MPS-10, MPS-11-1, MPS-11-2, and MPS-12, occurred within a 10 bp area. Rio (1990) predicted that this area encodes a LLVLL leucine zipper. MPS-10 is a nonsense mutation and MPS-11 changes a key position, the central valine (Figure 4-8).

Exon 2 had six mutations (MPS-13 through MPS-18) scattered evenly all over its length. Rio (1990) found a weak homology at the part of the protein that is encoded by the end of exon 1 and beginning of exon 2 with a bacterial helix-turn-helix (HTH) domain, a prokaryotic DNA binding domain. I found one mutation, MPS-13, near the end of this domain. Rio (1990) also predicted an ILLQL leucine zipper near the end of exon 2, which is where MPS-18 is mutated, in one of leucine zipper non-key positions.

MPS-19 was mutated at the end of the *P-Sal* coding sequence, which is considered part of an intron in a *P* element germ-line specific splicing pattern (Figure 4-8). This last mutation has a

moderate hypomorph phenotype, while the rest of the mutants are either amorphs or very strong hypomorphs. Results from the *In(1)w^{m4}* experiment support the presence of an *E(var)* mutation on this mutant.

Nine of 22 *P-Sal** mutants (MPS-1 through MPS-8) were mutated in the sequences common to both the *P-Sal* and *KP* elements. MPS-8 has a nonsense mutation at base pair 794 (codon 195) which is just 14 base pairs upstream of *KP* element deletion break point. This makes the MPS-8 encoded polypeptide just 14 amino acids shorter than the *KP* encoded protein. Considering that the last eight amino acids of the 208 amino acids of the KP protein come from an out of frame translation of part of exon 3, MPS-8 encoded polypeptide is very similar to the KP protein that is encoded by *P* element standard exons. However, the MPS-8 encoded polypeptide has lost the ability of the KP protein to induce PDS, at least at its current genomic location.

After the first round of sequencing, MPS-20 lacked any mutation in its sequence from base pair 23 up to base pair 2163. Sequencing it again from base pair 23 through 2819 also failed to identify any difference between parental *P-Sal* and MPS-20 sequences. To test the possibility of a second-site modifier that is located very near the *P-Sal* construct, I tried again to segregate the possible modifier from *P-Sal*. I crossed *y w; +; MPS-20/ TM6B Tb Hu;* + males to *w; dp; e;* + virgin females and selected *w; dp/ +; MPS-20/ e;* + virgin females in the progeny and backcrossed them to *w; dp; e; Pci* males. After scoring more than 1400 potential recombinant flies, I did not find any PDS phenotype among the progeny. This narrows down the location of any potential second-site modifier to 0.27 cM of *P-Sal*. The *Df(3)* experiment had already rejected the possibility of a dominant *Su(var)* loss-of-function mutation around *P-Sal*. Therefore, the most parsimonious explanation is that this is a mutation in the regulatory sequences that controls *P-Sal* expression, possibly in the regulatory region of the *Pak3* gene.

Table 4-8) Sequencing results of the *P-Sal** mutants.

Mutant Name	Sequenced	DNA Change	Amino Acid Change
MPS-1	23-2831	238 G>A	29 Trp>STOP
MPS-2	23-2093	240 G>A	30 Glu>Lys
MPS-3	92-2103	283 G>A	44 Cys>Tyr
MPS-4	73-2101	346 G>A	65 Arg>Lys
MPS-5-1	111-2097	500 G>A	SPLICING
MPS-5-2	99-2111	500 G>A	SPLICING
MPS-6	92-2103	665 G>A	152 Gly>Glu
MPS-7	108-2137	712 C>T	168 Leu>Phe
MPS-8	23-2796	794 G>A 1089A>T	195 Trp>STOP
MPS-9	68-2101	821 G>A	204 Gly>Glu
MPS-10	111-2104	1097 T>A	296 Leu>STOP
MPS-11-1	78-2093	1100 T>A	297 Val>Glu
MPS-11-2	140-2100	1100 T>A	297 Val>Glu
MPS-12	73-2103	1106 C>T	299 Ala>Val
MPS-13	73-2103	1228 C>T	322 Thr>Ile
MPS-14	73-2100	1352 A>T	363 Leu>Phe
MPS-15	73-2093	1473 G>A	404 Ala>Thr
MPS-16	136-2093	1645 C>T	461 Pro>Leu
MPS-17	73-2137	1750 G>T	496 Gly>Val
MPS-18	108-2137	1828 C>T	522 Thr>Ile
MPS-19	23-2137	1963 A>T	567 Lys>Ile
MPS-20	23-2819	None	

Footnote: Amino acid changes were predicted based on the published splicing pattern (O'Hare and Rubin, 1983).

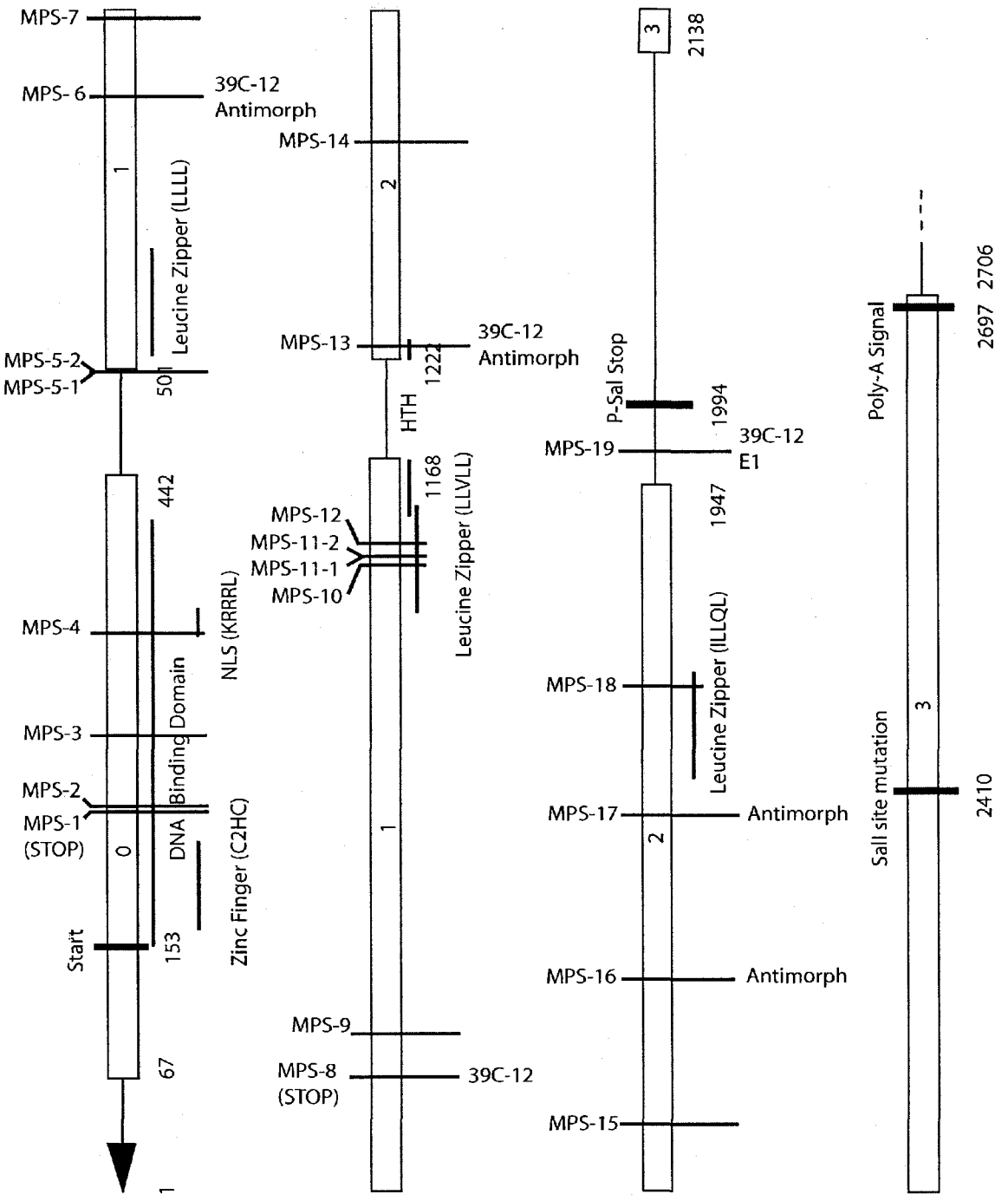


Figure 4-8) Location of *P-Sal mutations.**

Numbers refer to the location of the mutations in the complete *P* element sequence. The labels "39C-12" and "E1" refer to mutants that maintain their enhancing of *P{hsp26-pt-T}39C-12* and silencing of *E1* abilities respectively. The label "Antimorph" refers to mutants with antimorphic effects. Start: translation start codon, NLS: nuclear localization signal, HTH: helix-turn-helix domain, *P-Sal* Stop: translation stop codon for the *P-Sal* protein (type I repressor protein). Arrowhead indicates 31 inverted repeat. Numbers indicates location of each splicing site or domain on the complete *P* element sequence.

Note: Hodgetts and O'keefe (2001) reported that transcription starts at base pair 67.

Protein prediction programs

Information that was used for classification of the *P-Sal** mutants is recorded in the following appendices in detail. Appendix I contains different amino acid classifications, appendix II shows analysis of the parental type I repressor protein, and appendix III shows the detailed predictions of protein prediction programs for the mutants.

Table 4-1 shows the predicted amino acid changes of the *P-Sal** mutants. All mutations except MPS-4 caused changes in one or more of the following major amino acid characteristics. Out of 16 missense mutations, nine caused changes in amino acid polarity, seven caused changes in amino acid water affinity, and 14 caused changes in amino acid tendency for a different secondary structure.

To predict changes in the mutant type I repressor protein, I used seven computer programs: PROSITE, NetPhos 2.0, DomPred, Porter protein predict, PSIPRED, SABLE-2, and SIFT (Table 6-2). I analysed 17 out of 22 mutant amino acid sequences by these programs. MPS-1 encodes a 28 aa polypeptide that is too short for analysis by these programs. MPS-5-1 and MPS-5-2 are mutated at a splicing site and the new splicing pattern and amino acid sequence are unknown. Therefore, they can not be analyzed by above mentioned programs. MPS-20 did not show any change in its coding sequence, therefore, its amino acid sequence is the same as parental *P-Sal*. MPS-11-1 and MPS-11-2 have the same mutation so I used one of them for analysis.

PROSITE motif search (Gattiker *et al*, 2002) is an online tool of ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server from the Swiss Institute of Bioinformatics (SIB) that scans protein sequences to find patterns, profiles and motifs stored in its database. In the type I repressor protein it found seven N-glycosylation sites, one cAMP- and cGMP-dependent protein kinase phosphorylation site, eight protein kinase C phosphorylation sites, seven casein kinase II phosphorylation sites, nine N-myristoylation sites, one amidation site, and one leucine zipper pattern (Appendix II). Since none of the predicted N-glycosylation sites or N-myristoylation sites are located either at the N terminus of the protein or downstream of a protease site their predicted presence or absence is irrelevant to the phenotype. PROSITE predicted the loss of the protein kinase C phosphorylation site at position 522 in MPS-18 (Appendix III).

NetPhos 2.0 (Blom *et al*, 1999) is an online tool from Technical University of Denmark that predicts serine, threonine, and tyrosine phosphorylation sites in a protein. It found 22 phosphorylation sites in the type I repressor protein (Appendix II). Analysis of the *P-Sal**

mutations predicted a loss of phosphorylation site at position 522 for MPS-18, and a gain of a phosphorylation site at amino acid 404 for MPS-15 (Appendix III).

DomPred (Marsden, 2002) is a protein domain prediction server from University College London that predicts putative protein domains. In the type I repressor protein, DomPred predicted a domain around amino acid 160 and another domain around amino acid 320 (Appendix II). It predicted a change in pattern of domains in MPS-6 and MPS-9 (Appendix III).

Porter (Pollastri and McLysaght, 2005) is a bidirectional neural network–based protein secondary structure prediction program from University College Dublin. It is a modified version of SSpro server. Porter did not predict any change in amorphic mutants MPS-2, MPS-4, MPS-7 MPS-13, MPS-15, and MPS-18. All of the Porter predicted changes before amino acid 500 are in truncated polypeptides of MPS-8 and MPS-10. For all other mutants it predicted just two changes, 518 coil to strand change and 524 coil to helix. Both of these changes are located within 497-525 leucine zipper. MPS-14, MPS- 16, MPS-17, and MPS-19 showed exactly the same pattern on Porter prediction (Appendix III).

PSIPRED (Jones, 1999) is another tool in the protein domain prediction server from University College London. It predicts secondary structure by neural network analysis on output obtained from PSI-BLAST (Position Specific Iterated - BLAST). It could not find any change in protein secondary structure of amorphic mutants MPS-6 and MPS-12. It predicted exactly the same changes for MPS-7, MPS-9, and MPS-11. In addition to nonsense mutations that have been predicted to have several impacts on all domains of protein structure, PSIPRED predicted a change in the zinc finger domain in MPS-2 and MPS-3, a change in nuclear localization signal in MPS-2, MPS-3, and MPS-4, changes in the 100-122 leucine zipper domain in MPS-2, MPS-3, MPS-4, MPS-7, MPS-9, and MPS-11, and changes in 497-525 leucine zipper domain for MPS-16, MPS-17 and MPS-18. The most common predicted change by this program was a change from coil to helix at amino acid 427 that had been predicted for eight of the mutants (Appendix III).

SABLE-2 (Adamczak *et al*, 2005) is a protein structure prediction server in the Cincinnati Children's Hospital Medical Center that uses relative solvent accessibilities of amino acid residues in addition to evolutionary profiles to predict the overall packing and secondary structure of the protein. SABLE-2 was not able to find any change in MPS-6, MPS-11, and MPS-15. However, it predicted some changes in zinc finger domain for MPS-2, MPS-3, and MPS-7. It also predicted changes in the nuclear localization signal in MPS-3, MPS-4, and MPS-7. Based on SABLE-2 predictions 100-122 leucine zipper have changed in amorphic mutants MPS-7, MPS-

12, MPS-13, MPS-17, and MPS-18. The 497-525 leucine zipper domain has changed in MPS-16, MPS-17 and MPS-18. There are several common changes that have been predicted for four mutants or more but the most common ones are helix to coil changes at amino acid 170 and amino acids 220-221(Appendix III).

SIFT (Ng and Henikoff, 2001) is supported by the Fred Hutchinson Cancer Research Center and predicts the effects of an amino acid substitution on polypeptide function based on physical properties and homology between amino acids. SIFT predicted that the mutation would be tolerated in amorphic mutants MPS-4, MPS-6, and MPS-13, and hypomorphic MPS-19 (Appendix III).

In general, all of the mutants have some changes predicted by at least one of the above mentioned programs (Appendix IV). Table 6-2 shows the important changes that could lead to the observed phenotype for each mutant.

Summary of *P-Sal** mutations

In summary, 12 out of 22 *P-Sal** mutants failed to retain any of the *P-Sal* original function(s) in any of the tests. In other words, they lost all functions. Twenty-one out of 22 mutants (all except MPS-19) have lost their ability to affect *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ and *E1* as well as *Pci*. Because these three constructs are all inserted just upstream of the *ci* gene, this common loss of function suggests that the P repressor protein affects all these inserts through a common pathway. MPS-19 was an exception; it had a hypomorphic *P-Sal* phenotype on *Pci*, an amorphic *P-Sal* phenotype on *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, and a parental *P-Sal* phenotype on *E1*.

Nine of the mutants (MPS-2, MPS-6, MPS-7, MPS-8, MPS-13, MPS-15, MPS-16, MPS-19, and MPS-20) that had lost their ability to affect *Pci*, still were able to silence variegation in *P{hsp26-pt-T}39C-12* stronger than the balancer chromosome. Four out of these nine *P-Sal** mutants (MPS-6, MPS-8, MPS-13, and MPS-19) did not show any significant difference from the parental *P-Sal* in their effect on *P{hsp26-pt-T}39C-12* while they are significantly different from parental *P-Sal* in their effect on *Pci*. This indicates the repressor effects on *Pci* and *P{hsp26-pt-T}39C*-12 are at least partially independent. Five more mutants (MPS-2, MPS-7, MPS-13, MPS-15, and MPS-16) are different from both the *P-Sal* and *TM6B Tb Hu* balancer chromosomes in their effect on the *P{hsp26-pt-T}39C-12*. In other words, these mutants are weaker than parental *P-Sal* in their effect on the *P{hsp26-pt-T}39C-12* but have not totally lost it. These nine mutations are scattered all along *P-Sal* and can not be grouped into one single domain.

Identifying the base pair changes by sequencing revealed two possible domains in the middle and near the end of exon 1 of *P-Sal*. Rio (1990) predicted a leucine zipper in the area near

the end of exon 1 and I found four mutants within 10 bp of it. However, none of the protein conformation prediction programs predicted any significant changes in this region for any of the mutants. These data suggest that this 10 bp area is either a novel domain and/or a mutational hot spot. MPS-8 encoded a truncated polypeptide that is just 14 amino acids shorter than the KP protein. It is unable to induce PDS on *Pci*, *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, or *E1*, but it is still able to suppress variegation of *P{hsp26-pt-T}39C-12*. MPS-20 was the only mutant that did not show any mutation within the *P* element sequence of *P-Sal* construct, but I have mapped its mutation within 0.27 cM from *P-Sal*. This is probably due to a mutation in a nearby enhancer sequence or other regulatory sequences that control *P-Sal* and probably *Pak3* expression. All together, I used several computer programs to predict changes in protein structure of my mutants, but none of them produced reliable results.

Chapter 5: P repressor protein presence in mutants

Introduction

The loss of the PDS phenotype in the *P-Sal** mutants could be due to mutation of one of the critical function domains, or instability and degradation of the protein. If the first scenario is true thus the protein, although lacking function would probably still be present inside the cell. This would provide a connection between structure and function. In the second situation, protein instability and degradation would decrease its cellular levels and a connection between structure and function cannot be determined. To differentiate between these two possibilities I could use antibodies to detect the presence of the P repressor protein in the mutants in a Western blot experiment. An antibody could also be useful to study change in protein modifications.

Since there was no antibody available from commercial or academic sources, I decided to make an antibody against the protein. A polyclonal antibody was preferred to a monoclonal one as a more sensitive test rather than a domain specific test. To make polyclonal antibody against P repressor protein I produced a full-length type I repressor protein in a prokaryotic expression system. I denatured and purified the P repressor protein and injected two rabbits with it. The recovered rabbit sera with anti-repressor antibodies were tested against the denatured P repressor protein and fly protein extracts.

cDNA cloning vectors

I tried to clone cDNA from both a *P-Sal* element (type I repressor) and a *KP* element (type II repressor). For cloning I used two different vectors: *pET21-b(+)* (Novagen) and *pRSET-B* (Invitrogen). Both vectors added a 6xHis tag to the expressed protein that facilitated its purification by Ni-NTA resin (Qiagen). *pET21-b(+)* provided both a T7 tag at the 5' end (protein N-terminus) and a 6xHis tag at the 3' end (protein C-terminus). The inserted gene was under control of a T7 promoter. In addition to its Amp^R gene, this plasmid has a *lac I* gene that helps to control the gene expression in *BL21(DE3)* host cells. These cells have a *T7 RNA polymerase gene* under control of a *lac operon* operator sequence that allows control of the gene expression by IPTG (Novagen manual) (Figure 5-1-a). *pRSET-B* added 6xHis tag to the 5' end of the insert (N-terminus of protein). Again the insert was under control of a T7 promoter. This plasmid also has an Amp^R gene (Qiagen manual) (Figure 5-1-b).

P-Sal cDNA production and RT-PCR

RNA was extracted from 50 *y w; +; Sb P-Sal/ TM6b TB Hu;* + flies following Invitrogen's TRIzol protocol. The extracted RNA was used as template for a RT reaction by iScript Select cDNA Synthesis kit (BioRad) and Oligo$(dT)_{20}$ primer. The resulting cDNA was amplified by PCR. I used a modified form of *Pyrococcus furiosus* high fidelity thermostable DNA Polymerase (Gary Ritzel, personal communication) to amplify the product.

A PCR reaction using P-BamHI (AATGAGAGGATCCGAAATATTGCAAATTTTGCTGC) introduces a *Bam*HI restriction site (underlined in the primer sequence) in place of the *P* element start codon and amplifies the *P* element from second codon. This facilitates the insertion of the PCR product into the *pET21-b(+)* in frame with its T7 tag and into *pRSET-B* in frame with its 6xHis tag. As reverse primer of the PCR reaction PSal-NotI (ACCAATTGCGGCCGCTGGAATTA CATTTTTGTTTACGC) allows amplification of *P-Sal* until the very last amino acid codon, and replaces the *P-Sal* stop codon by a *Not*I restriction site (underlined in the primer sequence) that is necessary for insertion of the PCR product into the *pET21-b(+)* plasmid in frame with the plasmid encoded 6xHis tag and its stop codon. For insertion into *pRSET-B*, another reverse primer, PSal-NcoI (ACTGAAACCATGGTGAAACATATAGC), was used. It adds an *Nco*I restriction site (underlined in the primer sequence) to the 3' end of the *P-Sal* cDNA downstream of its stop codon. This facilitates insertion of *P-Sal* cDNA into *pRSET-B*. Since there is no tag at the 3' end, disruption of *P-Sal* stop codon is not necessary for insertion to *pRSET-B*.

KP cDNA production and RT-PCR

RNA was extracted from the single *KP* insert strain KP(2)-beta (*y w; KP(2)-beta/ CyO; +; +*). Fifty flies were used and cDNA was prepared following the same steps that were mentioned for *P-Sal* cDNA production. P-BamHI was used as forward primer to disrupt the start codon and introduce a *Bam*HI restriction site at the 5' end of the RT-PCR product. To insert the *KP* cDNA into *pET21-b(+)* I used KP-NotI (AGCTAGAAGCGGCCGCCTTGTTTATCAACATCGACG) as the RT-PCR reverse primer. This primer disrupts the *KP* stop codon in the same way that PSal-NotI disrupts the *P-Sal* stop codon. And for insertion of *KP* cDNA into *pRSET-B*, I used KP-NcoI (TACTTAACCATGGAATATATTC) as the RT-PCR reverse primer that is equivalent of PSal-NcoI for *P-Sal*.

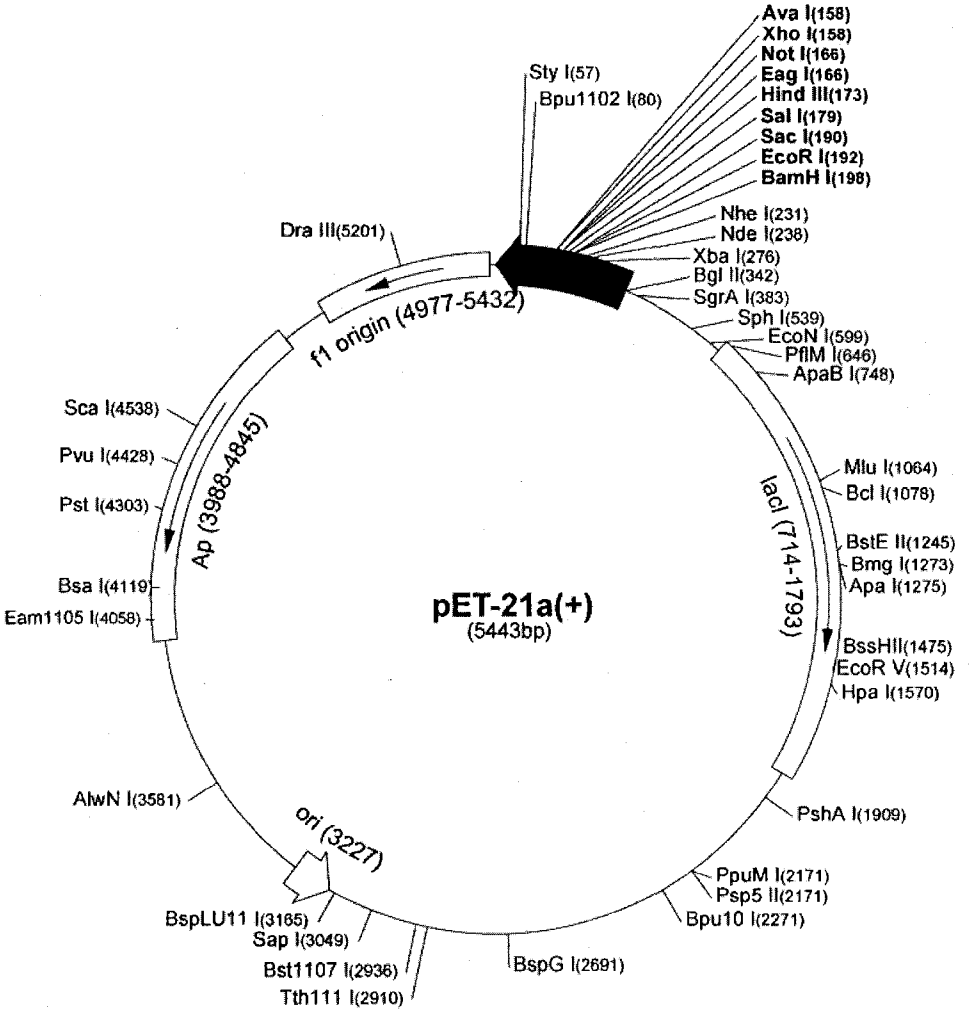
Test of cDNA

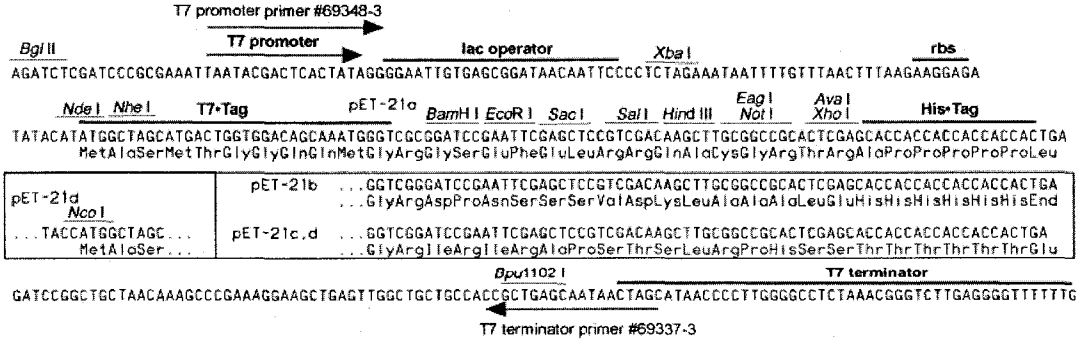
Each RT-PCR product was gel purified by Qiaquick kit (Qiagen) and tested by a KPF1/KPR2 nested PCR to make sure that the RT-PCR reaction had amplified cDNA and not the

genomic DNA (Figure 5-2). This set of primers amplified a 180 bp band from cDNA and a 239 bp band from genomic DNA in both *KP* and *P-Sal*. This nested PCR produced a ~180 bp band from both *P-Sal* RT-PCR reactions, *i.e.* it was produced from *P-Sal* cDNA, not genomic DNA. The nested PCR reaction for both *KP* RT-PCR reactions produced a ~239 bp band, *i.e.* RT-PCR did not amplify *KP* mRNA, probably due to lack of *KP* expression in adults. With these results, I discontinued further work with the *KP* and continued the experiment using *P-Sal* RT-PCR products.

Cloning

The *P-Sal* P-BamHI / PSal-NotI RT-PCR product and *pET21-b(+)* were doubly digested by *Bam*HI and *Not*I restriction enzymes (Invitrogen). The *P-Sal* P-BamHI / PSal-NcoI RT-PCR product and *pRSET-B* were doubly digested by *Bam*HI and *Nco*I restriction enzymes (Invitrogen). Both digestion reactions were at 37°C overnight. The digestion products were gel purified by Qiaquick kit (Qiagen). The purified appropriate segments were ligated together by T4 ligase (Invitrogen) at 16°C overnight. The resulting ligation products were used to transform *DH5α* competent cells (Invitrogen) following the producer's instructions. Transformed cells were spread on Ampicillin LB plates (Sambrook, 1989) and incubated at 37°C overnight. From each plasmid transformation, one colony that was tested positive for the presence of *P-Sal* cDNA by KPF1/KPR2 PCR (Figure 5-3), was incubated in 50 ml of liquid medium overnight at 37°C in a rotation shaker. One ml of the resulting culture was used to make a 15% glycerol stock and the rest was used for plasmid extraction following a small scale preparation of plasmid DNA protocol (Sambrook, 1989). To make sure no mutations were introduced during the process of RT-PCR and cloning and to make sure that each insert is in frame with its plasmid, I sequenced the inserts and more than 50 bp of adjacent plasmid sequences using extracted plasmids as template. Using the T7 promoter and T7 terminator primers (Novagen) I amplified the inserts and primed the sequencing reaction using T7 promoter, KPF1, 2033F, and 2632F primers. The sequencing was done by BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). No difference was found, using Genetools software (Biotools), between the inserts and published sequence (O'Hare and Rubin, 1983).

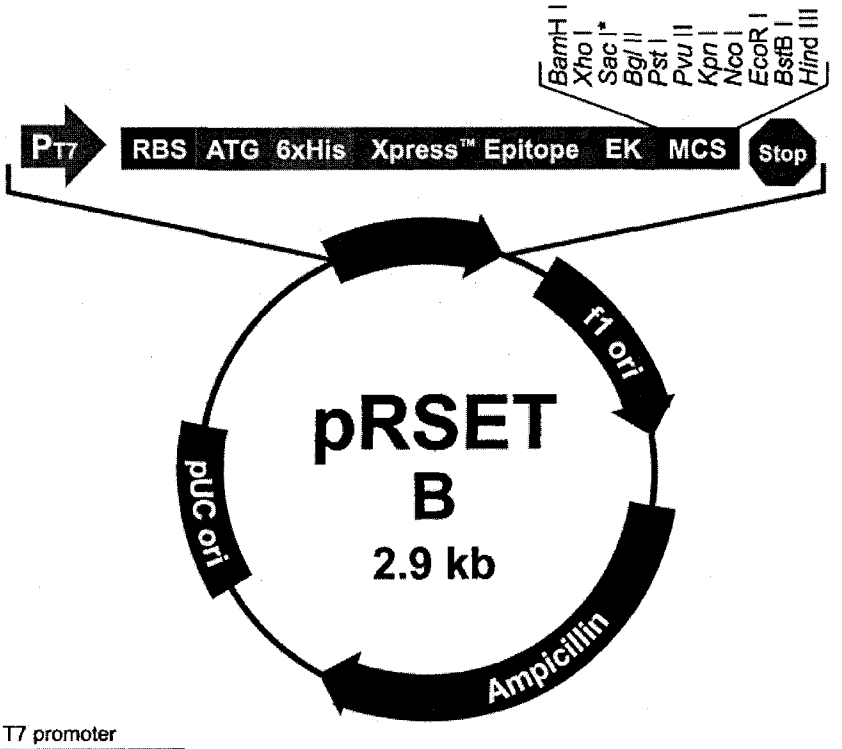




pET-21a-d(+) cloning/expression region

Figure 5-1-a) Map of *pET-21*(+) plasmid vectors (top) and the DNA sequence at the cloning site (Novagen pET manual).

I used pET -21b(+) version of this plasmid.



T7 promoter RBS

21 AATACGACTC ACTATAGGGA GACCACAACG GTTTCCCTCT AGAAATAATT TTGTTTAACT TTAAGAAGGA

Polyhistidine (6xHis) region

91 GATATACAT **ATG** CGG GGT TCT CAT CAT CAT CAT CAT CAT GGT ATG GCT AGC ATG ACT
Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Thr

T7 gene 10 leader Xpress™ Epitope *BamH* I *Xho* I *Sac* I

148 GGT GGA CAG CAA ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT CCG AGC TCG
Gly Gly Gln Gln Met Gly Arg Asp Leu Tyr Asp Asp Asp Asp Lys Asp Pro Ser Ser

EK recognition site EK cleavage site

Bgl II *Pst* I *Pvu* II *Kpn* I *Nco* I *Eco*R I *Bst*B I *Hind* III

205 AGA TCT GCA GCT GGT ACC ATG GAA TTC GAA GCT TGA TCCGGCTGCT AACAAAGCCC
Arg Ser Ala Ala Gly Thr Met Glu Phe Glu Ala ***

T7 reverse priming site

261 GAAAGGAAGC TGAGTTGGCT GCTGCCACCG CTGAGCAATA ACTAGCATAA

Figure 5-1-b) Map of *pRSET-B* expression vector (top) and the DNA sequence of the multiple cloning site (Invitrogen pRSET manual).

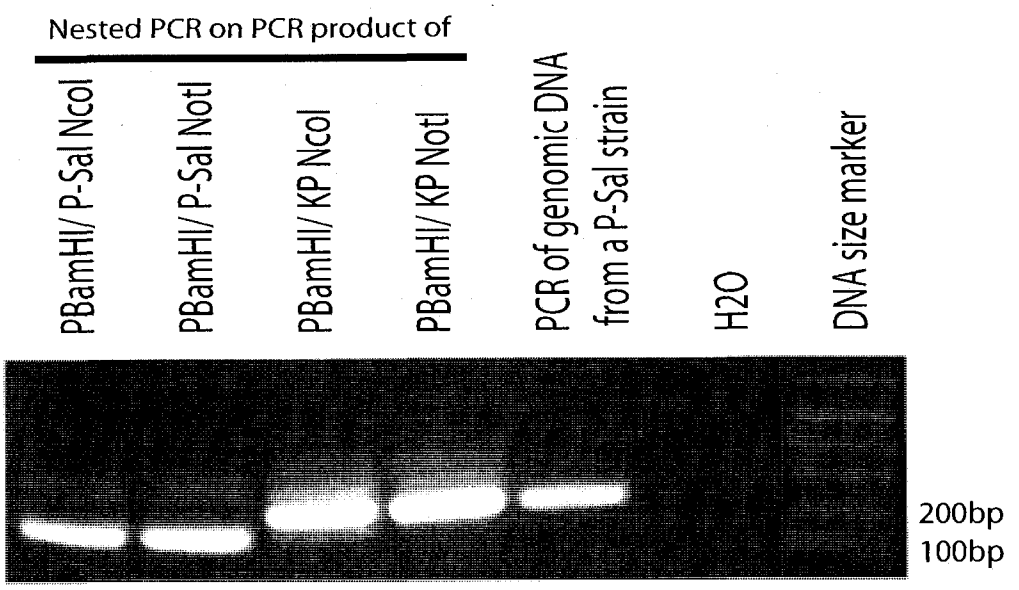


Figure 5-2) Testing for the presence of *P* element cDNA versus genomic DNA in RT-PCR product by KPF1/KPR2 nested PCR.

RNA from fly stocks carrying *KP(2)-beta* and *Sb P-Sal* chromosomes was used for a RT reaction and subsequent PCR by different primer sets. The product of this PCR reaction was subject to a nested KPF1/KPR2 PCR reaction to test for the accuracy of the RT reaction. KPF1/KPR2 PCR reaction would produce a 180 bp band from a cDNA and a 239 bp band from genomic DNA (See Figure 2-1 and Table 2-1).

Both *KP* cDNA preparations produced a band with the same size as PCR band from genomic DNA, while the PCR product from *P-Sal* cDNA preparations was smaller than the PCR band from genomic DNA. Therefore I continued my work with *P-Sal* cDNAs.

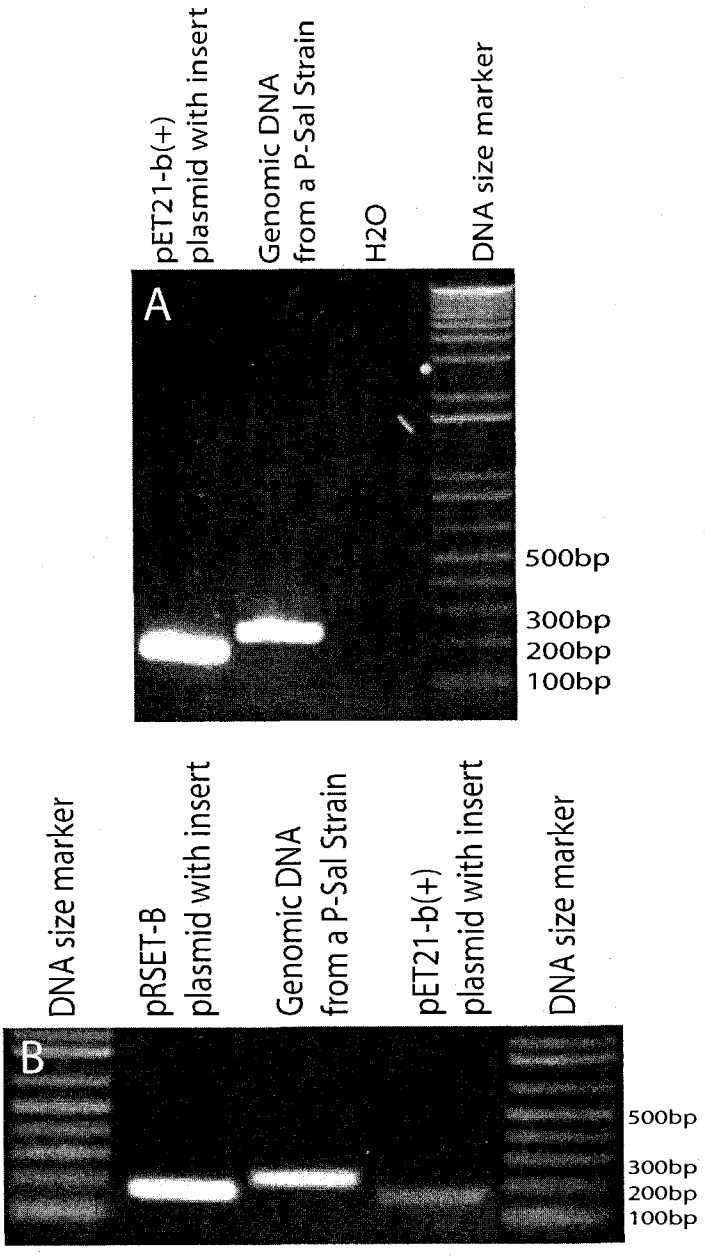


Figure 5-3) Test of the presence of *P* element cDNA in the cloned plasmid by KPF1/KPR2 PCR.

A KPF1/KPR2 PCR reaction would produce a 180 bp band from a cDNA and a 239 bp band from genomic DNA (See Figure 2-1 and Table 2-1). Panel A) The presence of the band that is smaller than the genomic DNA band confirms the presence of *P* element cDNA in the *pET21-b(+)* plasmid. Panel B compares *pRSET-B* and *pET21-b*(+) with each other.

Protein Production and Purification

Transformation of expression host cells

To express the inserted gene in two different plasmids I tried three different *BL21(DE3)* host cells: *BL21(DE3)pLysS* (Novagen), *Rosetta* (Novagen), and *BL21-RIL* (Stratagen). This gave me a choice of six different vector/host combinations. The P repressor protein is 66 kD (Laski *et al*, 1986, Rio *et al*, 1986) and 6xHis tag is about 2-3 kD (Qiaexpressionist protocol, Qiagen) giving a predicted size for the fusion protein that had both 6xHis tag and T7 tag of about 69 kD.

BL21(DE3) is an *E. coli* strain that is *OmpT* and *lon* protease deficient, and has a chromosomal copy of a *T7 RNA polymerase* gene (as a *DE3* lysogen) under the control of *lac operon* operator sequence which permits induction of expression of an inserted gene by Isopropyl β-D-1-thiogalactopyranoside (IPTG) addition (Novagen manual). *BL21(DE3) pLysS* is a modified *BL21(DE3)* strain that has a Cam^R plasmid with a *T7 lysozyme* gene that is a natural inhibitor of T7 RNA polymerase. This reduces the chance of expression of the target gene before induction. This is specifically useful in cases of toxic proteins that can kill or inhibit the host cell before the selected time of induction (Novagen manual). Both *Rosetta* (Novagen) and *BL21-RIL* (Stratagen) are modified *BL21(DE3)* strains that are corrected for eukaryotic-prokaryotic codon bias and have some extra tRNA genes on a Cam^R plasmid. This will increase yield of protein production from eukaryotic genes (Novagen and Stratagen manuals). I transformed *BL21(DE3) pLysS* and *Rosetta* following Novagen user protocol TB009 Rev F0104. And for *BL21-RIL* transformation I followed BL21CodonPlus Stratagen manual. The resulting clones were tested for the presence of the *P-Sal* cDNA *pET21-b(+)* plasmid by KPF1/KPR2 PCR (data not shown), and then cultured for glycerol stocks and time course experiments.

Evaluation of the best combination of plasmid and host cells

To find the best combination of plasmid and host cell and the best induction time length, I did a time course experiment following the Qiaexpressionist protocol (Qiagen). For each transformed host cell, I inoculated a pre-warmed 50 ml of bacterial culture (modified LB (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g glucose per liter plus Ampicillin 100 μg/ml and Chloramphenicol 30 μg/ml) with 2.5 ml of the fresh saturated culture and incubated it for 30 minutes at 37°C in a rotating incubator until the OD_{600} =~0.5. A 1 ml sample was gathered before adding IPTG to a final concentration of 1mM. Additional 1 ml samples were gathered at 3 hours and then again at 12 hours. This time course experiment was used to find the best time of induction for maximal protein expression. Based on preliminary results (Figure 5-4) I chose the

pET21-b(+) plasmid with *Rosetta* as the host cell. This combination was the only one that produced a strong ~70 kD band after 3h of induction. I tried to optimize the protein production by increasing the pre-induction cell concentration by an incubation period of 40min, increasing IPTG final concentration up to 2.5 mM, and decreasing the induction time to two hours to reduce the chance of cell death and protein degradation before harvesting. I increased the amount of Ampicillin and Chloramphenicol by 50% to a final concentration of 150 μg/ml and 45 μg/ml respectively, to select for the cells that retained the plasmids. I also added 20 ml Tris (pH 7.5) per liter of the bacterial culture as a buffer to prevent any pH drop in the culture and prevent from Ampicillin inactivation in the acidic condition.

Protein expression and denaturation

For the full scale expression, I transformed the *Rosetta* again and used fresh transformants for induction. I made a 50 ml saturated bacterial culture in LB medium (Sambrook, 1989) at 37°C in a rotating incubator overnight and added 10 ml of it to each of the five different one-liter flasks containing 200 ml of pre-warmed modified LB. To each flask I added Ampicillin and Chloramphenicol to the final concentration of 150 μg/ml and 45 μg/ml respectively. After 40 minutes of pre-induction incubation in a 37°C rotatory shaker, IPTG was added to each flask to a final concentration of 2.5 mM. After two hours of incubation in a 37°C rotatory shaker, the bacterial cultures were centrifuged. Total weight of pellets was ~3 g. The pellets were transferred to two 50 ml falcon tubes and 25 ml of 6 M Guanidine Hydrochloride (GuHCl) (pH 8.0) was added to each tube for cell lysis and protein denaturation. Protein denaturation was continued for 72 hours at 50°C until the liquid was translucent.

Protein purification

The lysates were centrifuged and supernatant transferred to new falcon tubes. Four milliliters of 50% Ni-NTA resin (Qiagen) was added to the supernatant of each tube. Tubes were inverted continuously at room temperature overnight. For each tube, Ni-NTA resin was separated from the lysate by centrifuging. The collected resin was washed in 12.5 ml of 8 M Urea (pH 7.5) for one hour by continuous inversion and then spun down again and rewashed for one hour in 12.5 ml of 8 M Urea (pH 6.3). Finally, the suspension from each tube was transferred to an empty Prep Sep R 1 ml chromatography column (Fisher Scientific) and the resin was packed by gravity. The flow through was gathered and the column rewashed with 5 ml of 8 M Urea (pH 6.3). For elution, I added 6 ml of 8 M Urea (pH 4.5) to each column and let it pass through the column by gravity. A second elution was done with 2 ml of 8 M Urea (pH 4.0) to make sure there is no protein remaining on the resin (Figure 5-5). As I mentioned before, the predicted size for the

fusion protein that had both a 6xHis tag and a T7 tag was about 69 kD. To separate protein from the urea solution, I dialyzed the eluate using Spectra/Por 3 dialysis membranes 3,500 MWCO (Spectrum Laboratories) in 10 liter of phosphate buffered saline (PBS) (Sambrook, 1989) at 4°C overnight. The protein precipitated in PBS buffer inside the bag was transferred to a new tube and was re-suspended in 1 ml of PBS buffer. The PBS buffer inside the bag was concentrated using 30 kD Amicon Ultra-4 Centrifugal Filter Units (Millipore) following the manufacturer's protocol to reduce its volume to 1 ml. Further, to measure the amount of protein in both the re-suspended precipitate and the concentrated PBS solution I electrophoresed a sample of each of them on SDS-PAGE (Sambrook, 1989) along with different amounts of 66 kD bovine serum albumin (BSA)-fraction V (Sigma). I used 0.5 µl out of 1 ml of re-suspended precipitate and 10 µl out of 1 ml of concentrated PBS solution to load on the gel (Figure 5-6). The amount of precipitated protein was approximately equal to 1 µg of BSA while amount of protein in concentrated solution was less than 0.5 µg of BSA. Thus the total amount of protein in the precipitate was about 2 mg, while the solution had 50 µg at most.

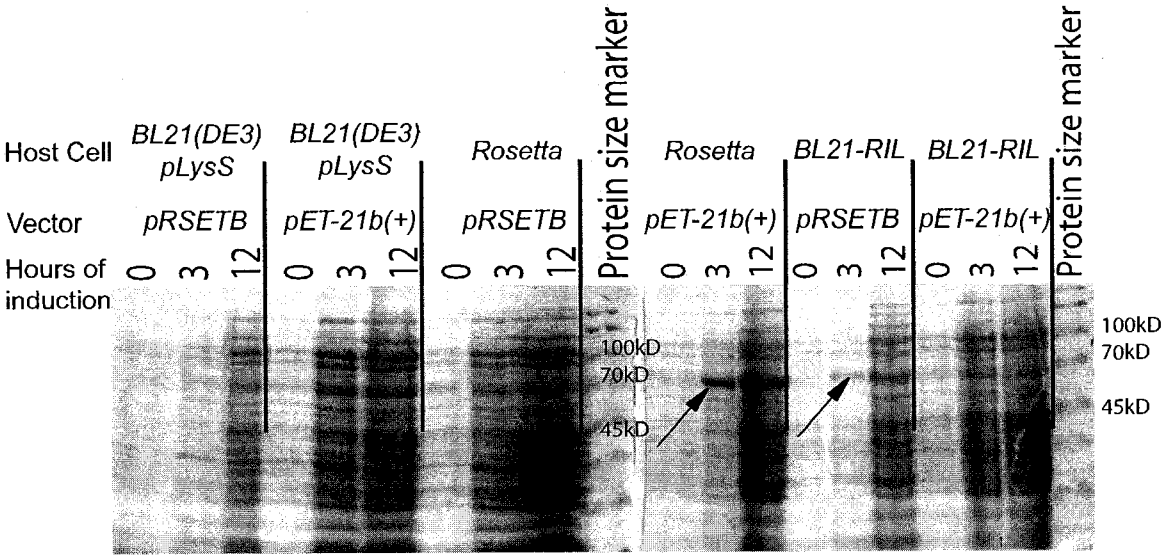


Figure 5-4) P repressor protein expression – time course study.

A Coomassie blue stained SDS-PAGE gel of bacterial lysates of different transformed cell cultures that were induced by IPTG to express the vector plasmid containing the *P-Sal* cDNA. 50 ml of bacterial culture were incubated for 30 minutes at 37°C in a rotating incubator until the OD_{600} =~0.5. A 1 ml sample was gathered before adding IPTG to a final concentration of 1 mM. Additional 1 ml samples were gathered at 3 hours and then again at 12 hours. This time course was used to find the best time of induction for maximal protein expression. Based on these results both the *pET21-b(+)* plasmid in a *Rosetta* host cell and the *pRESTB* plasmid in a *BL21-RIL* host cell can express a ~70 kD protein. The best combination is the *pET21-b(+)* plasmid in a *Rosetta* host cells.

Protein size marker: NEB broad range protein ladder

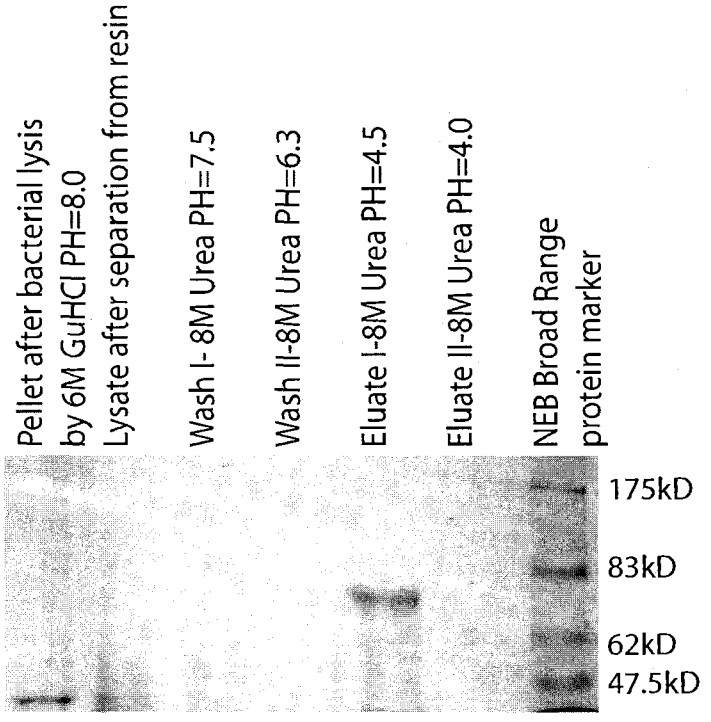


Figure 5-5) Recombinant type I repressor protein extraction, denaturation and purification.

SDS-PAGE gel stained with Coomassie blue showing the purification of the recombinant type I repressor protein by Ni-NTA resin. The induced bacterial culture was lysed and protein were denatured in 6 M Guanidine Hydrochloride (GuHCl) (pH= 8.0) for 72 hours at 50°C. The lysate was centrifuged and the supernatant added to Ni-NTA resin and mixed by inversion overnight. The suspension has centrifuged and the pellet was washed in 8 M Urea (pH =7.5) for one hour and rewashed in 8 M Urea (pH =6.3). Then the pellet was washed in 8 M Urea (pH =4.5) to elute the protein bound to the Ni-NTA. The second elution in 8 M Urea (pH =4.0) does not have any protein and showed that no residual protein was bound to the Ni-NTA.

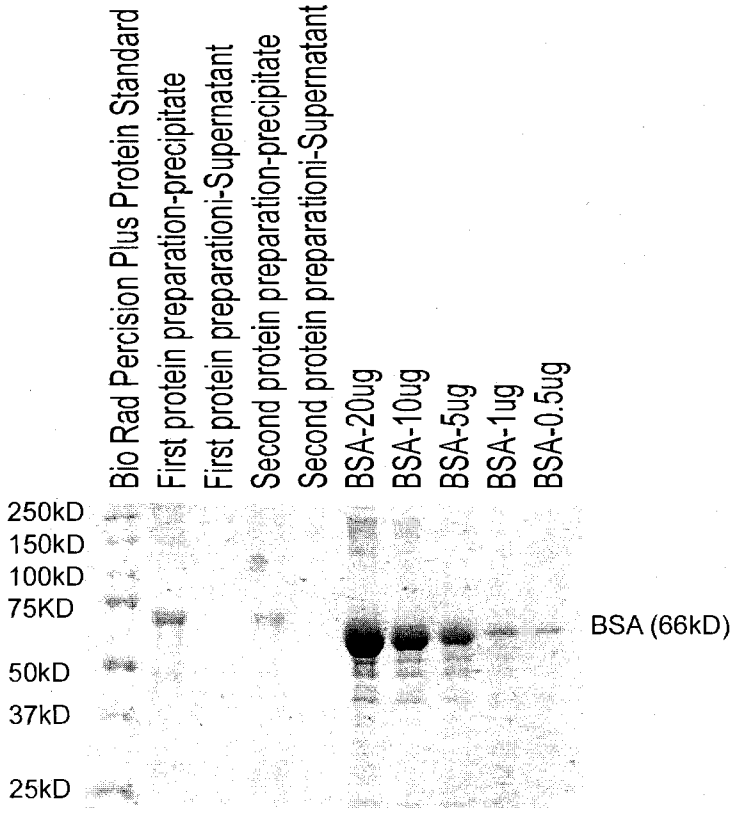


Figure 5-6) Concentration of P repressor protein in the final preparation.

The 2nd through the 5th lanes of the SDS-PAGE gel that is stained with Coomassie blue are loaded by the final denatured P repressor protein solution. The last sis lane are loaded by different concentrations of Bovine serum albumin (BSA)-fraction V (Sigma).

For each protein preparation, the first eluate of Ni-NTA columns was dialyzed in 3,500 MWCO dialysis membranes in 10 liter of phosphate buffered saline (PBS). The precipitates re-suspended in 1ml of PBS buffer and 0.5 µl of each was added to loading buffer, boiled and loaded to the corresponding lane. The supernatants were concentrated using Amicon Ultra-4 30 kD tubes and 10 µl of each was added to loading buffer, boiled and loaded on the gel. BSA was diluted in PBS in different concentrations, boiled and loaded to corresponding lanes. Amount of protein loaded in each lane is mentioned above that lane.

Rabbit Immunization

I used two New Zealand white rabbits from the animal facility of the Department of Biological Sciences for immunization with P repressor protein and Freund's adjuvant. They were each injected with 350 µg of denatured P repressor protein three times at one month intervals following University of Alberta and Department of Biological Sciences protocols. Blood samples were gathered before the first immunization (pre-immunization sample), and two weeks after each round of immunization. Blood samples were coagulated and kept at 37°C for one hour to let the blood clot contract. Then serum was transferred to new tubes and its cells and debris were removed by centrifugation. The supernatant was aliquoted and frozen. Table 5-1 shows the general process and dates for Rabbit immunization.

Table 5-1) Rabbit immunization.

Date	**Procedure**	**Comment**
Dec 5, 2006	Pre-immune bleed	5 ml
Dec 6, 2006	First injection	Complete Freund's adjuvant (FCA) 1.5 ml SQ in 3 locations
Jan 3, 2007	Second Injection	Incomplete Freund's adjuvant (FIA) 1.5 ml SQ in 3 locations
Jan 17, 2007	Test Bleed	5 ml
Jan 31, 2007	Third Injection	Incomplete Freund's adjuvant (FIA) 1.5 ml SQ in 3 locations
Feb 14, 2007	Test Bleed	5 ml
Feb 22, 2007	Termination Bleed	

Western Blots

I tested the sensitivity and specificity of the antibody in the serum of the second test bleed using a fly extract that was lysed in 100 µl of lysis buffer (Nelson, 2004) and centrifuged down to separate debris from lysate. The supernatant was added to 50 µl of 4x SDS-PAGE loading buffer (Sambrook, 1989) and boiled for 10 minutes to denature and solubilize proteins. For a fly lysate lacking P element proteins (negative control), I used *y w; +; +;* + which is an M cytotype fly stock (Scott Hanna, personal communication). As positive control, I used *y w; +; Sb P-Sal/ TM6B Tb Hu;* + that had the parental *P-Sal* insert. Since I used adult flies from this stock to make *P-Sal* cDNA successfully, I inferred that *P-Sal* was expressed in adults of this strain. The equivalent of one fly was added to each lane. I also added 0.3 µg and 0.0003 µg of the recombinant type I repressor protein to other lanes. I made six strips and in each strip the order of the lanes was the same (*y w; +; +;* + fly extract, *y w; +; Sb P-Sal/ TM6B Tb Hu;* + fly extract,

0.3 μg and 0.0003 μg of the recombinant type I repressor protein, and size marker -Precision Plus Protein Dual Color Standard from BioRad).

Samples were electrophoresed on SDS-PAGE and transferred to nitrocellulose membrane using a BioRad Mini Trans-Blot System. When transfer was completed the membrane was blocked overnight at 4°C in blocking solution (0.1% skim milk powder in PBT). I used three blots per rabbit: one for pre-immune serum and the other two for two different dilutions (1/50 and 1/1000) of the second test bleed serum. The sera were diluted in 0.1% skim milk powder in PBT and membrane incubated with diluted serum at room temperature for 90 minutes. Then I washed the membrane in 0.1% skim milk powder in PBT. I used 1/25000 dilution of anti-rabbit HRP antibody (Jackson Immunologicals) as secondary antibody and used Super Signal West Pico Chemiluminescent Substrate kit (Pierce Inc) to induce chemiluminescence. Radiography films were exposed to probed membrane for different periods of time for which a five second exposure gave the best result. Pre-immune sera did not detect the recombinant type I repressor protein, while even 1/1000 dilution of post-immune serum detected 0.0003 μg of the protein. In other words, the rabbits had produced the antibody against the antigen that had been injected to them (Figure 5-7). But in fly extract, antibody did not detect any specific band (Figure 5-7).

To have more specific results I repeated the Western blot experiment with some changes. First, I used a 1/50 dilution of pre-absorbed rabbit serum to increase specificity of the serum (see below). Second, I increased the amount of loaded protein by four times (equivalent of four flies in each lane) to compensate for any decrease in sensitivity of pre-absorbed serum. Third, I used a low bis-acrylamide SDS-AGE system (Scheid, 1999) to separate the bands better. Finally, I used more positive controls. In addition to the *P-Sal* strain, I used two other *P* strains that show a visible *P* element related phenotype in the eyes of adult flies. KP(D) (*y w; Cy Bl vg/* ap^{Xa}*; +, +*) induces strong PDS in flies. It has four *KP* inserts on a *Cy Bl vg* chromosome that encode the KP protein which is much smaller than a type I repressor (Bushey, 2004). The third positive control was *w; +; Sb e Δ2-3/TM6 Ubx e; +*. Due to the lack of the 2-3 intron in its structure, *Δ2-3* encodes transposase in somatic tissues. Based on the DNA sequence, the KP protein has 207 amino acids, the type I repressor has 576 amino acids and transposase has 751 amino acids, but all of them share the first 199 amino acids. Therefore, each of the positive controls should show a specific band, which is different in size from other controls. After electrophoresis on SDS-PAGE, I blotted the gel to nitrocellulose membrane using BioRad Mini Trans-Blot System. When transfer was completed the membrane was blocked overnight at 4°C in blocking solution (0.1% skim milk powder in PBT). To decrease any non-specific binding I tried to pre-absorb the serum

using a lysate derived from *y w, +; +; +* flies (negative control). For this lysate I squashed 50 frozen flies in 200 µl of PBS and boiled the lysate for 10 minutes to inactivate any protease. After cooling, I added 200 µl of rabbit serum and 600 µl of PBS to it and mixed thoroughly. After two hours at room temperature, centrifugation removed fly debris and the supernatant was used for the experiment. This process hypothetically should absorb all antibodies that attach to the proteins that are common between the positive and the negative controls and leave specific antibodies against P repressor proteins in the solution. Therefore these antibodies would be available for probing of the Western Blots. The supernatant were diluted in 0.1% skim milk powder in PBT to a final dilution of 1/50. I incubated the membrane with this diluted pre-absorbed serum for 90 minutes at room temperature and followed the steps that I did in the last Western blot. The resulting Western blot did not show any specific band in any of positive controls (Figure 5-8). This could be due to the lack of a sufficient amount of P element proteins in adult tissues. Another explanation may be that all of P protein antigenic epitopes are present in the adult fly proteome. In this case there is no specific antibody and during pre-absorption all of the antibodies that may detect P protein epitopes have attached to other adult fly proteins in the lysate. Further investigation is necessary to affinity purify the antibody and to optimize Western blot conditions.

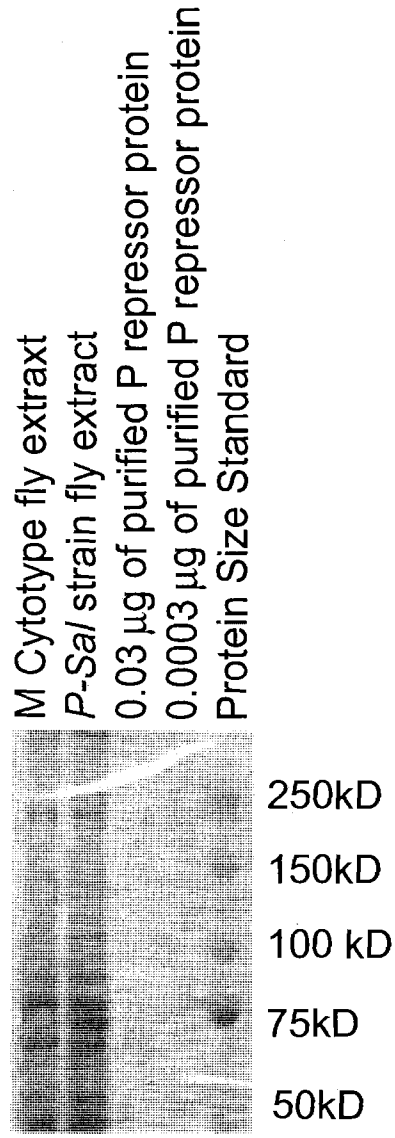


Figure 5-7-a) Loading control of Western blot test of rabbit serum specificity and sensitivity for P repressor protein.

The equivalent of one fly was added to 1st and 2nd lanes for M cytotype strain (*y w; +; +; +*) and *P-Sal* strain (*y w; +; Sb P-Sal/ TM6B Tb Hu; +*). Approximate amounts of P repressor protein in 3rd and 4th lanes are mentioned above each lane. Precision Plus Protein Dual Color Standard (BioRad) was used as protein size standard. The gel was stained by Coomassie blue after electrophoresis.

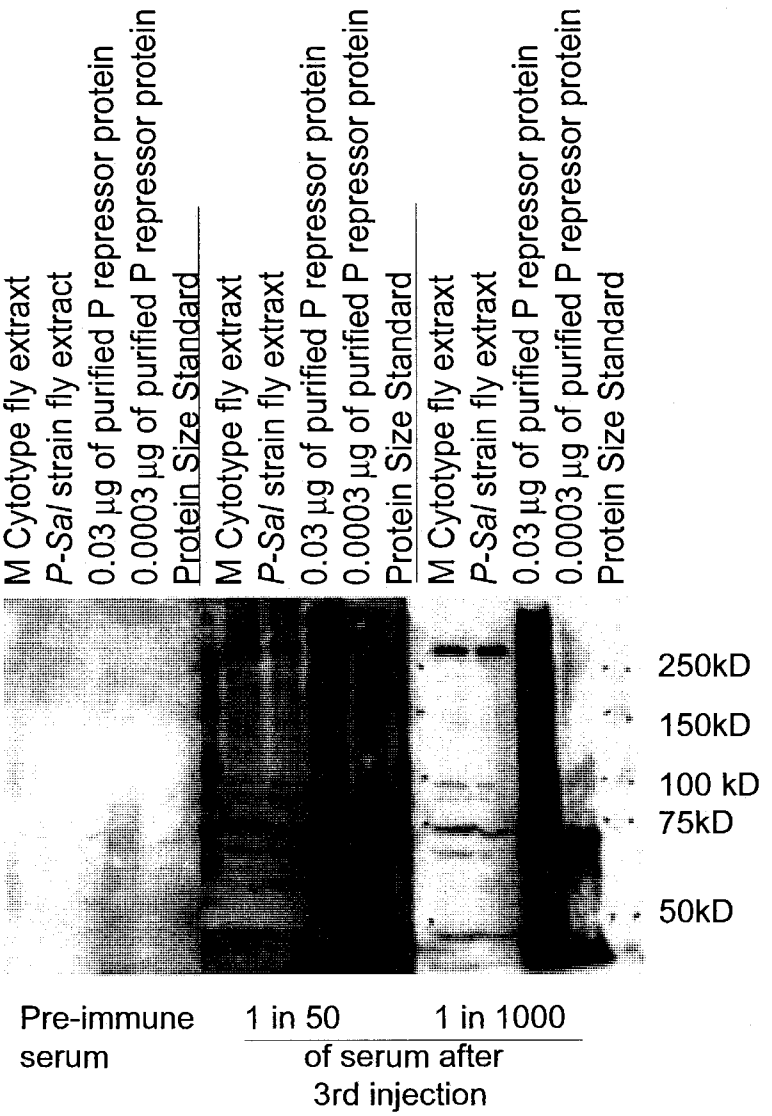


Figure 5-7-b) Western blot test of rabbit serum specificity and sensitivity for P repressor protein.

Three replicas of the Figure 5-7-a SDS-PAGE were transferred to nitrocellulose membrane to make blots. The first blot from left is incubated with pre-immune serum, the middle blot was incubated with 1/50 dilution of the rabbit serum and the right blot is incubated with 1/1000 dilution of the rabbit serum. Anti-rabbit HRP antibody was used as secondary antibody and Super Signal West Pico Chemiluminescent Substrate kit was used to induce chemiluminescence (Pierce Inc). The X ray film was exposed to blots for five seconds. Even in 1/1000 dilution, rabbit serum detected 0.0003 µg of purified P repressor protein strongly (right blot) but it could not differentiate between M strain fly protein extract and *P-Sal* strain fly protein extract.

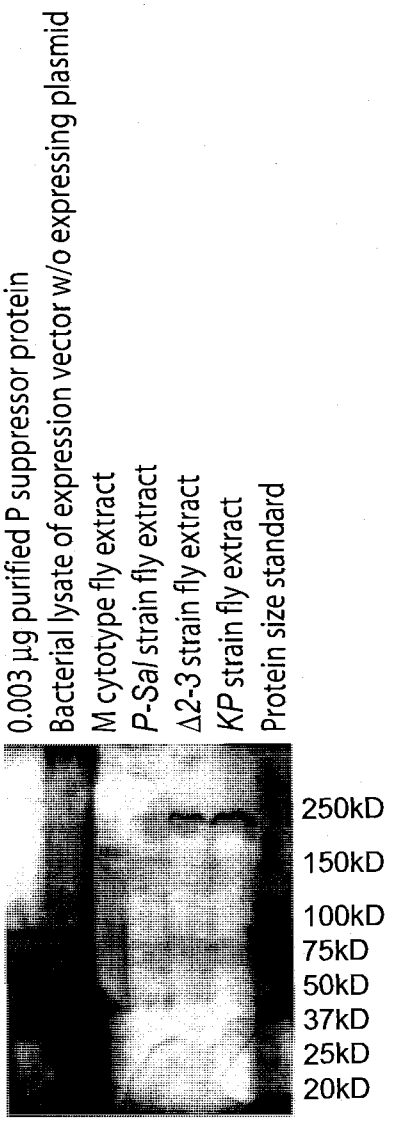


Figure 5-8) Western blot test of pre-absorbed serum using a low Bis-acrylamide (118:1) SDS-PAGE gel.

Each lane has the equivalent of four flies. After blotting, the membrane was incubated with pre-absorbed rabbit serum, washed and incubated with anti-rabbit HRP antibody and chemiluminescence kit (Pierce Inc). The pre-absorbed serum could not detect any specific band in any of positive controls while it was still able to detect 0.003 µg of the recombinant type I repressor protein. Precision Plus Protein Dual Color Standard (BioRad) was used as protein size standard.

M cytotype strain: *y w; +; +; +.*

P-Sal strain: *y w; +; Sb P-Sal/ TM6B Tb Hu; +.*

Δ2-3 strain: *w; +; Sb e Δ2-3/TM6 Ubx e; +* that does not have 2-3 intron and encodes transposase in all tissues.

KP strain: KP(D): *y w; Cy Bl vg/ ap^{Xa}; +, +* that has four *KP* inserts on a *Cy Bl vg* chromosome.

Chapter 6: Discussion

Overview

In *Drosophila melanogaster, P* element dependent silencing (PDS) has been described as silencing of *mini-white* gene expression in *P* element constructs due to the presence of other *P* elements such as *P-Sal* and *KP* in the genome (Bushey and Locke, 2004). I showed that PDS is produced by a dose-dependent mechanism, *i.e.* increase in the number of *KP* elements in the genome intensifies the silencing effect, although it does not follow a linear curve. The recovery and analysis of 21 point mutations in the *P-Sal* coding sequence establishes the role of the P repressor protein in PDS. These mutants behave differently when tested against different *P* inserts. Although these differences can be attributed to different functional domains, protein prediction programs did not predict any new domains, or any common change of pattern in most of the mutants. The differences among *P-Sal** mutant can also be explained by differences in protein stability that changes the intranuclear concentration of the protein, but the produced antibody against repressor protein was not specific enough to answer this question.

Study of *KP* induced PDS

The Number of *KP* genomic insertion sites is higher than what was expected.

So far all known *P* inserts that are affected by PDS are inserted on chromosome 4 within the distal regulatory sequence upstream of the *ci* gene (Bushey and Locke, 2004). In contrast, it appears that there are a large number of genomic locations for a modifying element.

Bushey and Locke (2004) showed that the *KP(D)* chromosome is able to induce PDS. This chromosome appeared to offer an excellent model for examining *KP* element effects on PDS. However, I could not segregate its multiple *KP* elements by recombination due to the presence of multiple inversions, so I transposed them to new genomic locations. My observation that 45% of the progeny from the transposition cross acquired the PDS phenotype suggests that the ability to induce PDS is not limited to a few genomic locations. This high frequency of new PDS mutants challenged the "golden location" model to explain *KP(D)* chromosome phenotype. Based on that model, I expected to see PDS only if *KP* is inserted in specific locations and the number of new mutants should be much lower than 45% of progeny. This high frequency indicates that many genomic locations are suitable for a *KP* insert to induce PDS. Since the type

II repressors, such as the KP protein, are usually considered weaker than the type I repressors (Gloor *et al*, 1993), it seems that this number could be even higher for the type I repressors.

One *KP* element is enough to induce PDS.

I examined 25 new *KP* strains whose PDS phenotype segregated with chromosomes X, 2, or 3. Using Southern transfer, I selected lines with single *KP* inserts on chromosome 2 and 3 that induced PDS in a *Pci* background. This indicates that one *KP* insert is sufficient for PDS induction. First, none of these new *KP* insert mutants enhances the hPEV phenotype in the $In(1)w^{m4}$, eliminating the possibility that their PDS effect is due to the presence of an *E(var)* mutation. Second, all of the newly produced single *KP* inserts are inserted in the euchromatic areas of chromosome 2 and 3. This argues against the involvement of heterochromatin expansion due to DNA sequence homology in the process of PDS.

PDS is dose-dependent.

With two exceptions, both male and female progeny of all four crosses showed significantly lower eye pigment A_{480} in flies that had two *KP* elements than flies that had just one *KP* element. In both exceptional cases (male progeny of the cross between KP(2)-beta and KP(3)-CG9007 and female progeny of the cross between KP(2)-beta and KP(3)-eff), KP(2)-beta is involved. It is possible that *KP(2)-beta*-induced PDS is so weak in comparison with both *KP(3)-eff* and *KP(3)-CG9007* elements that its presence is insignificant. However, in general, the additive effects support the hypothesis that the strength of PDS is related to the number of effective *KP* elements present in the genome, although this relationship is not linear (A_{480} results, Table 3-3). In other words, the A_{480} difference between the no insert and the double insert samples is smaller than the sum of differences between A_{480} in the no insert sample and each of the single insert samples of the same cross.

First, this non-linearity might be explained by a non-linearity in A_{480}-pigment relationship. If the pigment concentration in this experiment was out of the linear part of the A_{480}-pigment relationship curve, reducing the pigment concentration by half would not cause a similar drop in A_{480}. However, frequent tstsing of spectrophotometer readings of diluted pigment samples maintain their linearity over the range of A_{480} readings used here, so this is not a likely explanation (John Locke, personal communication). Second, eye pigment level-w^+ gene expression relationship may also be non-linear and this might explain the non-linearity between the *KP* copy number and the eye pigment levels as well. This might be the subject of another study. Third, the presence of an *E(var)3-9* mutation on the *TM6B Tb Hu* balancer can contribute to this difference as well. This would increase the PDS intensity that is induced by KP(2), and

reduce the measured A_{480} for this group. That is why I omitted this progeny group from analysis in the first place. The other explanation is that flies with two *KP* inserts had a very small amount of pigment and the PDS effect was pretty much near saturation in this group. If PDS intensity in relation to *KP* expression follows a sigmoid curve (which is the most probable scenario considering other response curves in biology), in this experiment the level of *KP* expression has put the PDS intensity near the plateau part of the curve and thus by increasing the amount of *KP* expression (by doubling the number of *KP* inserts), a similar decrease in eye pigment and A_{480} is not predicted.

The *KP* dose-dependent phenomenon of PDS is similar to *in vitro KP* dose dependent suppression of transposition (Lee *et al*, 1998). This is also in accordance with the conclusion of Jackson *et al* (1988) that intensity of the P cytotype is directly related to number of *KP* inserts present in the genome. They concluded there is a positive selection for higher *KP* elements due to lower mutation rate, sterility and mortality by the strengthening transposition suppression and the weakening of hybrid dysgenesis. However, Itoh *et al* (2007) questioned that conclusion. They reported an increase in *KP* element number in natural *Drosophila melanogaster* populations without any change in the level of transposition suppression and concluded that *KP* number increases due to its transposition advantage rather than an additive effect of *KP* elements in suppressing transposition.

PDS is dependent on repressor protein.

There are three major models that could provide an explanation for PDS. The first model involves DNA homology, such as has been described for telomeric *P* inserts (Ronsseray *et al*, 1998). The second model is an RNAi mediated transcription or post-transcription silencing as it is involved in transposition suppression by telomeric *P* inserts (Simmons *et al*, 2007) and suppressing some *P* element alleles of the *vg* gene(Anderson, 2008). The third model is suppression of *Pci* by a repressor protein that is encoded by *KP* or *P-Sal* as it is reported for P cytotype and transposition suppression (Gloor *et al*, 1993, Lee *et al*, 1996). Bushey (2004) suggested the protein model for PDS is correct, and I concur.

The basis for the first model, DNA homology, and second model, RNAi, are the studies of the left arm telomere of the X chromosome. Some complete and incomplete *P* inserts in this location are able to suppress other *P* elements based on DNA homology at the 5' end of the element (Ronsseray *et al*, 1998). The shortest reported homology is 1.8 kb (Marin *et al*, 2000). This effect is dependent on *Su(var)205* and *E(z)*, which are also involved in hPEV and heterochromatin initiation and maintenance (Roche and Rio, 1998; Haley *et al*, 2005). This

system is also sensitive to mutations in the *aubergine* gene but neither the *piwi* nor the *homeless* genes, which are involved in the RNAi pathway (Reiss *et al*, 2004; Simmons *et al*, 2007). None of these genes is involved in the establishment of the P cytotype (Haley *et al*, 2005; Simmons *et al*, 2007). Bushey (2004) showed that *Argonate-1* and *piwi* are not involved in PDS.

A DNA homology model is unlikely for PDS because the length of homologous sequence between *KP* and *P{lacW}* is only 582 bp, which is much shorter than the 1.8 kb that is reported to be required by Marin *et al* (2000). Next, none of the tested single *KP* mutants is inserted into a telomere. They are inserted into euchromatic regions of chromosomes 2 and 3. Therefore, PDS cannot be due to telomeric effect and DNA homology. Also, this model is not supported by single base substitutions altering the effect of *P-Sal* as discussed in the present study. *P-Sal* and *Pci* are co-linear for the first 582 bp. Sixteen out of the 22 *P-Sal** mutants have not mutated in this area and therefore the level of homology between *P-Sal* and *Pci* has not been affected by mutation but they have lost their ability to induce PDS.

The second model, post-transcriptional silencing of the *mini-white* gene through an RNAi system is unlikely because there is probalbly insufficient sequence homology between the *KP* transcript and the *mini-white* transcript to support the possibility of siRNA involvement in PDS. Moreover, *aubergine1* and *piwi* mutations do not affect PDS and this is against miRNA involvement (Bushey, 2004; Okamura, 2004). Furthermore, the fact that PDS is abolished by 16 point mutations in the *P-Sal* sequences that are not present in *Pci* argues against any homology based mechanism. After all, these mutations have not changed the homology between *Pci* and *P-Sal*.

If PDS is not mediated by DNA or RNA, P repressor protein involvement in PDS seems the most probable model. Single base substitutions in *P-Sal** mutants are enough to alter repressor protein and abolish PDS activity. Further, it was shown that repressor proteins are involved in transposition suppression (Lee *et al*, 1998).

Gene silencing in PDS can be explained by an expansion of a nearby heterochromatin center.

There are many regions along chromosome 4 that are considered heterochromatic because *mini-white* transgene inserts show variegated expression (Sun *et al*, 2000). There is supporting evidence for the presence of a heterochromatinization center near *Pci*. First, transvection in control of the *ci* gene expression shows the presence of a silencing center in the regulatory area of this gene (Locke and Tartof, 1994; Henikoff, 1997; Henikoff and Comai, 1998). Second, Bushey and Locke (2004) reported that the presence of a *gypsy* element upstream

of the *ci* gene suppresses the *mini-white* gene expression in *Pci*. Huang (2005) showed that not only this suppression happens in *cis*, but *gypsy* element can also suppress *Pci* on the homologous chromosome, in *trans*. Third, the heterochromatinization proteins, HP1 and SU(VAR)3-7, are involved in PDS, and increase in the *Su(var)3-7*$^{+}$ gene copy number induces variegation in *Pci* even in the absence of any other *P* element (Bushey and Locke, 2004).

This evidence indicates that PDS is mediated through heterochromatinization and thus there is a heterochromatinization center near *Pci*. Change in DNA sequence (*gypsy* insertion) or change in chromatin protein environment (attachment of the KP protein) leads to expansion of this center and silencing of *Pci*. It is interesting that the KP protein, a foreign protein, can co-opt the heterochromatinization system in a dose-dependent manner just as the endogenous heterochromatin protein SU(VAR)3-7 does.

Dose dependence is beneficial for both organism and transposon.

Uncontrolled mobile element transposition can cause mutations and affect viability. Thus, transposition regulation seems inevitable for maintaining cellular genome integrity. The regulation would also benefit the transposon because if it kills the first cell, it will be never transferred to the next generation. Beall *et al* (2002) made a hyperactive mutant form of transposase. The fact that wild type transposase is weaker than this mutant form indicates a selection against a more active transposase. Therefore, both the host cells and the mobile element will mutually benefit from controls of transposition.

A cell would benefit if transposition is silenced, but native gene expression is kept intact. Localized heterochromatinization is one way to control transposons, but it can expand from the silenced transposon to surrounding native genes. This could affect the organism adversely. A dose-dependent response can be helpful to fine tune heterochromatinization to a level that has the most benefit (suppression of transposition) and the least risk (suppression of surrounding genes). If this is the case, natural selection will favor organisms with an optimal number of repressor elements that is sufficient to suppress just the transposition without any significant effect on any other DNA sequence. Of course, each element has its own level of effect but it is the sum of all effects that determines the final results. For the transposon, it is better to have enough transposition to increase the number of elements but at the same time keep the organism alive and able to overcome the rival competition. Again there is a positive selection for a balance between suppression and transposition. If suppression happens in a dose-dependent manner, this balance can be achieved by balancing the number of repressor and transposase genes.

P-Sal Mutagenesis

Reasons for choosing *P-Sal* for mutagenesis.

I used *P-Sal*, a strong PDS inducer (Bushey and Locke, 2004), rather than *KP* because I was looking for mutants that had a clear reduction or a loss of PDS phenotype. Moreover, this would give me a chance to study the presence of unidentified domains in a type I repressor, *P-Sal*, which are not present in the *KP*, as a type II repressor. Some previous studies indicate the presence of these domains. First, Gloor *et al* (1993) did not find any *P* repressor element shorter than 1956 bp but larger than the *KP* (808 bp). Second, Bushey (2004) showed that exon 2, which is not present in the *KP*, is necessary for *P-Sal* ability to induce PDS. And third, Rio (1990) predicted two leucine zippers at amino acid positions 283-311 and 497-525 that are not present in the *KP* (*KP* has the first 199 amino acids of *P* element).

P-Sal is inserted in the 5'-UTR of the *Pak3*.

P-Sal is inserted in the 5'-UTR region of the first exon of *Pak3* and the direction of *P-Sal* insert transcription is the same as *Pak3*. Thus, it is possible that *P-Sal* follows a *Pak3* expression pattern, although this is unknown at present. *Pak3* encodes a receptor signaling serine/ threonine kinase-like protein (Morrison *et al*, 2000). It is known that the PAK3 protein is involved in actin protein organization and cytoskeleton (Kiger *et al*, 2003) and this suggests that the *Pak3* gene is widely expressed. Merriam (2005) reported an allele of *Pak3* that was mutagenized by *P* element activity and was recessive lethal. Since *P-Sal* homozygotes are viable, this insertion is not an amorphic mutation of *Pak3*. A study of the *P-Sal* expression pattern could help to uncover the *Pak3* expression pattern.

There is no second site modifier within 3 cM of *P-Sal*.

During the *P-Sal* mutagenesis screen I mapped 22 PDS deficient *P-Sal** mutants within 2.9 cM of *P-Sal* and I showed there is no dominant second site modifier gene in this 5.8 cM area (*P-Sal* ± 2.9 cM) by using chromosome 3 deficiencies (*Df(3)*s) for *P-Sal* ± 3 cM (See Table 4-2 for the list of *Df(3)*s). This experiment eliminated the chance of an amorphic PDS suppressor mutation, but there was still a small chance for neomorphic mutations. To study this, I tested the *P-Sal** mutants for their suppression effect on hPEV to check the presence of any *Su(var)* mutation.

*P-Sal** mutants do not suppress hPEV.

Bushey and Locke (2004) showed that there are at least two *Su(var)* genes that suppress both the hPEV and PDS systems. Although there is no known *Su(var)* gene within 3 cM from *P-Sal*, it is possible that the phenotype of *P-Sal** mutants is due to mutations in an unknown

Su(var) gene in this area. However, 21 out of 22 *P-Sal** mutants showed no effect on the $In(1)w^{m4}$ phenotype (Table 4-3). These data support the previous results from the *Df(3)* experiment. Only MPS-19 showed enhancement of the $In(1)w^{m4}$ variegated phenotype which was similar to the *TM6B Tb Hu*. Since *TM6B Tb Hu* has an *E(var)3-9* mutation (Weiler, 2002) MPS-19 may have an *E(var)* mutation. This could explain why this mutant appears as a hypomorph.

There is still a small chance of a neomorphic mutation in the area that is able to suppress PDS but not hPEV. This small chance will be important to explain the antimorphic effects of some of the mutants as I will explain later.

E(var)3-9 in the _TM6B Tb Hu_ chromosome does not induce _Pci_ variegation.

The *TM6B Tb Hu* chromosome carries an *E(var)3-9* mutation (Weiler, 2002) and its effect is detectable on $In(1)w^{m4}$. However this chromosome does not show any enhancing effect on *Pci*, $P\{hsp26\text{-}pt\text{-}T\}ci^{2\text{-}m1021.R}$, and *E1*. In all of the crosses *TM6B Tb Hu* was significantly different from parental *P-Sal* and was similar to amorphic *P-Sal** mutants in its effect on *Pci*, $P\{hsp26\text{-}pt\text{-}T\}ci^{2\text{-}m1021.R}$, and *E1*. This indicates that *E(var)3-9* does not affect *Pci* expression individually. However, this does not exclude its possible enhancement of PDS in the presence of *P* elements. It is already known that *Su(var)3-9* does not affect PDS (Bushey and Locke, 2004). *E(var)3-9* is another locus that affects hPEV, but not PDS. The implication here is that the composition of chromatin must be different in these two systems.

Four _P-Sal*_ mutants have an antimorphic phenotype.

Four out of 22 *P-Sal** mutants, MPS-6, MPS-13, MPS-16, and MPS-17, showed some antimorphic effect. MPS-6 and MPS-13 are still able to suppress *P{hsp26-pt-T}39C-12* variegation strongly and MPS-16 suppresses it weakly. This indicates the presence of residual repressor protein in the cell that has lost its PDS function. Although three of these mutants were mutated in exon 2, computer program predictions were contradictory. Only Porter predicted a change in the leucine zipper 497-525. It predicted changes for MPS-6, MPS-16 and MPS-17 but it spared MPS-13 even though this mutation is closer to the zipper than MPS-6. Porter also predicted a change in this zipper for six other mutants but none of them showed any antimorphic effect.

There are several possible explanations for the antimorphic effect. The most plausible one is a competitive inhibition of the parental type I repressor protein. If the mutant protein has lost its ability to induce PDS but is still able to bind to DNA, it will compete with parental repressor protein in binding to specific DNA sequences. This can be the case if we consider that none of these mutants were mutated in the DNA binding domain of the *P* element. Another

possible explanation is a dimerization-poisoning effect by the mutant protein. None of these four mutants were mutated in the 101-122 leucine zipper that is effective in the KP protein dimerization (Lee et al, 1996). If the mutant proteins still have this domain, they should be able to bind to the parental repressor protein and perhaps inactivate it. A third way to explain the antimorphic effect of these four mutants (MPS-6, MPS-13, MPS-16, and MPS-17) is by a gain of function mutation that induces expression from the *mini-white* gene of *Pci*, a mechanism similar to the suppression of variegation of *P{hsp26-pt-T}39C-12*. The last explanation is a second site modifier. A neomorphic mutation that suppresses PDS but does not affect hPEV can not be found by any of my screening experiments for second site modifiers. However, it can affect PDS phenotype both in *cis* and in *trans* (both the mutated and the parental *P-Sal* chromosomes).

_P-Sal_* mutants do not induce PDS on *P{hsp26-pt-T}ci$^{2\text{-}m1021.R}$*.

P{hsp26-pt-T}ci$^{2\text{-}m1021.R}$ is another PDS sensitive *P* insertion (Bushey and Locke, 2004). *Pci* and *P{hsp26-pt-T}ci$^{2\text{-}m1021.R}$* are inserted in adjacent locations (Figure1-5); therefore the chromatin status should be very similar for both of them. Thus, it is not surprising that the phenotype of *P-Sal** mutants are similar for both of them. There is one exception. MPS-19, a hypomorph for *Pci,* acts as an amorph here. This may indicate that *P{hsp26-pt-T}ci$^{2\text{-}m1021.R}$* is less sensitive to the presence of P repressor proteins than *Pci*.

Nine *P-Sal mutants suppress variegation of *P{hsp26-pt-T}39C-12.***

Eye color of *P{hsp26-pt-T}39C-12* carrier flies is variegated and this variegation can be suppressed by *Su(var)205* mutations, an increase in the *Su(var)205*$^{+}$ copy number and both the type I and the type II repressors (Sun et al, 2000; Haynes *et al*, 2007; Bushey, 2004). Bushey (2004) showed that *P-Sal* suppresses the variegated phenotype of *P{hsp26-pt-T}39C-12* to produce an essentially w^{+} phenotype. This is different from the *P-Sal* silencing effect on expression of *Pci* and *P{hsp26-pt-T}ci$^{2\text{-}m1021.R}$*. Nine *P-Sal** mutants suppress variegation of *P{hsp26-pt-T}39C-12.* MPS-6, MPS-8, MPS-13, and MPS-19 are similar to parental *P-Sal* in their suppression of variegation of *P{hsp26-pt-T}39C*-12. Moreover, two of them, MPS-6 and MPS-13, have an antimorphic effect against PDS induced by parental *P-Sal.*

The *P-Sal** mutant MPS-8 encodes a polypeptide that is just 14 amino acids shorter than the KP protein. This shortened version of the KP protein is similar to the wild type KP protein in its phenotype regarding *P{hsp26-pt-T}39C-12*, but is amorphic regarding PDS. This may indicate a critical role for the last 14 amino acids of the KP protein in effecting *Pci*. This area may be a functional domain or it may play a role in stabilizing the KP protein.

MPS-2, MPS-7, MPS-15, MPS-16, and MPS-20, showed a hypomorphic phenotype on *P{hsp26-pt-T}39C-12*. One of these five mutants also has a PDS antimorphic phenotype. MPS-2 is mutated in the DNA binding domain and has lost its ability to induce PDS. However, it is still able to partially suppress variegation of *P{hsp26-pt-T}39C-12*. This may imply that DNA binding is not necessary for this phenotype. Another possibility is that the specificity of the DNA binding has changed and that the MPS-2 product is still able to partially attach to DNA and induce its effect on *P{hsp26-pt-T}39C-12*.

The fact that MPS-20, which lacks a mutation in its *P* element sequence, is hypomorphic regarding *P{hsp26-pt-T}39C-12*, and it is amorphic on its effect on *Pci* and $P\{hsp26\text{-}pt\text{-}T\}ci^{2\text{-}m1021.R}$, suggests that MPS-20 is a reduced-expression version of the parental *P-Sal*. This could be due to a mutation in the regulatory sequence of the *Pak3* gene that may control *P-Sal* expression.

It seems that in all nine mutants, the mutant P-Sal* protein is still present inside the cell in enough concentration to affect *P{hsp26-pt-T}39C-12*, and to show an antimorphic phenotype against parental *P-Sal*. If *P{hsp26-pt-T}39C-12* is more sensitive than *Pci* to the presence of the type I repressor protein, its phenotype will change from w^{var} to w^{+} even with a decreased level of repressor protein (MPS-8 and MPS-20) or a decreased ability to bind to DNA (MPS-2) while the *Pci* phenotype will not change in any of these cases because the protein concentration is not enough to affect *Pci* expression.

P-Sal* mutants no longer enhance _E1_.

E1 ($P\{lacW\}ci^{E1}$) is a *Pci* with a distal *gypsy* transposon insert. *E1* has a w^{var} phenotype by itself and the presence of *P-Sal* will enhance its w^{var} phenotype (Bushey and Locke, 2004). All of the *P-Sal** mutants, except MPS-19 which was a PDS hypomorph, lost their ability to enhance the *E1* phenotype. As with $P\{hsp26\text{-}pt\text{-}T\}ci^{2\text{-}m1021.R}$, the lack of PDS in all *P-Sal** mutants on *E1* was predictable based on their lack of PDS on *Pci*. This indicates that a similar mechanism is involved in silencing of all of these different constructs in this region upstream of the *ci* gene on chromosome 4.

P-Sal* mutants have point mutations in _P-Sal_ sequence.

Sequencing results of the 22 *P-Sal** mutants that were mapped within 2.9 cM of *P-Sal* showed that most of the *P-Sal** mutants (20/22) had one single base substitution. MPS-8 had a double change and MPS-20 lacked any change in *P-Sal* (Table 4-8). The fact that point mutations in *P-Sal* abolish its PDS effect supports the contention that the P repressor protein is responsible for the gene silencing and weakens the possibility of DNA and RNAi involvement in PDS. Both DNA and RNAi mechanisms are based on homology between two sequences and a point

mutation does not change the homology between *P-Sal* and *P{lacW}* significantly, but can affect the protein structure/function completely. Moreover, while *P-Sal* and *P{lacW}* sequences are co-linear for only their first 582 bp, in 16 mutants the change occurs in a base pair after base pair 582 and does not affect homology between these two construct at all.

MPS-1 and *SP* element encoded polypeptides are similar.

Rasmusson *et al* (1993) reported a very short type II repressor called *SP* element. It is 0.5 kb long and has a deletion of base pairs 187-2576 of the *P* element. Theoretically it encodes a 14 aa polypeptide that has the first 11 amino acids of the transposase N-terminus and the last three amino acids are encoded by an out of frame translation of exon 3. *SP* is able to suppress hybrid dysgenesis and sn^{w} hypermutability through a zygotic effect when the source of transposase is a complete *P* element. *SP* does not suppress transposition from $P(ry^{+}, \Delta 2\text{-}3)99B$ and does not suppress the $vg^{21\text{-}3}$ phenotype (Rasmusson *et al*, 1993). There are no data available about the effect of this element on PDS or other *P* element related phenotypes. Since it does not have the complete DNA binding domain and does not have the nuclear localization signal (NLS), it is very unlikely that its phenotype depends on its ability to encode any polypeptide. MPS-1 is mutated at base pair 238 (codon 29), encodes a very short 28 aa polypeptide from the N-terminus of transposase, and is an amorph regarding PDS. This mutant does not affect *E1*, $P\{hsp26\text{-}pt\text{-}T\}ci^{2\text{-}m1021.R}$, and *P{hsp26-pt-T}39C-12*. It also did not have any antimorphic effect. MPS-1 provides a very useful tool to examine the effects of *SP* and its mechanism of suppression of transposition.

The DNA binding domain is mutated in four mutants but the zinc finger is spared.

The repressor protein DNA binding domain is a single zinc finger THAP domain that is located at the N-terminus of the protein. This domain is necessary for specific DNA binding of protein and its suppressive effect on transposition (Lee *et al*, 1998). There is a nuclear localization signal (NLS), near the end of the DNA binding domain at amino acid residues 64-68. Without an NLS, the protein won't be transported to nucleus, the place that silencing and heterochromatinization take place, and without a DNA binding domain transport to the nucleus is futile. These are critical functions in the protein and my expectation was that I would find many mutants in these domains, especially in the zinc finger since it is an important part of the DNA binding domain.

MPS-1, MPS-2, MPS-3, and MPS-4 are mutated in the DNA binding domain. MPS-1 is a nonsense mutation; therefore, only 3 out of 18 missense mutants (16.6%) are in the DNA binding domain. The DNA binding domain (base pairs 153-416) is 15.3% (264 bp) of total coding *P-Sal*

sequence (1728 bp). Thus, 16.6% of mutations happened in 15.3% of the sequence. This shows that no bias occurred towards this region. Moreover, none of these mutants was mutated in the zinc finger (CCHC) domain of the protein, which is encoded by base pairs 162-218 of *P-Sal* (57 bp or 3.3% of the coding region). The presence of four mutations in a 10 bp area near the end of exon 1 makes lack of any mutation in this 57 bp length of DNA strange. If these mutations were distributed randomly, they should follow a Poisson distribution along the *P* element. But a Poisson distribution shows this difference is unlikely to be due to random distribution. To do the Poisson probability calculation, the EMS mutagenesis rate was required for this study. Since EMS mutagenesis rate is dependent on EMS concentration and also on DNA sequence of the gene under investigation, the calculated mutagenesis rates from other studies could not be used here. Moreover, screening scheme in this study only selected amorphic mutants; thus I need to know the rate of mutations that change the phenotype to amorph and the plain mutagenesis rate was not useful here. Therefore, I used the data from my study to estimate the rate for *P-Sal* amorphic mutations. After screening of 14565 *P-Sal* heterozygote flies, 21 mutations were found in 2000 bp (as mentioned before, *P-Sal* was sequenced from base pair 100 through base pair 2100 for all mutants). These numbers gave a rate of 0.0105 amorphic mutation per base pair for the whole screen. As 14565 flies were scored the final mutagenesis rate was 1/140000 mutation per base pair per genome. Based on Poisson calculation the chance of having no mutation in a 57 bp sequence is 54.96% while the chance of 4 mutations in a 10 bp sequence is almost nil (0.00056%). Therefore, the distribution of these mutations is not random.

MPS-2 still has the ability to weakly suppress variegation of *P{hsp26-pt-T}39C-12*, suggesting it may retain some weak ability to bind to DNA although it is mutated in the DNA binding domain right after the zinc finger. MPS-4 is mutated in NLS. In this mutant, an arginine has been replaced by a similar amino acid, a lysine. But this is enough to change the NLS and deactivate the protein.

In addition to mutants that were mutated within the DNA binding domain, change in secondary structure of this domain has been predicted for some of the mutants by computer programs. PSIPRED predicted changes in the DNA binding domain for the nonsense mutations of MPS-8 and MPS-10, and SABLE-2 predicted these changes for MPS-7. Although these programs are not in accordance with each other in their predictions, their results may indicate a more general change in the protein folding pattern that affects the DNA binding domain without changing its amino acid sequence.

MPS-5 is mutated in a splicing acceptor site.

Both MPS-5-1 and MPS-5-2 are mutated at base pair 500, the acceptor splicing site of intron 0-1. This predicts a change in the splicing pattern. The altered polypeptide may have a different folding pattern or may be unstable with a shorter half life than required to affect PDS. It seems that the amino acid sequence encoded by exon 1 is important for repressor function in PDS and any change in the splicing pattern that affects it would change the function of protein even if the DNA binding domain, which is encoded by exon 0, is intact. This is in contrast with the conclusions of Lee *et al* (1998) who reported that the leucine zipper, which is encoded at the beginning of exon 1, is not necessary for the KP protein's ability to suppress transposition *in vitro*. Moreover, Anderson (2008) reported a new element, *P[21r36]*, which is shorter than *KP* and induces a uniform orange eye color in *Pci* carrier flies. *P[21r36]* encodes a 108 aa polypeptide that shares the first 92 amino acids of the KP protein, including the DNA binding domain and NLS but not the leucine zipper. The rest of the polypeptide is encoded by exon 3 out of frame (Anderson *et al*, 2006). Therefore, it is a construct capable of silencing *Pci*, but without having any sequence from exons 1 and 2. If this is the case, any splicing mutation in the acceptor site of intron 1-2 should end up with a protein that is larger than the P[21r36] protein. Furthermore, MPS-8 has a nonsense mutation and encodes a 194 aa polypeptide that is larger than the P[21r36] protein. Obviously having more sequence here is a disadvantage. Although nonsense-mediated mRNA decay may explain degradation of MPS-8 mRNA due to its premature stop codon, it cannot explain the phenotype of splicing mutants (reviewed by Shyu *et al*, 2008).

A functional domain in the middle of exon 1.

In exon 1, I did not find any mutation in the leucine zipper domain (at the beginning of exon 1). Lee *et al* (1998) did not find any change in the *in-vitro* transposition repression activity of the KP protein that lacks this dimerization domain either. In contrast to the lack of mutants at the beginning of exon 1, three mutants (MPS-6, MPS-7, and MPS-8) are mutated in a 130 bp area shared between *KP* and *P-Sal* which encodes the C-terminus of the KP protein. All three mutants are able to suppress variegation of *P{hsp26-pt-T}39C-12* while they have lost their ability to induce PDS. This suppression of *P{hsp26-pt-T}39C-12* indicates that the mutant repressor protein is still present inside the cell and it is relatively stable. Therefore, this region is not necessary for protein stability in the type I repressor protein. However, inability of these mutants to induce PDS indicates this area is necessary for silencing of *Pci*.

MPS-8 protein is just 14 amino acids shorter than the KP protein.

MPS-8 has two changes: a nonsense mutation at position 794 (at codon 195) and a missense mutation at 1089. The second mutation should be inconsequential as there is an upstream nonsense mutation at base pair 794 (codon 195). This stop codon is just 14 bp upstream of the *KP* element deletion break point, thus MPS-8 encodes a 194 aa polypeptide that is just 14 amino acids shorter than 208 aa KP protein. Since the last eight amino acids of the KP protein are from an out of frame translation of exon 3, the expectation was that they are just a random sequence of amino acids that do not have any function. Thus, the MPS-8 protein is very similar to the KP protein and just 6 amino acids shorter. The MPS-8 effect on *P{hsp26-pt-T}39C-12* indicates that the protein is still present inside the cell but cannot affect the *Pci* due to either damage of a domain or instability of the protein that decreases its cellular concentration. Degradation of MPS-8 mRNA due to nonsense-mediated mRNA decay (reviewed by Shyu *et al*, 2008) is another explanation for MPS-8 phenotype on *Pci*, but it can not explain why it retains the ability to affect *P{hsp26-pt-T}39C-12*.

Anderson *et al* (2006) reported a *P* element called *P[21r36]* that has a deletion of base pairs 430-2680. It potentially encodes a 108 aa polypeptide that is the same as *KP* in its first 92 amino acids and includes the DNA binding domain and NLS but not the 101-122 leucine zipper. The rest of the polypeptide is encoded out of frame by exon 3. It partially silences expression of *mini-white* gene of *Pci* (Anderson, 2008). *P[21r36]* can be a shorter version of *KP* (deleted at 808-2506, Black *et al*, 1987) and it would be interesting to transpose it to new positions and screen for inserts that might be able to induce PDS and affect other *P* element related phenotypes. MPS-8 has most of the coding sequence of *KP* and its encoded protein is very similar to the KP protein and larger than the P[21r36] protein but it does not affect *Pci*. The C-terminus of both KP and P[21r36] proteins are encoded by out of frame translation of part of exon 3, while MPS-8 does not have this part. This may indicate the importance of the C-terminus of the KP protein for its stability and/or function. As reported before, *KP* is a very frequent *P* element in *Drosophila melanogaster* populations gathered from all over the world (Itoh *et al*, 2007; Wook *et al*, 1996). This indicates a strong positive selection for it. This selection might be due to its effect as a transposition repressor factor. The fact that shorter sequences are not as frequent as the *KP* element in the natural populations may indicate that the *KP* element has the minimum required sequence to encode a functional, stable repressor protein and anything shorter than that would be unstable or non-functional regarding P cytotype. A comparison between *KP*, *P[21r36]* and MPS-

8 in their effects on different *P* element related phenotypes may provide a window on how the *KP* element functions.

Nine of the 22 *P-Sal** mutants, (MPS-1 through MPS-8) were mutated in sequences common to *P-Sal* and *KP*. The other 11 mutants are altered in the sequences that are necessary for *P-Sal* function, but are not present in *KP* at all. We know that these sequences are necessary for the type I repressor (Gloor *et al*, 1993; Bushey, 2004). Most probably this part of the P-Sal protein is important for its stability. The best way to test this would be to check the protein level in the mutants and compare them with the parental stock. The antibody produced against P repressor protein was not specific enough to do this job. Additional affinity purification of the antibody or remaking it using a shorter polypeptide hopefully might solve this problem. An *in vitro* experiment with these P-Sal proteins could also be helpful to test their DNA binding affinity, their transposition suppression effect, and their protein binding pattern in comparison to the complete type I repressor protein.

Four mutations alter one of Rio's predicted leucine zippers.

There is a cluster of four mutations in a 10 bp sequence near the end of exon 1 in the center of the predicted 283-311 leucine zipper (Rio, 1990). Four mutations within such a short sequence is quite different from the distribution of the rest of the mutations in *P-Sal*. This may indicate that there is in fact a domain (probably a leucine zipper) encoded by this sequence. This hypothesis is supported by the fact that this amino acid sequence is conserved in Diptera (Nouaud and Anxolabéhère, 1997). The heptad pattern of predicted LLVLL leucine zipper is located at amino acid positions 283 (leucine), 290 (leucine), 297 (valine), 304 (leucine), 311 (leucine) (Rio, 1990). MPS-10 is a nonsense mutation at position 296. Second and third, MPS-11-1and MPS-11-2 have changed the central valine, at position 297 in LLVLL, to Glutamic acid. Replacing a hydrophobic amino acid with a hydrophilic one changes the heptad pattern and should abolish the function of leucine zipper. The last mutant, MPS-12, is a change of alanine to valine at position 299 and makes a stretch of three valines near the center of the zipper. Clustering of the mutations, in addition to conservation data, indicates a crucial role for this domain in the PDS activity of the type I repressor protein. However, there is also some data that fails to support a functional leucine zipper here. First, Lee *et al* (1998) showed that the sole leucine zipper of the KP protein is not necessary for the suppression of transposition. Second, none of the protein prediction programs predicted any change in this area for any of the mutants, including the four clustered mutants. In fact, SABLE-2 did not predict any change for the whole length of the MPS-11 protein.

The alternative explanation is that we are dealing with an EMS mutational hot spot, which happens to have a function, but this is unlikely to be the case. Of the 22 *P-Sal** mutations, 15 were G to A transitions. The above mentioned 10 bp area has 4 Gs and only one of them was mutated (MPS-12). MPS-10 and MPS-11-1 and MPS-11-2 are T to A transversions, which is the second most common mutation in the mutagenesis screen (6 out of 22). Three out of six T to A mutations happened to be in this small 10 bp area. It may indicate that for some unknown reason, Ts are more prone to mutation in this area. Another explanation is that this small area plays an important role in the type I repressor protein function and/ or stability maybe as an unknown enzyme motif that was not predicted by any of the computer programs.

The missense mutations in exon 2 are distributed evenly.

Exon 2 is present in *P-Sal* but absent from *KP*. However, Gloor *et al* (1993) showed that exon 2 is necessary for the type I repressors. Moreover, Bushey (2004) made a *P-Sal* construct where exon 2 can be excised using a FLP recombination system (Bushey, 2004). He inserted this construct, *P[I SalI]*, in different locations and stocked eight chromosomes with at least one modified *P[I SalI]* construct. Only one of these mutants was able to induce PDS before excision of exon 2 and its phenotype was weaker than *P-Sal*. After excision of exon 2, this PDS inducer lost its ability to induce PDS and the phenotype of PDS non-inducers did not change. This may indicate that exon 2 is necessary for PDS induction by *P-Sal*. During *P-Sal* mutagenesis I found six mutants that were mutated in this exon. These mutations were scattered evenly along the length of exon 2, suggesting that there is not a specific domain in this exon. This distribution, in addition to the fact that *KP* can induce PDS without exon 2 sequences, leads to the possibility that this exon is important for stability of the type I repressor protein rather than containing a domain with PDS specific function. This is in contrast with Rio's prediction for a third leucine zipper at position 497-525 in this exon (Rio, 1990).

Rio's predicted leucine zippers

Tang *et al* (2007) showed that transposase works as a tetramer *in vivo*. Rio (1990) predicted three different leucine zippers based on hydrophobic amino acid heptads in the type I repressor protein and suggested that they play a role in P protein hetero- or homo-dimerization. These repeats are located at following amino acid positions:

- 101 (leucine), 108 (leucine), 115 (leucine), 122 (leucine): LLLL
- 283 (leucine), 290 (leucine), 297 (valine), 304 (leucine), 311 (leucine): LLVLL
- 497 (isoleucine), 504 (leucine), 511 (leucine), 518 (glutamine), 525 (leucine): ILLQL

A study of conserved amino acid residues of *P* element (Nouaud and Anxolabéhère, 1997) found heterogeny in the first leucine zipper (101-122) which is also present in the KP protein. Lee *et al* (1996; 1998) also showed that this leucine zipper 101-122 plays a role in the KP protein dimerization but does not have any effect on suppressing transposition. Moreover, I did not find any mutation in the 101-122 leucine zipper. Further, the protein structure prediction programs (Porter, PSIPRED, and SABLE-2) do not agree on their prediction for this zipper. Porter did not predict any change in the 101-122 leucine zipper for any of the *P-Sal** mutants. PSIPRED and SABLE-2 gave similar results only for MPS-3, and MPS-7. However, MP-3 is mutated within the DNA binding domain and this is probably a better explanation for its phenotype. Putting all predictions together, it seems that the 101-122 leucine zipper does not play a role in repressor protein function.

The second zipper (283-311) is conserved (Nouaud and Anxolabéhère, 1997), and I discussed it in detail earlier. The third zipper (497-525) is also conserved (Nouaud and Anxolabéhère, 1997). MPS-17 is a 496 G->V change. Introduction of a new valine (a hydrophobic amino acid) just upstream of the first isoleucine of the zipper may change the α-helix pattern of zipper and inactivate it. MPS-18 is 522T–>I mutation and again introduces a hydrophobic amino acid in the middle of the heptad that could well affect its function. Porter predicted nine mutants including MPS-17 to have changes in the 497-525 leucine zipper, but interestingly MPS-18 is not one of them. PSIPRED found changes in this zipper only for MPS-17. And SABLE-2 confirmed the Porter finding for MPS-16 and MPS-17 and predicted MPS-18 has also changes in this zipper. Therefore all three programs agreed on a change in this area for MPS-17.

Table 6-1) Summary of findings about Rio's leucine zippers

Location at polypeptide	Amino acid sequence	Conserved	Mutations	Programs that predicted changes for this zipper
101-122	LLLL	No	-	PSIPRED, SABLE-2
283-311	LLVLL	Yes	MPS-10, MPS-11, & MPS-12	-
497-525	ILLQL	Yes	MPS-18	Porter, PSIPRED, SABLE-2

Rio's predicted helix-turn-helix motif

Rio (1990) found a weak homology between the helix-turn-helix motif of the DNA binding domain of some of the bacterial proteins and amino acid positions 308-327 of the type I repressor protein. Nouaud and Anxolabéhère (1997) reported that this motif is conserved in 12 *P* elements that were separated from seven species of Diptera. This indicates selection for this motif during evolution. MPS-13 has a mutation at position 322 (within Rio's domain) and has lost its

PDS ability, but it is still is able to suppress variegation of *P{hsp26-pt-T}39C-12* strongly and have some antimorphic effect. This indicates that the MPS-13 mutant protein is still present inside the cell. Despite evolutionary conservation of this motif, none of the protein structure prediction programs, predicted any changes in helix-turn-helix motif in any of the *P-Sal** mutants, even in MPS-13 that has a mutation within this motif.

MPS-15 is mutated in a conserved domain.

The mutation in MPS-15 changed a conserved alanine at position 404 to a threonine. This mutation introduced a novel phosphorylation site at this position, predicted by NetPhos. This mutation is in the middle of a conserved domain (discussed below) and near a kinase enzyme motif. Beall *et al* (2002) found an "ATM-family DNA damage checkpoint protein kinase" motif that phosphorylates transposase at position 405. Mutation of this motif affects the transposase activity significantly (Beall *et al*, 2002). There is a conserved six amino acid domain (ATQLFS) at position 404-409 of the *P* element (Hagemann and Pinsker, 2001). In 11 out of 12 tested *P* elements in Diptera, this domain is ATQLFS and in *Drosophila tsacasi* it is AAQLFS (Nouaud and Anxolabéhère, 1997). This domain is present in the *P* element human homolog, Phsa protein. Four out of six amino acids of this motif are also present in Pdre2 (Zebra fish, AAQLFS) and Pgga (chicken, AAPLFS) (Hammer *et al*, 2005) (Table 6-1).

Table 6-2) Results of conservation studies for amino acid positions 404-409 of the type I repressor protein

ATQLFS	404-409 type I repressor protein, *Drosophila melanogaster*
ATQLFS	P element protein, 11 out 12 tested members of the order *Diptera*
ATQLFS	Phsa, Human
AAQLFS	Pdre2, zebra fish
AAPLFS	Pgga, chicken

While the type I repressor protein and Phsa protein have a threonine at position 405 that can be phosphorylated, Pgga and Pdre2 have an alanine that can not be phosphorylated. Although MPS-15 still has this phosphorylation site, it is possible that its mutation at position 404 that introduces a new phosphorylation site affects the availability of position 405 for phosphorylation. MPS-15 still is able to suppress the variegation of *P{hsp26-pt-T}39C-12*. This indicates that the protein is still present in the cell. Although none of the protein structure prediction programs predicted any change for secondary structure of the type I repressor protein at positions 404-409 of MPS-15, the conservation of this short sequence in the wide range of species, from a transposon in the Drosophilidae family to genes in mammals and birds, points to the importance

of this short sequence for stability and/or function of the protein, and it is not unexpected that its mutation causes loss of protein function.

MPS-19 is mutated in the C-terminus of the P-Sal protein.

The last 15 amino acids of the P-Sal protein are encoded by intron 2-3, which is spliced out in the germ-line specific splicing pattern. Gloor *et al* (1993) reported that a type I repressor protein needs at least 1956 bp from the 5' end of the *P* element, *i.e.* the type I repressor has to have at least three amino acids encoded by intron 2-3 to be able to suppress transposition. Parental P-Sal protein, a typical type I repressor protein, is 12 amino acids longer than what Gloor mentioned as the minimum required sequence for a type I repressor. Therefore, it seems its last 12 amino acids do not play any significant role in the suppression of transposition. However, MPS-19 is mutated within this 12 amino acid sequence. This mutation changes a lysine at position 567 to isoleucine. Based on Nouaud and Anxolabéhère study (1997), two out of 12 studied elements have lysine at this location, three have glutamic acid, and seven have asparagine. All of these amino acids are hydrophilic while isoleucine (MPS-19 mutation) is a hydrophobic amino acid. This drastic change in C-terminus of protein may destabilize its structure that leads to its hypomorph phenotype.

MPS-19 has a hypomorphic PDS phenotype while the rest of the mutants are either amorphs or very strong hypomorphs. This mutant was also different from the other mutants in enhancing variegation of $In(1)w^{m4}$. Its phenotype is similar to the *TM6B Tb Hu* chromosome that has an *E(var)3-9* mutation and this suggests MPS-19 carries an *E(var)* mutation too. MPS-19 affects *E1* and *P{hsp26-pt-T}39C-12* to the same extent as the parental *P-Sal*, but has lost its ability to affect $P\{hsp26\text{-}pt\text{-}T\}ci^{2\text{-}m1021.R}$. The retention of the parental phenotype on *E1* and *P{hsp26-pt-T}39C-12* indicates that the C-terminus mutation has a trivial effect on protein stability and the protein is present inside the cell. Although the enhancement of variegation of *E1* can be due to the possible second site *E(var)* mutation, suppression of variegation of *P{hsp26-pt-T}39C-12* can not be explained by such an *E(var)* mutation. Segregation of this *E(var)* mutation from MPS-19 *P-Sal** will let us to study its effects on other *P-Sal** mutants and its ability to intensify their PDS phenotype. This can be the first reported *E(var)* that affects PDS.

The coding sequence is intact in MPS-20.

For MPS-20, I mapped the loss of PDS mutation within 0.27 cM of *P-Sal*. However, sequencing did not reveal any changes in the DNA sequence between base pair 23 and base pair 2819. Therefore, the promoter and the coding sequence of MPS-20 are intact. Since I already eliminated the chance of a dominant loss-of-function *Su(var)* mutation near *P-Sal* by the *Df(3)*

experiment, the cause of the amorphic phenotype of MPS-20 might be a mutation in the regulatory sequences of the *Pak3* gene that probably directs *P-Sal* expression. One way to test this would be to sequence the adjacent *Pak3* gene and regulatory sequences. An alternative way to test this is transposition of the *P[SalI]*. If the mutation is in the *Pak3* regulatory sequences and the *P[SalI]* insert is unaffected, transposed *P[SalI]* in new locations will be able to induce PDS. Lack of PDS in the new insertions will suggest the mutation is within the transposon (not in *Pak3*). Finding a mutation in *Pak3* regulatory sequence offers the opportunity to examine expression of *Pak3* gene and its promoter and regulatory sequences.

Comparing *P-Sal mutants with phylogenic studies.**

Nouaud and Anxolabéhère (1997) studied the repressor protein encoded by12 autonomous and non-autonomous *P* elements from seven species of Diptera. I compared the *P-Sal** mutants to the conserved sequences and found that out of 16 missense mutations in the coding sequence, 12 of them happened to be in amino acids that are the same at all 12 studied elements. Also, MPS-12 is mutated at an alanine that is conserved in 11 out of 12 sequences. In MPS-4 an arginine at position 65 is replaced by lysine, an amino acid with similar biochemical specification such as polarity and water affinity. In 9 out of 12 studied elements there is an arginine in this position, and in other three, it is a lysine. One of these three lysine bearing elements is an autonomous element in *Drosophila bifasciata.* It seems both amino acids in this location can be functional. This is the second amino acid at a nuclear localization signal (NLS). Maybe this slight change in the NLS is the cause of the MPS-4 amorphic phenotype. In MPS-13 the threonine at position 322 is replaced by isoleucine. Threonine is present in four studied sequences and asparagine in seven of them but one autonomous sequence has an isoleucine in that position (Nouaud and Anxolabéhère, 1997; Miller *et al*, 1999). The presence of isoleucine in this location in this functional element implies that the MPS-13 322 T>I mutation in this location should be tolerated, which is in contradiction with the fact that MPS-13 is not able to induce PDS. MPS-19 has been discussed in detail earlier.

Protein prediction programs

Changes in amino acid physical characteristics

Table 4-1 shows predicted amino acid changes based on the change in the DNA sequence. Nine changes in amino acid polarity, seven changes in amino acid water affinity, and 14 changes in amino acid tendency for secondary protein structure were found in 15 missense mutations. MPS-4 is the only missense mutation that does not have a change in any of these amino acid characteristics (Appendix I). Moreover, SIFT program predicted that this mutation

should be tolerated by the protein. Maybe the fact that MPS-4 was mutated in a nuclear localization signal, which has a very specific amino acid sequence, can explain its mutant phenotype although, as I mentioned before, phylogenic studies contradict this (Miller *et al*, 1999).

Given that these mutations affect *P-Sal* PDS induction ability, I predicted the effects of these mutations on protein enzyme motif, secondary structure, and mutation tolerance using seven computer programs: PROSITE, NetPhos 2.0, DomPred, Porter protein predict, PSIPRED, SABLE-2, and SIFT (Table 6-2).

Predicted enzyme motifs and phosphorylation sites

PROSITE and NetPhos 2.0 predicted a change in enzyme modification sites in two of the mutants: gain of a phosphorylation site at amino acid 404 for MPS-15, and loss of protein kinase C phosphorylation site at position 522 in MPS-18 (Appendix III). Although there is no reported study of post-translational modifications of the P repressor proteins at these motifs, based on the mutant phenotype it is reasonable to predict that these motifs play a significant role in the type I repressor protein function. Since the KP protein does not any of these motifs, they should play a specific role for the type I repressor protein, but not the KP repressor protein.

Beall *et al* (2002) showed P transposase activity is regulated by phosphorylation. They found eight "ATM-family DNA damage checkpoint protein kinase" motifs in transposase that affect the transposase activity significantly. These motifs are amino acid positions 41, 51, 62, 94, 96, 129, 143, and 405. Mutation S129A resulted in a hyperactive transposition and the rest of the mutants were hypoactive. I used two programs (PROSITE and NetPhos 2.0) to find potential phosphorylation sites. PROSITE predicted a total of 25 phosphorylation sites in the type I repressor protein and among them position 62 (Protein kinase C phosphorylation site) and position 94 (Casein kinase II phosphorylation site) match with the results of Beall *et al* (2002). NetPhos 2.0 predicted a total of 22 sites including positions 62 and 129 from the Beall *et al* study (2002) (Appendix II). Although I did not find any mutation in the corresponding sequences in *P-Sal** mutants, I found mutations very close to them. MPS-4 is mutated at position 65 that can affect phosphorylation at position 62. MPS-15 is mutated at position 404, adjacent to 405 and based on the NetPhos prediction program this induces a novel phosphorylation site at position 404.

Predicted changes in protein secondary structure

I used DomPred, Porter, PSIPRED, and SABLE2.0 to predict changes in secondary structure of the mutant proteins (Appendix III). Among these four programs, DomPred was the least informative program. It found new domains in just two mutants: MPS-6 and MPS-9. It

provided results in graphic format without any specification of each domain. Together, these problems made DomPred an undesirable program for prediction of structural changes in proteins with unknown secondary structure.

Porter predicted changes in 11 out of 17 tested mutants, nine of them are predicted to have changes at position 518 or 524, or both. Porter predicted similar changes for four of exon 2 mutants. MPS-14, MPS-16, MPS-17, and MPS-19 changed for both locations 518 and 524. These predictions happened to be in a small area very near C-terminus of protein. This domain was predicted by Rio (1990) as an ILLQL heptad pattern at 497, 504, 511, 518, 525 that works as a leucine zipper. In fact, Porter predicted change in protein structure at position 518 (one of the critical amino acids in the zipper) for six mutants (MPS-11, MPS-12, MPS-14, MPS-16, MPS-17, and MPS-19). It also predicted a change at position 524 which is within the zipper for MPS-3, MPS-6, and MPS-9. PSIPRED predicted MPS-17 to have changes in this zipper and SABLE-2 predicted changes within this zipper for MPS-16, MPS-17, and MPS-18. Therefore all three programs agree on MPS-17 changes in this zipper (Appendix III). On the other hand, the fact that Porter mainly picked up changes in this zipper and ignored the rest of protein indicates there is a bias within the software and its prediction scheme. Moreover, Porter did not find any change in MPS-18 which is mutated in position 522 that is within the 497-525 leucine zipper. Considering all of this, Porter is not efficient enough to predict different type of changes in the mutant type I repressor proteins.

PSIPRED 2.0 appears better than the previous two programs. It predicted some changes in all of the mutants, except MPS-6 and MPS-12, and it predicted several changes near the location of each mutation. In the missense mutants, PSIPRED predicted a change of the zinc finger in two, of the DNA binding domain in four, and of NLS in three. Changes of the 101-122 leucine zipper are predicted in five missense mutants. PSIPRED did not predict any changes for the rest of exon 1. This is in contrast to the sequencing results that predicted two domains in this area. It is also contrary to Rio's prediction (1990) for a leucine zipper at position 283-311. However, it found changes in Rio's third leucine zipper (497-525) for one mutant (MPS-17). The most common predicted change by PSIPRED 2.0 is a coil to helix change at amino acid 427 that has been predicted for eight out of 17 tested mutants. Position 427 is encoded by exon 2 and is not present in the KP protein. Although it seems interesting that eight mutants are predicted to be changed at this specific position, the biological significance of a change in the secondary structure of just one amino acid is under question.

SABLE-2 is the last program that I used to predict the secondary protein structure of the *P-Sal** mutants. It was not able to find any change in MPS-6, MPS-11, or MPS-15. However, it predicted zinc finger domain changes in three mutants (MPS-2, MPS-3, and MPS-7) and nuclear localization signal changes in MPS-3, MPS-4, and MPS-7. In Rio's (1990) predicted leucine zippers, SABLE-2 predicted the 101-122 leucine zipper changes for amorphic mutants MPS-7, MPS-12, MPS-13, MPS-17, and MPS-18. It did not predicted any change for the 283-311 leucine zipper and found changes in 497-525 leucine zipper in three mutants (MPS-16, MPS-17 and MPS-18). There are several common changes that have been predicted for four mutants or more by SABLE-2. The most common ones are helix to coil changes at amino acid 170 and amino acids 220-221 that have been predicted for ten mutant proteins (Appendix III). This program predicted more changes in the part of the type I repressor protein that is common with the KP protein than any other prediction program. Prediction of PSIPRED 2.0 and SABLE-2 was similar for MPS-3, MPS-4, MPS-7, MPS-14, and MPS-17 (Table6-2). Therefore, predictions for these five mutants are more likely to represent a biological change than predictions for other mutants.

The software that I used did not predict any changes at Rio's leucine zipper at position 283-311 even for the mutants that were mutated within this domain (Rio, 1990). They also did not predict any change in the conserved domain at 404-409 (ATQLFS) even in the MPS-15 that is mutated in this domain (Hagemann and Pinsker, 2001).

Predicted mutation tolerance based on amino acid changes

SIFT is the last program that I used to study the P-Sal* mutant proteins. SIFT is different from the other programs in that it is focused on the protein function not its structure. It calculates the chance that the protein can tolerate a specific amino acid substitution. SIFT predicted that the mutation would be tolerated in amorphic mutants MPS-4, MPS-6, and MPS-13, and hypomorphic MPS-19 (Appendix III). This result was expected for MPS-4 (arginine to lysine) since its mutation did not change amino acid polarity, water affinity, or tendency for secondary structure. Phylogenic studies support this conclusion as well (Miller *et al*, 1999). Although in the case of MPS-4, the mutation happened to be in nuclear localization signal, which likely explains its phenotype. Based on SIFT results, glycine to glutamic acid (MPS-6 mutation) and threonine to isoleucine (MPS-13 mutation) are well tolerated in nature. MPS-6, and MPS-13 do not have anything in common in amino acid changes, but phenotypically both of them are antimorphic and have ability to affect *P{hsp26-pt-T}39C-12* similar to parental *P-Sal*. MPS-19 is mutated at protein C-terminus and this mutation changed the MPS-19 phenotype to a PDS hypomorph although it still has the ability to affect *E1* and *P{hsp26-pt-T}39C*-12.

Table 6-3) Predicted reason for mutant phenotype in *P-Sal** mutants based on the results of sequencing and protein prediction softwares.

Mutant	Amino acid	Predicted structure change	Supported by
MPS-1	29 Trp>STOP	Premature protein truncation	DNA sequencing
MPS-2	30 Glu>Lys	Mutation in the DNA binding domain	DNA sequencing
MPS-3	44 Cys>Tyr	Mutation in the DNA binding domain	DNA sequencing
		Change in protein secondary structure at amino acids 21, 62-64, 84-85	• PSIPRED 2.0 • SABLE-2
MPS-4	65 Arg>Lys	Mutation in the DNA binding domain	DNA sequencing
		Mutation in Nuclear Localization Site	
		Change in protein secondary structure at amino acids 64	• PSIPRED 2.0 • SABLE-2
MPS-5	SPLICING	Change in splicing site	DNA sequencing results
MPS-6	152 Gly>Glu		
MPS-7	168 Leu>Phe	Change in protein secondary structure at amino acids 44, 47, and 99-102	• PSIPRED 2.0 • SABLE-2
MPS-8	195 Trp>STOP	Premature protein truncation	DNA sequencing
MPS-9	204 Gly>Glu	A new domain around amino acid 140	DomPred
MPS-10	296 Leu>STOP	Premature protein truncation	DNA sequencing
MPS-11	297 Val>Glu	Mutation in a new domain or enzyme modification site	DNA sequencing
MPS-12	299 Ala>Val		
MPS-13	322 Thr>Ile		
MPS-14	363 Leu>Phe	Change in protein secondary structure at amino acid 356	• PSIPRED 2.0 • SABLE-2
MPS-15	404 Ala>Thr	Gain of phosphorylation site at 404	NetPhos
MPS-16	461 Pro>Leu	Change in protein secondary structure at amino acids 523-524	• Porter • SABLE-2
MPS-17	496 Gly>Val	Change in protein secondary structure at amino acids 481, 493	• PSIPRED 2.0 • SABLE-2
MPS-18	522 Thr>Ile	Loss of a TSR PKC phosphorylation site at 522	• PROSITE motif search • NetPhos
MPS-19	567 Lys>Ile		
MPS-20	None		

P repressor protein presence in the mutants

The loss of the PDS phenotype in *P-Sal** mutants could be due to change of one of its critical functional domains. In this case, the protein is probably still present inside the cell but lacks function. The second explanation is that a change in the protein stability and reduction of the protein half-life has occurred. The fact that I was able to amplify *P-Sal* cDNA from adult flies indicates that *P-Sal* gene is expressed in them. I tried to develop antibodies to test for the presence of the repressor protein in the *P-Sal** carriers. If the protein was still present, the mutation had affected its function, and if it was not present the mutation had likely affected protein stability and its effect on any function could not be determined. I was interested in detecting all forms of the P proteins; and a polyclonal antibody against whole type I repressor protein was induced in rabbits. Even at a 1/1000 dilution, the resulting rabbit sera could detect the denatured purified protein. However, the number of non-specific bands in resulting Western blots was overwhelming. Rabbit sera could not detect the type I repressor protein, the KP protein or the transposase in the fly extracts. It is possible that either the protein expression in adults is not enough to be detected or the antibody is not specific enough to detect them. Affinity purification of post-immune sera might select for antibodies that specifically and more strongly bind to P repressor protein. The alternative solution is an antibody against a shorter peptide sequence of repressor protein to make the epitope and the antibody more specific and reduce the background.

***P-Sal** mutations fine tuned previous predictions about the type I repressor domains.**

All of the *P-Sal** mutants, except MPS-19, have lost their ability to silence *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$ and *E1*, as well as *Pci*. It suggests that the *P-Sal* product affects all three constructs, that are inserted just upstream of the *ci* gene, through a common pathway. The only exception is MPS-19. MPS-19 is mutated in a location that Gloor *et al* (1993) stated was unnecessary for the type I repressor function. MPS-19 acts as a hypomorph on *Pci*, an amorph on *P{hsp26-pt-T}ci*$^{2\text{-}m1021.R}$, and a typical *P-Sal* on *E1*. It also probably has an *E(var)* mutation because it enhances *In(1)w*^{m4} variegation. This mutant shows the importance of the C-terminus of the type I repressor protein that is encoded by the 2-3 intron. To understand the role of the protein C-terminus in function and stability of the type I protein, the *E(var)* mutation should be segregated from the *P-Sal** mutation.

Nine mutants that lost their ability to affect *Pci* were still able to silence variegation in *P{hsp26-pt-T}39C-12*. This indicates that the *P-Sal* effect on *Pci* is separable from its effect on *P{hsp26-pt-T}39C-12*. This suggests that there are two independent functions. The alternative

explanation is a difference in the sensitivity of these two systems to concentration of repressor protein function. Putting information from MPS-8, and MPS-19 together, it is more likely that the difference between PDS and suppression of *P{hsp26-pt-T}39C-12* variegation is in their sensitivity to cellular repressor concentration. The difference in response could be explained if *P{hsp26-pt-T}39C-12* is more sensitive to the presence of the repressor protein. That is, it is suppressed even at lower levels of the type I repressor protein function, while PDS requires higher levels. Therefore, if a mutation destabilizes the protein and partially reduces its cellular levels, there is still enough repressor protein inside the cell to suppress variegation in *P{hsp26-pt-T}39C-12* and change its phenotype while PDS can not be induced any more. At first glance this theory doesn't explain the opposite action (silencing in PDS vs. expression in *P{hsp26-pt-T}39C-12*). However, it can if we consider that both of these events happen through a change in chromatin structure at the loci in which the binding of the repressor protein to DNA plays a role.

Twelve out of 22 *P-Sal** mutants showed no activity in any of the genetic tests. Thus, they have lost their ability to induce PDS and to silence variegation on *P{hsp26-pt-T}39C-12*. This could indicate a significant change in protein stability that reduces protein concentration below a sensitivity threshold or a change in an important functional domain or a combination of both.

The localization of base pair changes in MPS-6 through MPS-12 putatively identifies two new domains in exon 1 of the *P* element. The first one is a 130 bp sequence in the middle of exon 1 that is present in the *KP* element as well. I found three mutations, MPS-6, MPS-7, and MPS-8, in this domain. All three of them affect *P{hsp26-pt-T}39C-12* but lost their PDS ability. One of these mutants, MPS-8, encodes a truncated polypeptide that is just 14 amino acids shorter than the KP protein. This indicates the importance of last few amino acids of the KP protein in its PDS induction. It also points to the difference between P repressor protein effects on *Pci* and *P{hsp26-pt-T}39C-12*. The second protein domain encoded by exon 1 is a 10 bp area with four mutations in the middle of the 283-311 leucine zipper (Rio, 1990). None of the protein structure prediction programs found any of these domains. It is possible that we are dealing with one or two novel protein domains, or enzyme motifs that play an important role in the process of PDS and possibly suppression of P transposition.

As it was mentioned earlier, MPS-20 was the only mutant with an intact *P-Sal* coding sequence and promoter. Its loss of PDS phenotype mapped to 0.27 cM from *P-Sal*. Since the chromosome 3 deficiencies of this region (3 cM around *P-Sal*) are unable to suppress PDS, second site modifiers are unlikely in MPS-20. Its mutation is probably a change in the sequence

that regulates *Pak3* and *P-Sal* expressions. This can be tested by mobilization of its *P[SalI]* construct to new genomic locations and test it for PDS function.

Future research about PDS and other effects of *P* elements on *P* inserts of chromosome 4

The current study showed that PDS is a dose-dependent heterochromatinization mechanism. This may indicate the presence of a general dose-dependent heterochromatinization for P cytotype and transposition control. More studies regarding the presence of such a mechanism and its effect on control of transposable elements can open a window to genome maintenance. The role of different RNAi machineries in potentially targeting heterochromatinization in this system is another subject that needs clarification. RNAi is a recent and evolving topic and testing of new RNAi mutants against PDS should give us a better idea about RNAi involvement in PDS and P cytotype. As mentioned before dose-dependent heterochromatinization can be beneficial for both organism and transposon. Screening for similar mechanisms in nature can lead us to a new era of gene expression control.

My mutagenesis of *P-Sal* showed some new domains in type I repressor protein. Some of the mutations are present in the areas that were considered unnecessary by previous studies. A more detailed description of these domains by *in vitro* mutagenesis would be useful. Do these domains really do what we thought they do is a question that should be answered to fine tune our protein domain predictions. For example, there is a conserved domain at amino acid positions 404-409 (ATQLFS) (Hagemann and Pinsker, 2001). Although it has a kinase motif in the type I repressor protein, its homologues in chicken and zebra fish do not have this motif. MPS-15, which has a mutation in this motif, still has some of the *P-Sal* functions and may help to study this domain.

*P-Sal** mutants provide an opportunity to study effects of different mutant repressor proteins on different phenotypes including suppression of transposition and somatic *P* element dependent phenotypes such as $vg^{21\text{-}3}$, sn^{w}, RIGS, and telomeric effect. These mutants would also facilitate study of some of the type II repressors. MPS-1 provides a good opportunity to study the *SP* element, and MPS-8 is very similar to the *KP* element. Although we should consider the role of nonsense mRNA decay in phenotype of these two mutants, the fact that MPS-8 still has some of the capabilities of *KP* element is an indicator of presence of some of its mRNA in the cell, and thus casts doubt on this mechanism as a general explanation for lack of PDS in these mutants.

The effect of hPEV genes on PDS is another interesting area to study. There are about 50 different known *Su(var)* genes. I already know that two of them (*Su(var)205* and *Su(var)3-7*)

affect PDS, while *Su(var)3-9* does not. I showed that *E(var)3-9* does not affect PDS either. Testing other *Su(var)* and *E(var)* mutants against PDS can open a window to how this system works as an example of heterochromatinization on chromosome 4 of *Drosophila melanogaster*.

Bibliography

ADAMCZAK, R., A. POROLLO and J. MELLER, 2005 Combining Prediction of Secondary Structure and Solvent Accessibility in Proteins. Proteins: Structure, Function and Bioinformatics 59: 467-475.

ANDERSON, K., M. DAVIS and R. HODGETTS, 2006 Novel events associated with phenotypic reversion of a *P* element mutant in *Drosophila melanogaster*. Genome 49: 1184-1192.

ANDERSON, K. J., 2007 Investigating the interactions between *P* elements and *vestigial* alleles in *Drosophila melanogaster*, PhD thesis, pp. 158 in *Molecular Biology and Genetics*. University of Alberta, Edmonton.

ANDREWS, J., and G. GLOOR, 1995 A role for the *KP* leucine zipper in regulating *P* element transposition in *Drosophila melanogaster*. Genetics 141: 587-594.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

ASHBURNER, M., and F. LEMEUNIER, 1976 Relationships within the *melanogaster* species subgroup of the genus *Drosophila (Sophophora)*. Proc. r. Soc. B. 193: 137-157.

BARTOLOMÉ, C., X. MASIDE and B. CHARLESWORTH, 2002 On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol Biol Evol 19: 926-937.

BEALL, E., A. ADMON and D. C. RIO, 1994 A *Drosophila* protein homologous to the human p70 Ku autoimmune antigen interacts with the *P* transposable element inverted repeats. Proc Natl Acad Sci U S A 91: 12681-12685.

BEALL, E., M. MAHONEY and D. RIO, 2002 Identification and analysis of a hyperactive mutant form of *Drosophila* P-element transposase. Genetics 162: 217-227.

BEALL, E., and D. RIO, 1997 *Drosophila* P-element transposase is a novel site-specific endonuclease. Genes Dev 11: 2137-2151.

BIER, E., H. VAESSIN, S. SHEPHERD, K. LEE, K. MCCALL *et al.*, 1989 Searching for pattern and mutation in the *Drosophila* genome with a *P-lacZ* vector. Genes Dev. 3: 1273-1287.

BIESSMANN, H., S. PRASAD, M. WALTER and J. MASON, 2005 Euchromatic and heterochromatic domains at *Drosophila* telomeres. Biochem Cell Biol 83: 477-485.

BLACK, D. M., M. S. JACKSON, M. G. KIDWELL and G. A. DOVER, 1987 *KP* elements repress *P*-induced hybrid dysgenesis in *Drosophila melanogaster*. EMBO J. 6: 4125–4135.

BLOM, N., S. GAMMELTOFT and S. BRUNAK, 1999 Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology 294: 1351-1362.

BUSHEY, D., 2004 Characterizing *P* element dependent silencing in *Drosophila melanogaster*, PhD thesis, pp. 302 in *Faculty of Science*. University of Alberta, Edmonton.

BUSHEY, D., and J. LOCKE, 2004 Mutations in *Su(var)205* and *Su(var)3-7* Suppress *P*-Element-Dependent Silencing in *Drosophila melanogaster*. Genetics 168: 1395-1411.

CASTRO, J., and C. CARARETO, 2004 *Drosophila melanogaster P* transposable elements: mechanisms of transposition and regulation. Genetica 121: 107-118.

CLOUAIRE, T., M. ROUSSIGNE, V. ECOCHARD, C. MATHE, F. AMALRIC *et al.*, 2005 The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. Proc Natl Acad Sci U S A 102: 6907-6912.

CRAYMER, L., 1984 New mutants report. D. I. S. 60: 234-236.

DAVIES, R. W., 2001 Sequence 285 from Patent WO0118547, pp. GenBank/EMBL/DDBJ.

DIMITRI, P., N. CORRADINI, F. ROSSI and F. VERNÌ, 2005 The paradox of functional heterochromatin. Bioessays 27: 29-41.

DORER, D., and S. HENIKOFF, 1994 Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. Cell 77: 993-1002.

DORER, D., and S. HENIKOFF, 1997 Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans*. Genetics 147: 1181-1190.

EBERL, D., B. DUYF and A. HILLIKER, 1993 The role of heterochromatin in the expression of a heterochromatic gene, the rolled locus of *Drosophila melanogaster*. Genetics 134: 277-292.

EBERT, A., S. LEIN, G. SCHOTTA and G. REUTER, 2006 Histone modification and the control of heterochromatic gene silencing in *Drosophila*. Chromosome Res 14: 377-392.

ENGELS, W., W. BENZ, C. PRESTON, P. GRAHAM, R. PHILLIS *et al.*, 1987 Somatic effects of *P* element activity in *Drosophila melanogaster*: pupal lethality. Genetics 117: 745-757.

ENGELS, W., and C. PRESTON, 1980 Components of hybrid dysgenesis in a wild population of *Drosophila melanogaster*. Genetics 95: 111-128.

ENGELS, W., C. PRESTON, P. THOMPSON and W. EGGLESTON, 1989 In situ hybridization of *Drosophila* polytene chromosomes with digoxigenin-dUTP-labeled probes. Trends Genet 5: 366.

EPHRUSSI, and HEROLD, 1944 Studies of Eye Pigments of *Drosophila.* Methods of extraction and quantitative estimation of the pigment components. Genetics 29: 148-175.

FANTI, L., D. DORER, M. BERLOCO, S. HENIKOFF and S. PIMPINELLI, 1998 Heterochromatin protein 1 binds transgene arrays. Chromosoma 107: 286-292.

GATTIKER, A., E. GASTEIGER and A. BAIROCH, 2002 ScanProsite: a reference implementation of a PROSITE scanning tool. Applied Bioinformatics: 107-108.

GLOOR, G., and W. R. ENGELS, 1992 Single-Fly Preps for PCR. DIS 71: 148-149.

GLOOR, G. B., C. R. PRESTON, D. M. JOHNSON-SCHLITZ, N. A. NASSIF, R. W. PHILLIS *et al.*, 1993 Type I Repressors of *P* Element Mobility. Genetics 135: 81-95.

GOTTLIEB, T., JACKSON, SP, 1993 The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. Cell 72: 131-142.

GREWAL, S., and S. ELGIN, 2007 Transcription and RNA interference in the formation of heterochromatin. Nature 447: 399-406.

HAGEMANN, S., and W. PINSKER, 2001 *Drosophila P* transposons in the human genome? Mol Biol Evol 18: 1979-1982.

HALEY, K., J. STUART, J. RAYMOND, J. NIEMI and M. SIMMONS, 2005 Impairment of cytotype regulation of *P*-element activity in *Drosophila melanogaster* by mutations in the *Su(var)205* gene. Genetics 171: 583-595.

HAMMER, S., S. STREHL and S. HAGEMANN, 2005 Homologs of *Drosophila P* transposons were mobile in zebrafish but have been domesticated in a common ancestor of chicken and human. Mol Biol Evol 22: 833-844.

HAYNES, K., E. GRACHEVA and S. ELGIN, 2007 A Distinct type of heterochromatin within *Drosophila melanogaster* chromosome 4. Genetics 175: 1539-1542.

HENIKOFF, S., 1997 Nuclear organization and gene expression: homologous pairing and long-range interactions. Curr Opin Cell Biol 9: 388-395.

HENIKOFF, S., and L. COMAI, 1998 Trans-sensing effects: the ups and downs of being together. Cell 93: 329-332.

HODGETTS, R., and S. O'KEEFE, 2001 The mutant phenotype associated with *P*-element alleles of the *vestigial* locus in *Drosophila melanogaster* may be caused by a readthrough transcript initiated at the *P*-element promoter. Genetics 157: 1665-1672.

HOSKINS, R., C. SMITH, J. CARLSON, A. CARVALHO, A. HALPERN *et al.*, 2002 Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. Genome Biol 3(12: RESEARCH0085.

HUANG, N., 2005 Characterization of *cis* and *trans* silencing of a *P{lacW}* transgene on chromosome 4 of *Drosophila melanogaster*, MSc thesis, pp. 98 in *Molecular Biology and Genetics*. University of Alberta, Edmonton.

ITOH, M., M. IWABUCHI, K. YOSHIDA and S. HORI, 1989 Four tandem defective *P* elements associated with positive regulation of the *Drosophila melanogaster* glucose-6-phosphate dehydrogenase gene. Biochem Genet. 27: 699-718.

ITOH, M., N. TAKEUCHI, M. YAMAGUCHI, M. YAMAMOTO and I. BOUSSY, 2007 Prevalence of full-size *P* and *KP* elements in North American populations of *Drosophila melanogaster*. Genetica 131: 21-28.

JAMES, T., J. EISSENBERG, C. CRAIG, V. DIETRICH, A. HOBSON *et al.*, 1989 Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of *Drosophila*. Eur J Cell Biol 50: 170-180.

JONES, D., 1999 Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195-202.

JOSSE, T., A. BOIVIN, D. ANXOLABÉHÈRE and S. RONSSERAY, 2002 *P* element-encoded regulatory products enhance Repeat-Induced Gene Silencing (RIGS) of *P-lacZ-white* clusters in *Drosophila melanogaster*. Mol Genet Genomics 268: 311-320.

KARESS, R., and G. RUBIN, 1984 Analysis of *P* transposable element functions in *Drosophila*. Cell 38: 135-146.

KAUFMAN, P., R. DOLL and D. RIO, 1989 *Drosophila P* element transposase recognizes internal *P* element DNA sequences. Cell 59: 359-371.

KAUFMAN, P. D., and D. C. RIO, 1991 *Drosophila P*-Element Transposase is a Transcriptional Repressor *in vitro*. PNAS 88: 2613-2617.

KE, R., R. JD and S. MJ, 1993 Repression of hybrid dysgenesis in *Drosophila melanogaster* by individual naturally occurring *P* elements. Genetics 133: 605-622.

KELLUM, R., and B. ALBERTS, 1995 Heterochromatin protein 1 is required for correct chromosome segregation in *Drosophila* embryos. J Cell Sci 108: 1419-1431.

KIGER, A., B. BAUM, S. JONES, M. JONES, A. COULSON *et al.*, 2003 A functional genomic analysis of cell morphology using RNA interference. J Biol 2: 27.

LABRADOR, M., and V. G. CORCES, 1997 Transposable element-host interactions: regulation of insertion and excision. Annu Rev Genet 31: 381-404.

LASKI, F. A., D. C. RIO and G. M. RUBIN, 1986 Tissue specificity of *Drosophila P* element transposition is regulated at the level of mRNA splicing. Cell 44: 7-19.

LEE, C., Y. MUL and D. RIO, 1996 The *Drosophila P*-element KP repressor protein dimerizes and interacts with multiple sites on *P*-element DNA. Mol Cell Biol 16: 5616-5622.

LEE, C. C., E. L. BEALL and D. C. RIO, 1998 DNA binding by the KP repressor protein inhibits P-element transposase activity *in vitro*. EMBO J. 17: 4166-4174.

LEMAITRE, B., and D. COEN, 1991 P Regulatory Products Repress *in vivo* the P Promoter Activity in *P-lacZ* Fusion Genes. PNAS 88: 4419-4423.

LEMAITRE, B., S. RONSSERAY and D. COEN, 1993 Maternal repression of the *P* element promoter in the germline of *Drosophila melanogaster*: a model for the P cytotype.

Genetics 135: 149-160.

LIEW, C., M. CROSSLEY, J. MACKAY and H. NICHOLAS, 2007 Solution structure of the THAP domain from *Caenorhabditis elegans* C-terminal binding protein (CtBP). J Mol Biol 366: 382-390.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press Inc, San Diego.

LOCKE, J., and K. TARTOF, 1994 Molecular analysis of *cubitus interruptus* (*ci*) mutations suggests an explanation for the unusual *ci* position effects. Mol Gen Genet 243: 234-243.

LU, B., P. EMTAGE, B. DUYF, A. HILLIKER and J. EISSENBERG, 2000 Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in *Drosophila*. Genetics 155: 699-708.

MARIN, L., M. LEHMANN, D. NOUAUD, H. IZAABEL, D. ANXOLABÉHÈRE *et al.*, 2000 *P*-Element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric *P* copy. Genetics 155: 1841-1854.

MARSDEN, R. L., L. J. MCGUFFIN and D. T. JONES, 2002 Rapid protein domain assignment from amino acid sequence using predicted secondary structure. Protein Science 11: 2814-2824.

MILLER, W., J. MCDONALD, D. NOUAUD and D. ANXOLABÉHÈRE, 1999 Molecular domestication--more than a sporadic episode in evolution. Genetica 107: 197-207.

MISRA, S., R. BURATOWSKI, T. OHKAWA and D. RIO, 1993 Cytotype control of *Drosophila melanogaster P* element transposition: genomic position determines maternal repression. Genetics 135: 785-800.

MISRA, S., and D. RIO, 1990 Cytotype control of *Drosophila P* element transposition: the 66 kd protein is a repressor of transposase activity. Cell 62: 269-284.

MORRISON, D., M. MURAKAMI and V. CLEGHON, 2000 Protein kinases and phosphatases in the *Drosophila* genome. J Cell Biol 150: F57-62.

MUNROE, D. J., 1996 IRS-Bubble PCR: An Effective Method for Representative Amplification of Human Genomic DNA Sequences from Complex Sources. Methods 9: 106-112.

NAKAYAMA, J., J. RICE, B. STRAHL, C. ALLIS and S. GREWAL, 2001a Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292: 110-113.

NAKAYAMA, J., J. RICE, B. STRAHL, C. ALLIS and S. GREWAL, 2001b Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292: 110-113.

NELSON, M. R., A. M. LEIDAL and C. A. SMIBERT, 2004 *Drosophila Cup* is an eIF4E-binding protein that functions in Smaug-mediated translational repression. EMBO J. 23:

150-159.

NG, P. C., and S. HENIKOFF, 2001 SIFT: Predicting deleterious substitutions. Genome Res. 11: 863-874.

NIEMI, J., J. RAYMOND, R. PATREK and M. SIMMONS, 2004 Establishment and maintenance of the P cytotype associated with telomeric *P* elements in *Drosophila melanogaster*. Genetics 166: 255-264.

NOUAUD, D., and D. ANXOLABÉHÈRE, 1997 *P* element domestication: a stationary truncated *P* element may encode a 66-kDa repressor-like protein in the *Drosophila montium* species subgroup. Mol Biol Evol 14: 1132-1144.

O'HARE, K., and G. M. RUBIN, 1983 Structures of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell 34: 25.

OKAMURA, K., A. ISHIZUKA, H. SIOMI and M. C. SIOMI, 2004 Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev 18: 1655-1666.

PINSKER, W., E. HARING, S. HAGEMANN and W. J. MILLER, 2001 The evolutionary life history of *P* transposons: from horizontal invaders to domesticated neogenes. Chromosoma 110: 148-158.

POLLASTRI, G., and A. MCLYSAGHT, 2005 Porter: a new, accurate server for protein secondary structure prediction. Bioinformatics 21: 1719-1720.

QUESNEVILLE, H., D. NOUAUD and D. ANXOLABEHERE, 2005 Recurrent recruitment of the THAP DNA-binding domain and molecular domestication of the *P*-transposable element. Mol Biol Evol 22: 741-746.

RASMUSSON KE, RAYMOND JD and S. MJ, 1993 Repression of hybrid dysgenesis in *Drosophila melanogaster* by individual naturally occurring *P* elements. Genetics 133: 605-622.

REISS, D., T. JOSSE, D. ANXOLABÉHÈRE and S. RONSSERAY, 2004 *aubergine* mutations in *Drosophila melanogaster* impair P cytotype determination by telomeric *P* elements inserted in heterochromatin. Mol Genet Genomics 272: 336-343.

REUTER, G., and P. SPIERER, 1992 Position effect variegation and chromatin proteins. BioEssays 14: 605-612.

RIDDLE, N. C., and S. C. R. ELGIN, 2008 A Role for RNAi in Heterochromatin Formation in *Drosophila*, pp. 185-209 in *RNA Interference*, edited by P. J. PADDISON and P. K. VOGT. Springer-Verlag, Berlin Heidelberg.

RILEY, J., R. BUTLER, D. OGILVIE, R. FINNIEAR, D. JENNER *et al.*, 1990 A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res 18: 2887-2890.

RIO, D., 1990 Molecular mechanisms regulating *Drosophila P* element transposition. Annu Rev

Genet 24: 543-578.

RIO, D. C., F. A. LASKI and G. M. RUBIN, 1986 Identification and immunochemical analysis of biologically active *Drosophila* P element transposase. Cell 44: 21-32.

RIO, D. C., and G. M. RUBIN, 1988 Identification and purification of a *Drosophila* protein that binds to the terminal 31-base-pair inverted repeats of the *P* transposable element. Proc Natl Acad Sci U S A 85: 8929-8933.

ROBERTSON, H., C. PRESTON, R. PHILLIS, D. JOHNSON-SCHLITZ, W. BENZ *et al.*, 1988 A stable genomic source of P element transposase in *Drosophila melanogaster*. Genetics 118: 461-470.

ROBERTSON, H. M., and W. R. ENGELS, 1989 Modified *P* Elements That Mimic the P Cytotype in *Drosophila melanogaster*. Genetics 123: 815-824.

ROCHE, S. E., and D. C. RIO, 1998 Trans-Silencing by *P* Elements Inserted in Subtelomeric Heterochromatin Involves the *Drosophila* Polycomb Group Gene, *Enhancer of zeste*. Genetics 149: 1839-1855.

RONSSERAY, S., D. ANXOLABÉHÈRE and G. PÉRIQUET, 1984 Hybrid dysgenesis in *Drosophila melanogaster*: influence of temperature on cytotype determination in the P-M system. Mol Gen Genet 196: 17-23.

RONSSERAY, S., A. BOIVIN and D. ANXOLABÉHÈRE, 2001 *P*-Element repression in *Drosophila melanogaster* by variegating clusters of *P-lacZ-white* transgenes. Genetics 159: 1631-1642.

RONSSERAY, S., T. JOSSE, A. BOIVIN and D. ANXOLABÉHÈRE, 2003 Telomeric transgenes and trans-silencing in *Drosophila*. Genetica 117: 327-335.

RONSSERAY, S., L. MARIN, M. LEHMANN and D. ANXOLABÉHÈRE, 1998 Repression of Hybrid Dysgenesis in *Drosophila melanogaster* by Combinations of Telomeric *P*-Element Reporters and Naturally Occurring *P* Elements. Genetics 149: 1857-1866.

ROUSSIGNE, M., S. KOSSIDA, A. LAVIGNE, T. CLOUAIRE, V. ECOCHARD *et al.*, 2003 The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase. Trends Biochem Sci 28: 66-69.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Habour Laboratory Press, New York.

SAROT, E., G. PAYEN-GROSCHENE, A. BUCHETON and A. PELISSON, 2004 Evidence for a *piwi*-dependent RNA silencing of the *gypsy* endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. Genetics 166: 1313-1321.

SAVITSKY, M., O. KRAVCHUK, L. MELNIKOVA and P. GEORGIEV, 2002 Heterochromatin protein 1 is involved in control of telomere elongation in *Drosophila melanogaster*. Mol Cell Biol 22: 3204-3218.

SCHEID, M. P., K. M. SCHUBERT and V. DURONIO, 1999 Regulation of Bad Phosphorylation and Association with Bcl-xL by the MAPK/Erk Kinase. J Biol Chem 274: 31108-31113.

SCHOTTA, G., A. EBERT, V. KRAUSS, A. FISCHER, J. HOFFMANN *et al.*, 2002 Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J 21: 1121-1131.

SHYU, A., M. WILKINSON and A. VAN HOOF, 2008 Messenger RNA regulation: to translate or to degrade. EMBO J 27: 471-481.

SIMMONS, M., K. HALEY, C. GRIMES, J. RAYMOND and J. FONG, 2002a Regulation of P-element transposase activity in *Drosophila melanogaster* by *hobo* transgenes that contain *KP* elements. Genetics 161: 205-215.

SIMMONS, M., K. HALEY, C. GRIMES, J. RAYMOND and J. NIEMI, 2002b A *hobo* transgene that encodes the *P*-element transposase in *Drosophila melanogaster*: autoregulation and cytotype control of transposase activity. Genetics 161: 195-204.

SIMMONS, M., K. HALEY and S. THOMPSON, 2002c Maternal transmission of *P* element transposase activity in *Drosophila melanogaster* depends on the last *P* intron. Proc Natl Acad Sci U S A 99: 9306-9309.

SIMMONS, M., J. NIEMI, D. RYZEK, C. LAMOUR, J. GOODMAN *et al.*, 2007a Cytotype Regulation by Telomeric *P* Elements in *Drosophila melanogaster*: Interactions with *P* Elements from M's Strains. Genetics 176: 1957-1966.

SIMMONS, M., J. RAYMOND, J. NIEMI, J. STUART and P. MERRIMAN, 2004 The P cytotype in *Drosophila melanogaster*: a maternally transmitted regulatory state of the germ line associated with telomeric *P* elements. Genetics 166: 243-254.

SIMMONS, M. J., D.-F. RYZEK, C. LAMOUR, J. W. GOODMAN, N. E. KUMMER *et al.*, 2007b Cytotype Regulation by Telomeric *P* Elements in *Drosophila melanogaster*: Evidence for Involvement of an RNA Interference Gene. Genetics 176: 1945-1955.

SPOFFORD, J. B., 1976 *Position-effect variegation in Drosophila*. Academic Press, London.

SUN, F.-L., M. H. CUAYCONG, C. A. CRAIG, L. L. WALLRATH, J. LOCKE *et al.*, 2000 The fourth chromosome of *Drosophila melanogaster*: Interspersed euchromatic and heterochromatic domains. PNAS 97: 5340-5345.

SVED, J. A., 1987 Hybrid dysgenesis in *Drosophila melanogaster*: evidence from sterility and southern hybridization tests that P cytotype is not maintained in the absence of chromosomal *P* factors. Genetics 115: 121-127.

TALBERT, P., and S. HENIKOFF, 2006 Spreading of silent chromatin: inaction at a distance. Nat Rev Genet 7: 793-803.

TANG, M., C. CECCONI, C. BUSTAMANTE and D. RIO, 2007 Analysis of *P* element transposase protein-DNA interactions during the early stages of transposition. J Biol

Chem 282: 29002-29012.

TANG, M., C. CECCONI, H. KIM, C. BUSTAMANTE and D. RIO, 2005 Guanosine triphosphate acts as a cofactor to promote assembly of initial *P*-element transposase-DNA synaptic complexes. Genes Dev 19: 1422-1425.

TARTOF, K. D., C. BISHOP, M. JONES, C. A. HOBBS and J. LOCKE, 1989 Towards an understanding of position effect variegation. Developmental Genetics 10: 162-176.

TZENG, T., C. LEE, L. CHAN and C. SHEN, 2007 Epigenetic regulation of the *Drosophila* chromosome 4 by the histone H3K9 methyltransferase dSETDB1. Proc Natl Acad Sci U S A 104: 12691-12696.

WALLRATH, L., and S. ELGIN, 1995 Position effect variegation in *Drosophila* is associated with an altered chromatin structure. Genes Dev 9: 1263-1277.

WEILER, K., and B. WAKIMOTO, 1998 Chromosome rearrangements induce both variegated and reduced, uniform expression of heterochromatic genes in a development-specific manner. Genetics 149: 1451-1464.

WEILER, K. S., 2002 Identification of the *E(var)3-9* gene, pp. in *Drosophila research conference 43*, San Diego.

WEINERT, B., B. MIN and D. RIO, 2005 *P* element excision and repair by non-homologous end joining occurs in both G1 and G2 of the cell cycle. DNA Repair (Amst) 4: 171-181.

WILLIAMS, J., S. PAPPU and J. BELL, 1988 Molecular analysis of hybrid dysgenesis-induced derivatives of a *P*-element allele at the *vg* locus. Mol Cell Biol 8: 1489-1497.

WOOK, K., K. JEONG-SOO and S. DONG-JIK, 1996 Molecular Analysis of *KP* Elements Derived from Korean Populations of *Drosophila melanogaster*. Korean J. Zool 39: 164-172.

Appendix I: Classifications of Amino Acids

Amino acids codes

One letter codes	Three letter codes	Name
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Classifications of amino acids

Table 1) Classification of amino acids based on their polarity.

Polar amino acids were sub classified based on their side chain to Acidic (carboxylic side chain) and Basic (amino group side chain).

Polarity	Amino Acid
Nonpolar	Glycine Alanine Valine Leucine Isoleucine Proline Methionine Phenylalanine Tryptophan
Polar	Serine Threonine Asparagine Glutamine Cysteine Tyrosine
Acidic (Polar)	Aspartic Acid Glutamic Acid
Basic (Polar)	Lysine Arginine Histidine

Table 2) Classification of amino acids based on number of their carboxylic and amino group side chains.

Superstructure	Structure	Amino Acid
Monoamino, moncarboxylic		Glycine Alanine
	Unsubstituted	Valine Leucine Isoleucine
	Heterocyclic	Proline Phenylalanine
	Aromatic	Tyrosine Tryptophan
	Thioether	Methionine
	Hydroxy	Serine Threonine
	Mercapto	Cysteine
	Carboxamide	Asparagine Glutamine
Monamino, dicarboxylic		Aspartate Glutamate
Diamino, monocarboxylic		Lysine Arginine Histidine

Table 3) Classification of amino acids based on structure of their side chain structure.

Side Chain	Amino Acid
Alphatic (do not contain N,O,S in side chain)	Glycine Alanine Valine Leucine Isoleucine
Sulfur-containing	Cysteine Methionine
Aromatic (benzene ring in side chain)	Phenylalanine Tyrosine Tryptophan
Neutral (hydroxyl or amide groups in side chain)	Serine Threonine Asparagine Glutamine
Acidic (carboxylate groups in side chain)	Aspartic acid Glutamic acid
Basic	Lysine Arginine
Imino acid (special case)	Proline

Appendix II: Results of protein prediction programs for the parental P-Sal protein

PROSITE motif search

Pattern-ID: **ASN_GLYCOSYLATION**
Pattern-DE: N-glycosylation site
Pattern: N[^P][ST][^P]
49 NDSQ
105 NKSL
376 NKSD
410 NTTA
428 NATE
501 NASL
514 NFSM

Pattern-ID: **CAMP_PHOSPHO_SITE**
Pattern-DE: cAMP- and cGMP-dependent protein kinase phosphorylation site
Pattern: [RK]{2}.[ST]
361 KKLT

Pattern-ID: **PKC_PHOSPHO_SITE**
Pattern-DE: Protein kinase C phosphorylation site
Pattern: [ST].[RK]
62 TFK
107 SLR
135 SLR
265 SWK
364 TKK
415 SIR
522 TSR
536 SMR

Pattern-ID: **CK2_PHOSPHO_SITE**
Pattern-DE: Casein kinase II phosphorylation site
Pattern: [ST].{2}[DE]
36 SLGE
94 TQTE
107 SLRE
197 SDVD
242 SSAD
276 TRMD
367 TIQE

Pattern-ID: **MYRISTYL**
Pattern-DE: N-myristoylation site
Pattern: G[^EDRKHPFYW].{2}[STAGCN][^P]
34 GCSLGE
92 GSTQTE
151 GGQRAT
217 GVDDAD
261 GLKKSW
316 GISESK
355 GLTING
496 GIIVNN
535 GSMRSR

Pattern-ID: **AMIDATION**
Pattern-DE: Amidation site
Pattern: .G[RK][RK]
359 NGKK

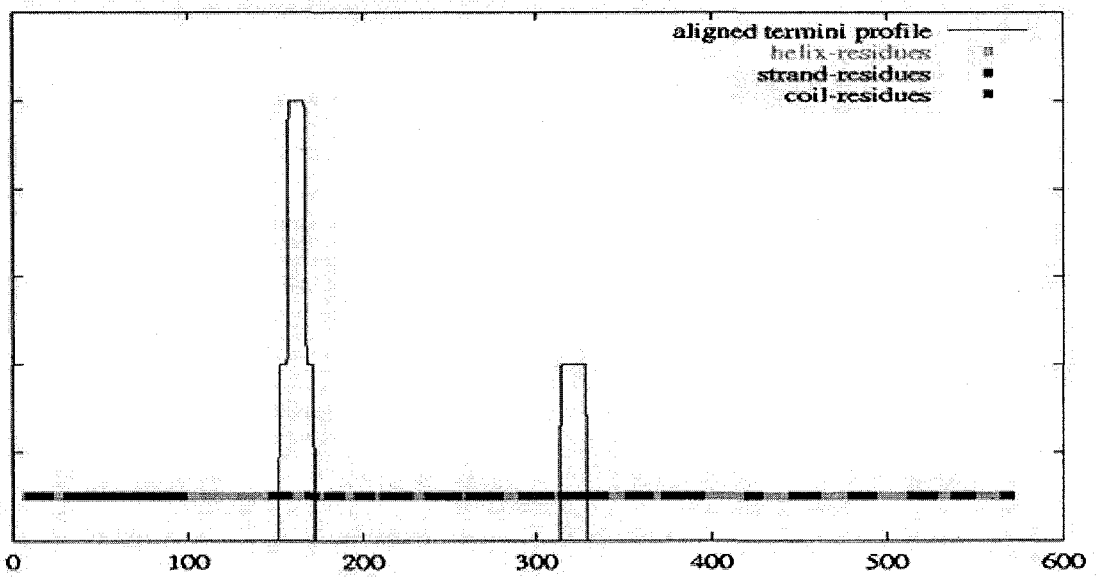
Pattern-ID: **LEUCINE_ZIPPER**
Pattern-DE: Leucine zipper pattern
Pattern: L.{6}L.{6}L.{6}L
101 LFNENKSLREKIRTLEYEMRRL

NetPhos 2.0

Predicted phosphorylation sites for type I repressor protein:

- Ser at positions: 91, 107, 129, 135, 197, 215, 242, 265, 318, 414, 415, 451, 477, 539
- Thr at positions: 62, 191, 522
- Tyr at positions: 89, 240, 253, 351, 424

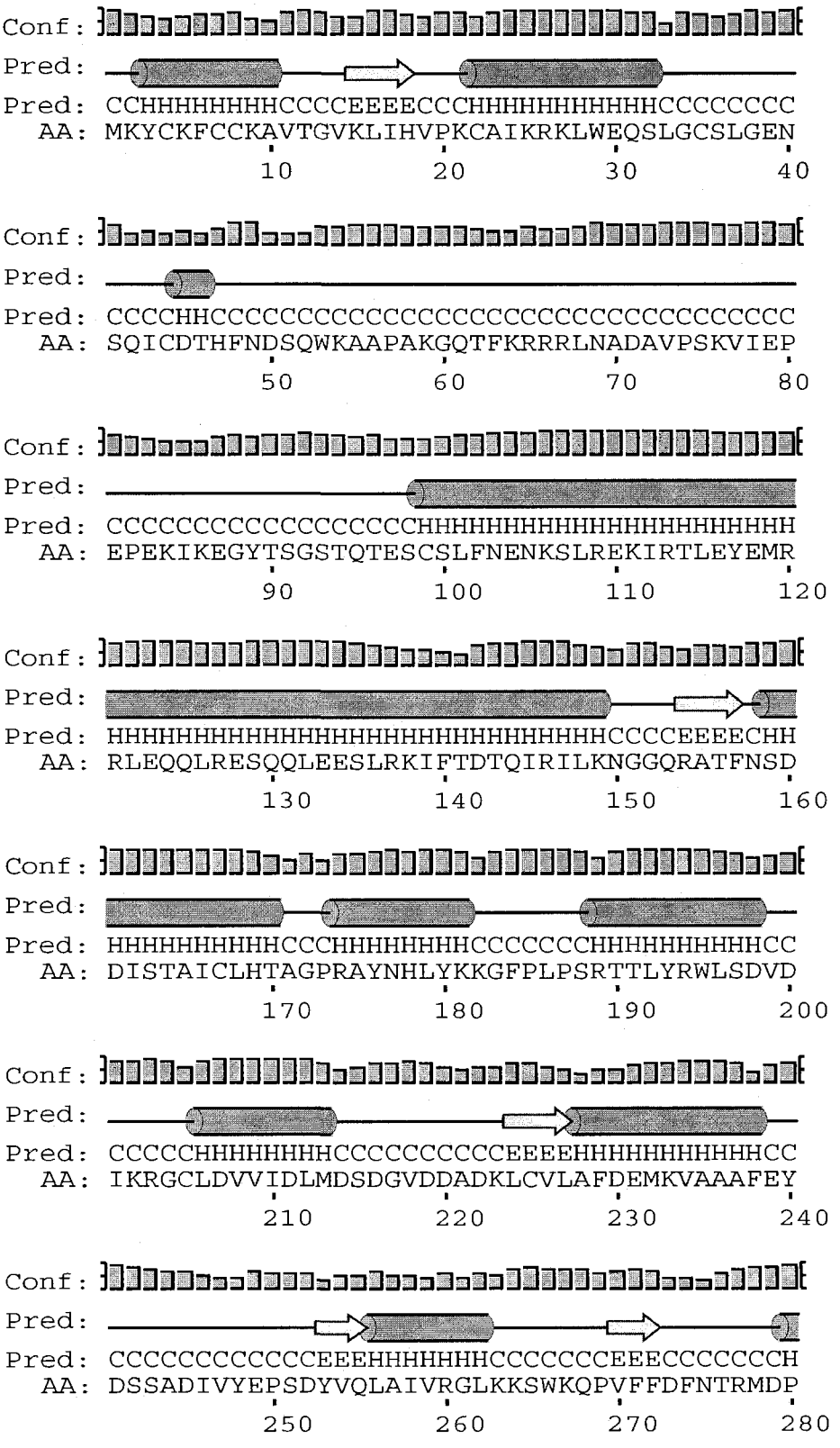
DomPred

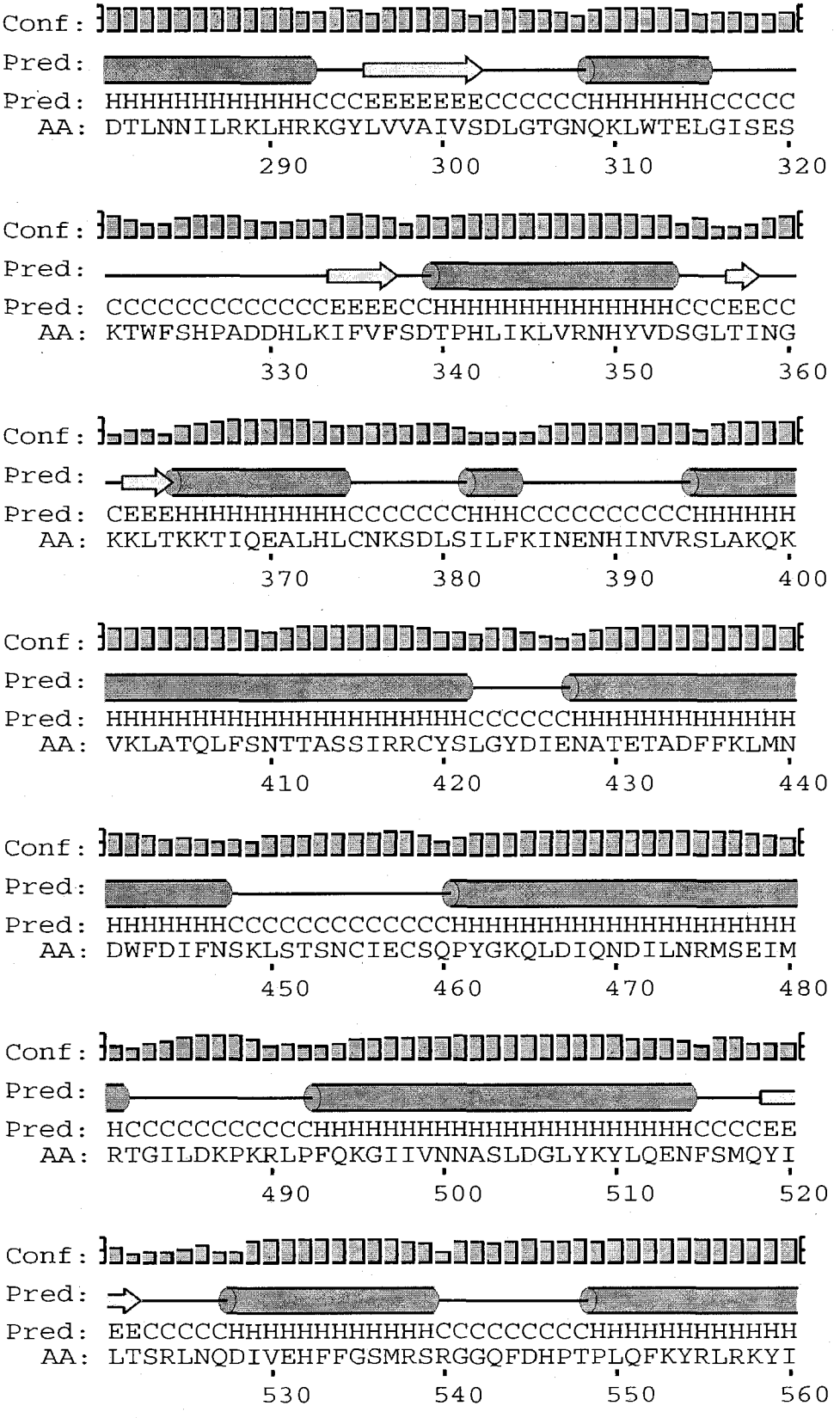


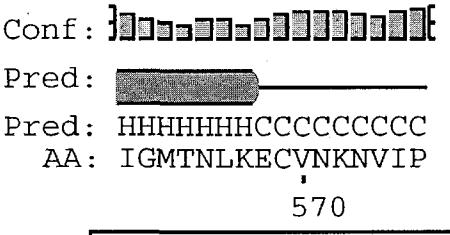
Porter

```
MKYCKFCCKAVTGVKLIHVPKCAIKRKLWEQSLGCSLGENSQICDTHFNDSQWKAAPAKG
CCCCCCCCCCCCCCEEEECCCCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCC

QTFKRRRLNADAVPSKVIEPEPEKIKEGYTSGSTQTESCSLFNENKSLREKIRTLEYEMR
CEEEHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHH

RLEQQLRESQQLEESLRKIFTDTQIRILKNGGQRATFNSDDISTAICLHTAGPRAYNHLY
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCEEEECHHHHHHHHHHHHCCHHHHHHH

KKGFPLPSRTTLYRWLSDVDIKRGCLDVVIDLMDSDGVDDADKLCVLAFDEMKVAAAFEY
HCCCCCCCHHHHHHHHHHCCCCCCHHHHHHHHHCCCCCCCCCCHHHHHHHHHHHHHHHHC

DSSADIVYEPSDYVQLAIVRGLKKSWKQPVFFDFNTRMDPDTLNNILRKLHRKGYLVVAI
CCCCCCCCCCCCCHEEEEEHHCCCCCCCCCCEEEECCCCCCHHHHHHHHHHHHHHCCCCEEEE

VSDLGTGNQKLWTELGISESKTWFSHPADDHLKIFVFSDTPHLIKLVRNHYVDSGLTING
EECCCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCEEEECCHHHHHHHHHHHHHHHCCCEECC

KKLTKKTIQEALHLCNKSDLSILFKINENHINVRSLAKQKVKLATQLFSNTTASSIRRCY
CCCCHHHHHHHHHHHCCCCCCCEEEECCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHH

SLGYDIENATETADFFKLMNDWFDIFNSKLSTSNCIECSQPYGKQLDIQNDILNRMSEIM
HCCCCCHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHH

RTGILDKPKRLPFQKGIIVNNASLDGLYKYLQENFSMQYILTSRLNQDIVEHFFGSMRSR
HCCCCCCCCCCCCHHHHHHCHHHHHHHHHHHHHHCCCCEEECCCHCHHHHHHHHHHHHHC

GGQFDHPTPLQFKYRLRKYIIGMTNLKECVNKNVIP
CCCCCCCCHHHHHHHHHHHHHHHHHHCCCCCCCCCCC
```

PSIPRED


Conf:
Pred:
Pred: CCHHHHHHHHCCCCEEEECCCHHHHHHHHHHHCCCCCCCC
AA: MKYCKFCCKAVTGVKLIHVPKCAIKRKLWEQSLGCSLGEN
10 20 30 40
Conf:
Pred:
Pred: CCCCHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: SQICDTHFNDSQWKAAPAKGQTFKRRRLNADAVPSKVIEP
50 60 70 80
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHH
AA: EPEKIKEGYTSGSTQTESCSLFNENKSLREKIRTLEYEMR
90 100 110 120
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCEEEECHH
AA: RLEQQLRESQQLEESLRKIFTDTQIRILKNGGQRATFNSD
130 140 150 160
Conf:
Pred:
Pred: HHHHHHHHHHCCCHHHHHHHHCCCCCCCHHHHHHHHHHCC
AA: DISTAICLHTAGPRAYNHLYKKGFPLPSRTTLYRWLSDVD
170 180 190 200
Conf:
Pred:
Pred: CCCCCHHHHHHHHCCCCCCCCCCEEEEHHHHHHHHHHHHCC
AA: IKRGCLDVVIDLMDSDGVDDADKLCVLAFDEMKVAAAFEY
210 220 230 240
Conf:
Pred:
Pred: CCCCCCCCCCCCEEEHHHHHHHCCCCCCCEEECCCCCCCH
AA: DSSADIVYEPSDYVQLAIVRGLKKSWKQPVFFDFNTRMDP
250 260 270 280


Conf:
Pred:
Pred: HHHHHHHHHHHHCCCEEEEEEECCCCCCHHHHHHHCCCCC
AA: DTLNNILRKLHRKGYLVVAIVSDLGTGNQKLWTELGISES
290 300 310 320
Conf:
Pred:
Pred: CCCCCCCCCCCCCEEEECCHHHHHHHHHHHHHHCCCEECC
AA: KTWFSHPADDHLKIFVFSDTPHLIKLVRNHYVDSGLTING
330 340 350 360
Conf:
Pred:
Pred: CEEEHHHHHHHHHHCCCCCCCHHHCCCCCCCCCCHHHHHH
AA: KKLTKKTIQEALHLCNKSDLSILFKINENHINVRSLAKQK
370 380 390 400
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHCCCCCCHHHHHHHHHHHH
AA: VKLATQLFSNTTASSIRRCYSLGYDIENATETADFFKLMN
410 420 430 440
Conf:
Pred:
Pred: HHHHHHHCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHH
AA: DWFDIFNSKLSTSNCIECSQPYGKQLDIQNDILNRMSEIM
450 460 470 480
Conf:
Pred:
Pred: HCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHCCCCEE
AA: RTGILDKPKRLPFQKGIIVNNASLDGLYKYLQENFSMQYI
490 500 510 520
Conf:
Pred:
Pred: EECCCCCHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHH
AA: LTSRLNQDIVEHFFGSMRSRGGQFDHPTPLQFKYRLRKYI
530 540 550 560



Legend:

= helix

= strand

= coil

Conf: = confidence of prediction

\- +

Pred: predicted secondary structure

AA: target sequence

SABLE-2

```
1                                                                    60
     MKYCKFCCKAVTGVKLIHVPKCAIKRKLWEQSLGCSLGENSQICDTHFNDSQWKAAPAKG
     CCEEHHHHHHCCCCEEEEECCCCHHHHHHHHHCCCCCCCCCEECCCCCCCCCCCCCCCC
     654445444468757888577655567788875887888865644455776667788888
>  61                                                                120
     QTFKRRRLNADAVPSKVIEPEPEKIKEGYTSGSTQTESCSLFNENKSLREKIRTLEYEMR
     CCCEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCHHHHHHHHHHHHHH
     754345647877777788888655677877888888855565787899999999999999
> 121                                                                180
     RLEQQLRESQQLEESLRKIFTDTQIRILKNGGQRATFNSDDISTAICLHTAGPRAYNHLY
     HHHHHHHHHHHHHHHHHHHCCHHHHHHHHHCCCEEEECHHHHHHHHHHHHCCCHHHHHHH
     877777558899988777556667889875787566577777776654446856889988
> 181                                                                240
     KKGFPLPSRTTLYRWLSDVDIKRGCLDVVIDLMDSDGVDDADKLCVLAFDEMKVAAAFEY
     HCCCCCCCHHHHHHHHHHCCCCCCCHHHHHHHHHCCCCCHHCCEEEEECCHHHHHHHCCC
     679988874579999986788766489999998678976545457765444436665457
> 241                                                                300
     DSSADIVYEPSDYVQLAIVRGLKKSWKQPVFFDFNTRMDPDTLNNILRKLHRKGYLVVAI
     CCCCCCCCCCCCCHHEHHHHHHHCCCCCCCEEEECCCCCCHHHHHHHHHHHHHCCCEEEEE
     887543467864444555665456777755666466678899999999999779859999
> 301                                                                360
     VSDLGTGNQKLWTELGISESKTWFSHPADDHLKIFVFSDTPHLIKLVRNHYVDSGLTING
     EECCCCCCHHHHHHCCCCCCCCCCCCCCCCCCCEEEECCCHHHHHHHHCCCCCCCEEECC
     967898866778875887788777888887775888678767899876565456567678
> 361                                                                420
     KKLTKKTIQEALHLCNKSDLSILFKINENHINVRSLAKQKVKLATQLFSNTTASSIRRCY
     CEECHHHHHHHHHHCCCCCCCCEEEECCCCCCCCCHHHHHCHHHHHHHHHHHHHHHHHHH
     656567899999865788776577756888778857764556788887666899999875
> 421                                                                480
     SLGYDIENATETADFFKLMNDWFDIFNSKLSTSNCIECSQPYGKQLDIQNDILNRMSEIM
     CCCCCCHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHH
     678875556789999999987766546778888888987678899999999999999887
> 481                                                                540
     RTGILDKPKRLPFQKGIIVNNASLDGLYKYLQENFSMQYILTSRLNQDIVEHFFGSMRSR
     HCCCCCCCCCEEECCHHHHHHHHHHHHHHHHHHHHCCCEEEEECCCCCHHHHHHHHHHHH
     567778988746533455667789999999887588768875445645799999999875
> 541                                                                576
     GGQFDHPTPLQFKYRLRKYIIGMTNLKECVNKNVIP
     CCCCCCCCHHHHHHHHHHHHHCCCCCCCCCCCCCCC
     788888888889999999987577666778888888
```

Appendix III: Part 1: Results of Protein Prediction programs for the Mutant P-Sal* Proteins

Note: Since MPS-1 encodes a 28 aa polypeptide that is too short for protein prediction programs, and MPS-5 and MPS-20 do not have any mutation in their coding sequence, I did not use them for protein prediction programs.

PROSITE motif search

http://www.expasy.org/tools/scanprosite/

The ScanProsite tool allows one to scan protein sequence(s) (either from UniProt Knowledgebase (Swiss-Prot/TrEMBL) or PDB or provided by the user) for the occurrence of patterns, profiles and rules (motifs) stored in the PROSITE database, or to search protein database(s) for hits by specific motif(s)

Mutant	Amino acid Change	Affected enzyme motif predicted by "PROSITE motif search"
MPS-1	29 Trp>**STOP**	**N/A**
MPS-2	30 Glu>Lys	**None**
MPS-3	44 Cys>Tyr	**None**
MPS-4	65 Arg>Lys	**None**
MPS-5	**SPLICING**	**N/A**
MPS-6	152 Gly>Glu	**MYRISTYL** 151 GGQRAT
MPS-7	168 Leu>Phe	**None**
MPS-8	195 Trp>**STOP**	**ASN_GLYCOSYLATION** 376 NKSD 410 NTTA 428 NATE 501 NASL 514 NFSM **CAMP_PHOSPHO_SITE** 361 KKLT **PKC_PHOSPHO_SITE** 265 SWK 364 TKK 415 SIR 522 TSR 536 SMR **CK2_PHOSPHO_SITE** 197 SDVD 242 SSAD 276 TRMD 367 TIQE **MYRISTYL** 217 GVDDAD 261 GLKKSW 316 GISESK 355 GLTING 496 GIIVNN 535 GSMRSR **AMIDATION** 359 NGKK
MPS-9	204 Gly>Glu	**None**

Mutant	Amino acid Change	Affected enzyme motif predicted by "PROSITE motif search"
MPS-10	296 Leu>STOP	**ASN_GLYCOSYLATION** 376 NKSD 410 NTTA 428 NATE 501 NASL 514 NFSM **CAMP_PHOSPHO_SITE** 361 KKLT **PKC_PHOSPHO_SITE** 364 TKK 415 SIR 522 TSR 536 SMR **CK2_PHOSPHO_SITE** 367 TIQE **MYRISTYL** 316 GISESK 355 GLTING 496 GIIVNN 535 GSMRSR **AMIDATION** 359 NGKK
MPS-11	297 Val>Glu	**None**
MPS-12	299 Ala>Val	**None**
MPS-13	322 Thr>Ile	**None**
MPS-14	363 Leu>Phe	**None**
MPS-15	404 Ala>Thr	**None**
MPS-16	461 Pro>Leu	**None**
MPS-17	496 Gly>Val	**MYRISTYL** 496 GIIVNN
MPS-18	522 Thr>Ile	**PKC_PHOSPHO_SITE** 522 TSR
MPS-19	567 Lys>Ile	**None**
MPS-20	**None**	**N/A**

NetPhos 2.0

http://www.cbs.dtu.dk/services/NetPhos/

The NetPhos 2.0 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins.
Kinase specific phosphorylation predictions are available at:

Original P protein
Ser at positions: 91, 107, 129, 135, 197, 215, 242, 265, 318, 414, 415, 451, 477, 539
Thr at positions: 62, 191, 522
Tyr at positions: 89, 240, 253, 351, 424

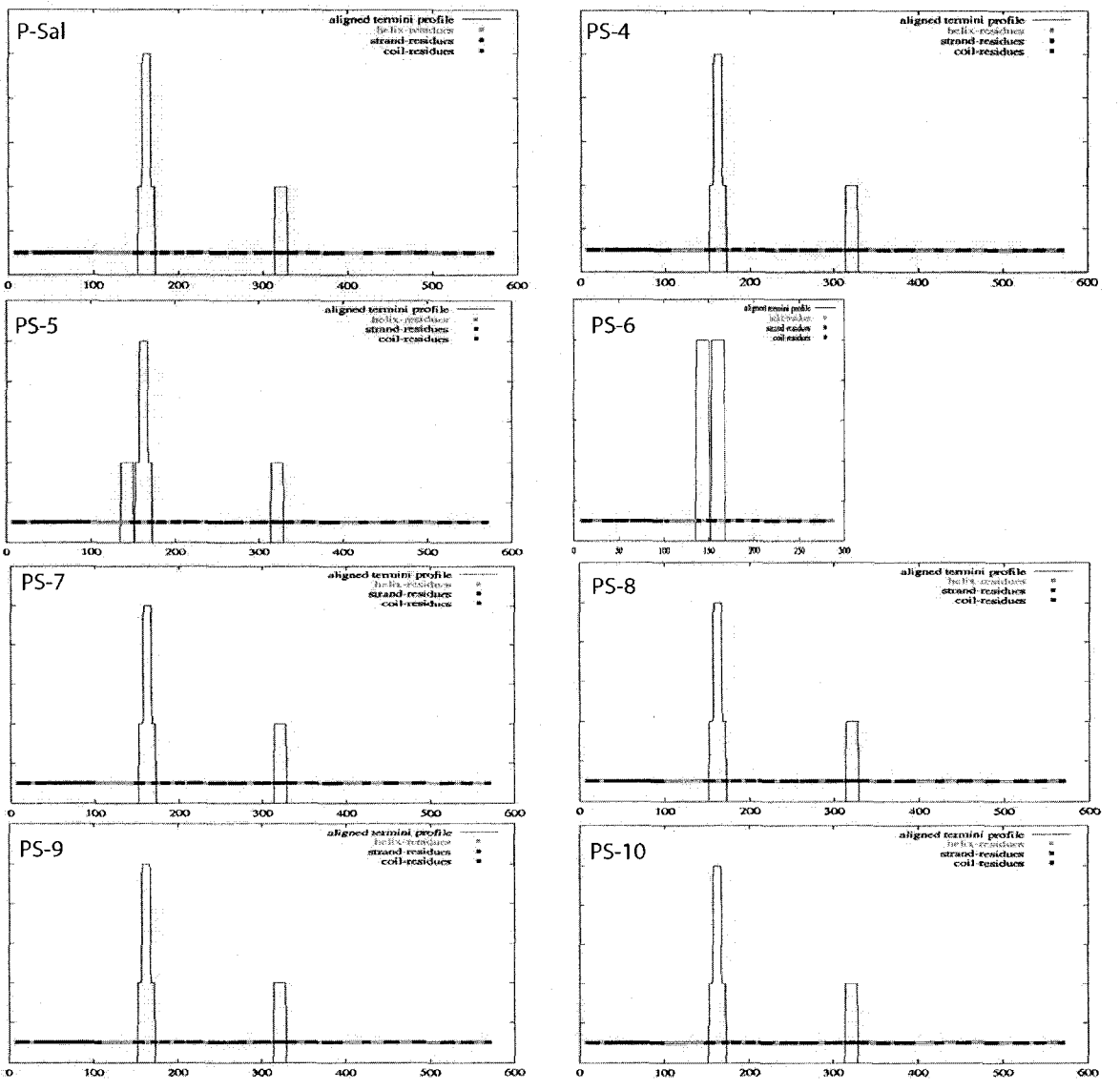
MPS-15: New phosphorylation site at 404 Ala>Thr
MPS-18: Loss of phosphorylation site at 522 Thr>Ile

DomPred Protein Domain Prediction Server

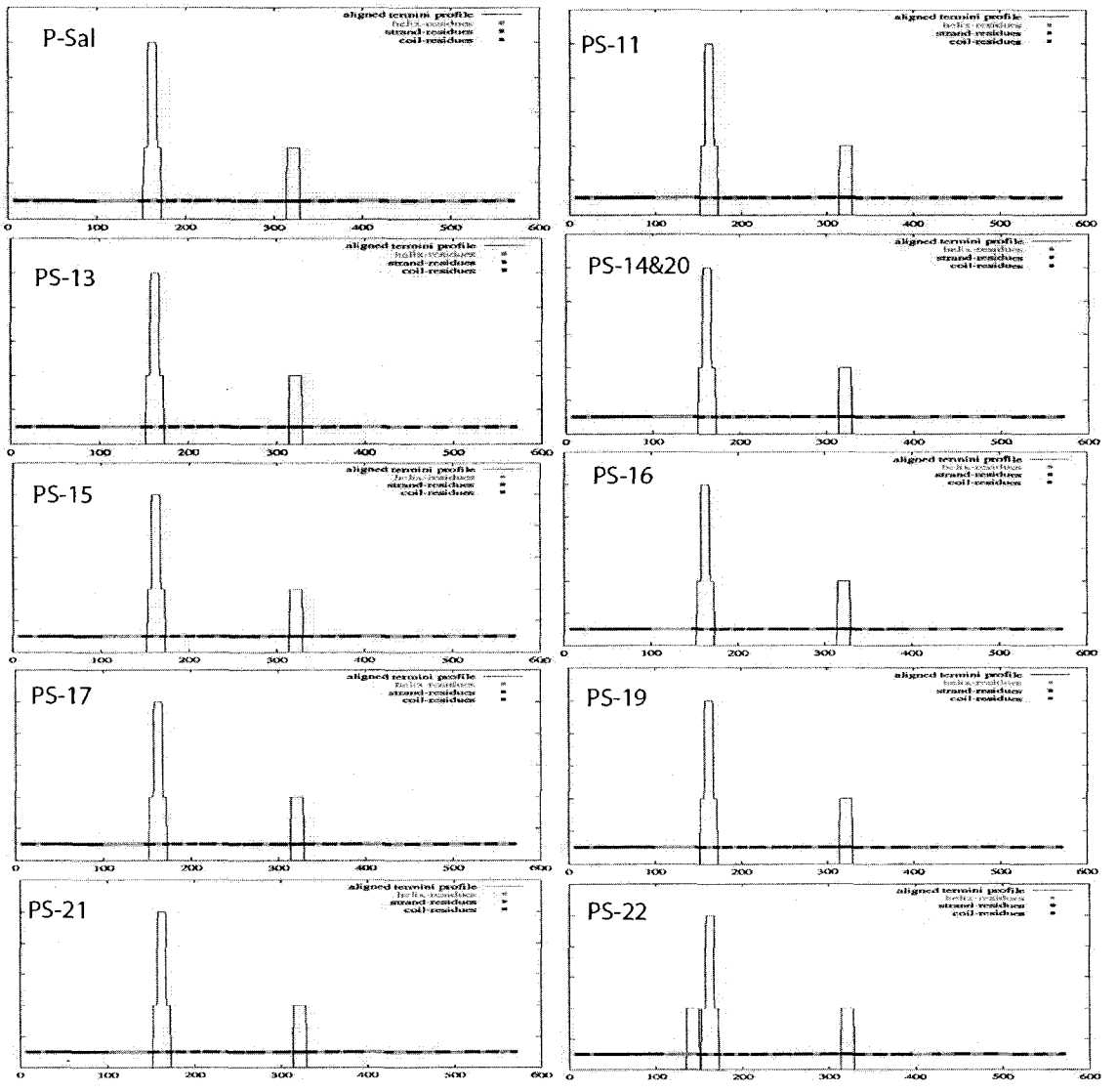
http://bioinf.cs.ucl.ac.uk/dompred/

DomPred is a server designed to predict putative protein domains and their boundaries for a given protein sequence. The server works on several levels, from identifying obvious similarities to Pfam-A domain sequences to predicting domains using DomSSEA in cases where sequence searching has yielded no results.

Mutant	Aminoacid Change	DomPred predicted changes from original P protein	Affected Domain
MPS-1	29 Trp>STOP	**N/A**	
MPS-2	30 Glu>Lys	**None**	
MPS-3	44 Cys>Tyr	**None**	
MPS-4	65 Arg>Lys	**None**	
MPS-5	SPLICING	**N/A**	
MPS-6	152 Gly>Glu	A new domain around aa 140	
MPS-7	168 Leu>Phe	**None**	
MPS-8	195 Trp>STOP	**N/A**	
MPS-9	204 Gly>Glu	A new domain around aa 140	
MPS-10	296 Leu>STOP	2 new domain side by side around aa 150	
MPS-11	297 Val>Glu	**None**	
MPS-12	299 Ala>Val	**None**	
MPS-13	322 Thr>Ile	**None**	
MPS-14	363 Leu>Phe	**None**	
MPS-15	404 Ala>Thr	**None**	
MPS-16	461 Pro>Leu	**None**	
MPS-17	496 Gly>Val	**None**	
MPS-18	522 Thr>Ile	**None**	
MPS-19	567 Lys>Ile	**None**	
MPS-20	None	**N/A**	



Predicted amino acid sequence of different mutants have been analyzed by DomPred online program (University College London) which searches for homology between the query sequence and P-fam A known domains database. The horizontal axis shows the place of amino acid in the sequence whereas the vertical axis represents homology to known domains. Different colors in the line represent probability of different secondary structures.


P-Sal
PS-11
PS-13
PS-14&20
PS-15
PS-16
PS-17
PS-19
PS-21
PS-22

PORTER

Protein Secondary Structure Prediction at University College Dublin

http://distill.ucd.ie/porter/

Mutant	Aminoacid Change	Porter predicted changes from original P protein	Affected Domain
MPS-1	29 Trp>**STOP**	**N/A**	
MPS-2	30 Glu>Lys	**None**	
MPS-3	44 Cys>Tyr	524C>H	Leucine zipper (497-525)
MPS-4	65 Arg>Lys	**None**	
MPS-5	**SPLICING**	**N/A**	
MPS-6	152 Gly>Glu	524C>H	Leucine zipper (497-525)
MPS-7	168 Leu>Phe	**None**	
MPS-8	195 Trp>**STOP**	7-8 C>E 85-87C>E 98H>C 159-160H>C 189-194C>H	Zinc finger(4-22), DNA binding domain (1-88) DNA binding domain (1-88)
MPS-9	204 Gly>Glu	524C>H	
MPS-10	296 Leu>STOP	7C>E 255-258E>H	Zinc finger(4-22), DNA binding domain (1-88)
MPS-11	297 Val>Glu	518C>E	Leucine zipper (497-525)
MPS-12	299 Ala>Val	518C>E	Leucine zipper (497-525)
MPS-13	322 Thr>Ile	**None**	
MPS-14	363 Leu>Phe	518C>E 524C>H	Leucine zipper (497-525)
MPS-15	404 Ala>Thr	**None**	
MPS-16	461 Pro>Leu	518C>E 523-524C>H	Leucine zipper (497-525)
MPS-17	496 Gly>Val	518C>E 524C>H	Leucine zipper (497-525)
MPS-18	522 Thr>Ile	**None**	
MPS-19	567 Lys>Ile	518C>E 524C>H	Leucine zipper (497-525)
MPS-20	**None**	**N/A**	

Note: High prevalence of predicted change at secondary structure of amino acids number 518C>E and 524C>H.

C: Coil
H: Helix
E: Strand

Predicted changes by Porter				
Change in 2ndry Structure		**Mutant Affected**	**Note**	
None		MPS- 2, 4, 7, 13, 15, 18		
7	C>E	**MPS-10**	**MPS-10:** 296 Leu>STOP	Zinc finger(4-22), DNA binding domain (1-88)
7-8	C>E	**MPS-8**	**MPS-8:** 195 Trp>**STOP**	DNA binding domain (1-88)
85-87	C>E	**MPS-8**		
98	H>C	**MPS-8**		
159-160	H>C	**MPS-8**		
189-194	C>H	**MPS-8**		
255-258	E>H	**MPS-10**		
518	C>E	MPS-11, 12, 14, 16, 17, 19	Leucine zipper (497-525)	
524	C>H	MPS-3, 6, 9, 14, 16, 17, 19		

MPS-8 and MPS-10 have nonsense mutation in the *P* element.

PSIPRED 2.0 protein structure prediction server

http://bioinf.cs.ucl.ac.uk/psipred/

PSIPRED is a simple and reliable secondary structure prediction method, incorporating two feed-forward neural networks which perform an analysis on output obtained from PSI-BLAST (Position Specific Iterated - BLAST).

Version 2.0 of PSIPRED includes a new algorithm which averages the output from up to 4 separate neural networks in the prediction process to further increase prediction accuracy.

Using a very stringent cross validation method to evaluate the method's performance, PSIPRED 2.0 is capable of achieving an average Q_3 score of nearly 78%. Predictions produced by PSIPRED were also submitted to the CASP4 server and assessed during the CASP4 meeting, which took place in December 2000 at Asilomar. PSIPRED 2.0 achieved an average Q_3 score of 80.6% across all 40 submitted target domains with no obvious sequence similarity to structures present in PDB, which ranked PSIPRED top out of 20 evaluated methods (an earlier version of PSIPRED was also ranked top in CASP3 held in 1998).

It is important to realise, however, that due to the small sample sizes, the results from CASP are not statistically significant, although they do give a rough guide as to the current "state of the art". For a more reliable evaluation, the EVA web site at Columbia University provides a continuous evaluation. Also see the EVA servlet to visualize a breakdown of specific types of errors made by PSIPRED and other secondary structure prediction methods.

Mutant	Amino acid Change	PSIPRED 2.0 predicted changes from original P protein	Affected Domain
MPS-1	29 Trp>**STOP**	**N/A**	
MPS-2	30 Glu>Lys	21C>H 63-64C>H 99H>C 427C>H	Zinc finger(4-22), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) Leucine zipper?(100-122)
MPS-3	44 Cys>Tyr	21C>H 62-64C>H 84-85C>E 99H>C 103-105H>C 251E>H	Zinc finger(4-22), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) DNA binding domain (1-88) Leucine zipper? (100-122) Leucine zipper (100-122)
MPS-4	65 Arg>Lys	63-64C>H 99H>C	NLS (64-68), DNA binding domain (1-88) Leucine zipper? (100-122)
MPS-5	**SPLICING**	**N/A**	
MPS-6	152 Gly>Glu	**None**	
MPS-7	168 Leu>Phe	44C>H 47C>H 99-107H>C 427C>H	DNA binding domain (1-88) DNA binding domain (1-88) Leucine zipper (100-122)
MPS-8	195 Trp>**STOP**	21C>H 33C>H 42-44C>E 45-46H>C 62-67C>H 159H>C 170H>C 189-194H>C	Zinc finger(4-22), DNA binding domain (1-88) DNA binding domain (1-88) DNA binding domain (1-88) DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88)
MPS-9	204 Gly>Glu	44C>H 47C>H 99-107H>C 427C>H	DNA binding domain (1-88) DNA binding domain (1-88) Leucine zipper (100-122)
MPS-10	296 Leu>STOP	47C>H 50-51C>H 61-64C>H 83-86C>E 103-105H>C 139-141H>C 227E>C 228H>C 246-248C>E 253-255E>H	DNA binding domain (1-88) DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) DNA binding domain (1-88) Leucine zipper (100-122)
MPS-11	297 Val>Glu	44C>H 47C>H 99-107H>C 427C>H	DNA binding domain (1-88) DNA binding domain (1-88) Leucine zipper (100-122)
MPS-12	299 Ala>Val	**None**	
MPS-13	322 Thr>Ile	427C>H	

Mutant	Amino acid Change	PSIPRED 2.0 predicted changes from original P protein	Affected Domain
MPS-14	363 Leu>Phe	356C>H 357-358E>H 375-377C>H 381C>H 426-427C>H	
MPS-15	404 Ala>Thr	426-427C>H	
MPS-16	461 Pro>Leu	427C>H 481H>C	
MPS-17	496 Gly>Val	427C>H 481H>C 493H>C 515C>H 528H>C	Leucine zipper (497-525)
MPS-18	522 Thr>Ile	528H>C	
MPS-19	567 Lys>Ile	427C>H 562-567H>C	
MPS-20	**None**	**N/A**	

Note: MPS-7, MPS-9 and MPS-11 have exactly the same pattern.

C: Coil
H: Helix
E: Strand

Predicted changes by PSIPRED 2.0			
Change in 2ndry Structure		**Affected Mutants**	**Note**
None		MPS-6, 12	
21	C>H	MPS-2, 3, **8**	Zinc finger(4-22), DNA binding domain (1-88)
33	C>H	**MPS-8**	DNA binding domain (1-88)
42-44	C>E	**MPS-8**	
44	C>H	MPS-7, 9, 11	
45-46	H>C	**MPS-8**	
47	C>H	MPS-7, 9, **10**, 11	
50-51	C>H	**MPS-10**	
61-64	C>H	**MPS-10**	NLS (64-68), DNA binding domain (1-88)
62-64	C>H	MPS-3	
62-67	C>H	**MPS-8**	
63-64	C>H	MPS-2, 4	
83-86	C>E	**MPS-10**	DNA binding domain (1-88)
84-85	C>E	MPS-3	
99	H>C	MPS-2, 3, 4	Leucine zipper (100-122)
99-107	H>C	MPS-7, 9, 11	
103-105	H>C	MPS-3, **10**	
139-141	H>C	**MPS-10**	**MPS-8**: 195 Trp>**STOP** **MPS-10**: 296 Leu>STOP
159	H>C	**MPS-8**	
170	H>C	**MPS-8**	
189-194	H>C	**MPS-8**	
227	E>C	**MPS-10**	
228	H>C	**MPS-10**	
246-248	C>E	**MPS-10**	
251	E>H	**MPS-8**	
253-255	E>H	**MPS-10**	
356	C>H	MPS-14	MPS-14: 363 Leu>Phe
357-358	E>H	MPS-14	
375-377	C>H	MPS-14	
381	C>H	MPS-14	
427	C>H	MPS-2, 7, 9, 11, 13, 16, 17, 19	
426-427	C>H	MPS-14, 15	
481	H>C	MPS-16, 17	MPS-17 496 Gly>Val Leucine zipper (497-525)
493	H>C	MPS-17	
515	C>H	MPS-17	
528	H>C	MPS-17, 18	
562-567	H>C	MPS-19	

MPS-8 and MPS-10 have nonsense mutation in the *P* element.

SABLE-2

SABLE-2 protein structure prediction server can be used to predict relative Solvent AccessiBiLitiEs, secondary structures and transmembrane domains for proteins of unknown structure.

http://sable.cchmc.org/

Mutant	Amino acid Change	SABLE-2 predicted changes from original P protein	Affected Domain
MPS-1	29 Trp>**STOP**	**N/A**	
MPS-2	30 Glu>Lys	7H>C 10H>C 45-47C>H 170H>C 205C>H 220-221H>C 231-233H>C 255E>H 494C>E	Zinc finger(4-22), DNA binding domain (1-88) Zinc finger(4-22), DNA binding domain (1-88) DNA binding domain (1-88)
MPS-3	44 Cys>Tyr	21-23C>H 45-47C>H 63-64E>C 65E>H 68E>C 82-86C>H 102H>C 170H>C 274C>E	Zinc finger(4-22), DNA binding domain (1-88) DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) DNA binding domain (1-88) Leucine zipper (100-122)
MPS-4	65 Arg>Lys	64E>C 65E>H 82-86C>H 170H>C 205C>H 220-221H>C 231-232H>C 455C>H	NLS (64-68), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) DNA binding domain (1-88)
MPS-5	**SPLICING**	**N/A**	
MPS-6	152 Gly>Glu	**None**	
MPS-7	168 Leu>Phe	4E>H 21-23C>H 45-46C>H 62-63C>H 64-66E>H 68E>C 82-87C>H 99-102H>C 170H>C 189H>C 252C>H 255E>H 274C>E	Zinc finger(4-22), DNA binding domain (1-88) Zinc finger(4-22), DNA binding domain (1-88) DNA binding domain (1-88) NLS? (64-68), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) DNA binding domain (1-88) Leucine zipper (100-122)

Mutant	Amino acid Change	SABLE-2 predicted changes from original P protein	Affected Domain
MPS-8	195 Trp>**STOP**	22-23C>H 41E>C 45-46C>H 64E>C 65E>H 68E>C 99H>C 173C>H 194-195H>C	Zinc finger(4-22), DNA binding domain (1-88) DNA binding domain (1-88) DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) Leucine zipper? (100-122)
MPS-9	204 Gly>Glu	170H>C 214H>C 220-221H>C 231-232H>C 495C>H	
MPS-10	296 Leu>STOP	4E>H 64E>C 68E>C 83-85C>H 99-100H>C 127-128H>C 139H>C 170H>C 174H>C 181H>C 214H>C 220-221H>C 231-232H>C 233-237H>E 238-239C>E 253C>H 256E>H 263H>C	Zinc finger(4-22), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) NLS (64-68), DNA binding domain (1-88) DNA binding domain (1-88) Leucine zipper (100-122)
MPS-11	297 Val>Glu	**None**	
MPS-12	299 Ala>Val	99-102H>C 170H>C 205C>H 220-221H>C 231-232H>C 495C>H	Leucine zipper (100-122)
MPS-13	322 Thr>Ile	99-102 H>C 170H>C 189H>C 205C>H 252C>H 255E>H 495C>H	Leucine zipper (100-122)

Mutant	Amino acid Change	SABLE-2 predicted changes from original P protein	Affected Domain
MPS-14	363 Leu>Phe	170H>C 205C>H 220-221H>C 231-232H>C 349-353C>H 356E>C 375C>H 386E>C 395H>C 495C>H	
MPS-15	404 Ala>Thr	**None**	
MPS-16	461 Pro>Leu	170H>C 274C>E 460C>H 491E>C 495C>H 514H>C 523-524C>E 540H>C 562C>H	 Leucine zipper (497-525) Leucine zipper (497-525)
MPS-17	496 Gly>Val	99-102H>C 189H>C 252C>H 274C>E 461H>C 481H>C 491E>C 496-498H>E 513-514H>C 523-524C>E 540H>C	Leucine zipper (100-122) Leucine zipper (497-525) Leucine zipper (497-525) Leucine zipper (497-525)
MPS-18	522 Thr>Ile	99-102H>C 274C>E 523C>E 528H>C 540H>C 562C>H	Leucine zipper (100-122) Leucine zipper (497-525)
MPS-19	567 Lys>Ile	99H>C 102H>C 173C>H 220-221H>C 231-232H>C 255E>H	Leucine zipper (100-122) Leucine zipper (100-122)
MPS-20	**None**	**N/A**	

C: Coil
H: Helix
E: Strand

Predicted changes by SABLE 2			
Change in 2ndry Structure		**Affected Mutants**	**Note**
None		MPS-6, 11, 15	
4	E>H	MPS-7, **10**	Zinc finger(4-22), DNA binding domain (1-88)
7	H>C	MPS-2	
10	H>C	MPS-2	
21-23	C>H	MPS-3, 7	
22-23	C>H	**MPS-8**	
41	E>C	**MPS-8**	DNA binding domain (1-88)
45-46	C>H	MPS-7, **8**	
45-47	C>H	MPS-2	
62-63	C>H	MPS-7	NLS(64-68), DNA binding domain (1-88)
63-64	E>C	MPS-3	
64	E>C	MPS-4, **8, 10**	
64-66	E>H	MPS-7	
65	E>H	MPS-3, 4, **8**	
68	E>C	MPS-3, 7, **8, 10**	
82-86	C>H	MPS-3, 4	DNA binding domain (1-88)
82-87	C>H	MPS-7	
83-85	C>H	**MPS-10**	
99	H>C	MPS-**8**, 19	Leucine zipper (100-122)
99-100	H>C	**MPS-10**	
99-102	H>C	MPS-7,12,13,17,18	
102	H>C	MPS-3, 19	
127-128	H>C	**MPS-10**	**MPS-10**: 296 Leu>STOP
139	H>C	**MPS-10**	
170	H>C	MPS-2, 3, 4, 7, 9, **10**, 12, 13, 14, 16	
173	C>H	MPS-8, 19	
174	H>C	**MPS-10**	
181	H>C	**MPS-10**	
189	H>C	MPS-7, 13, 17	
194-195	H>C	**MPS-8**	
205	C>H	MPS-2, 4, 12, 13, 14	
214	H>C	MPS-9, **10**	
220-221	H>C	MPS-2, 4, 9, **10**, 12, 14, 19	
231-232	H>C	PS-4, 9, **10**, 12, 14, 19	
231-233	H>C	MPS-2	
233-237	H>E	**MPS-10**	
238-239	C>E	**MPS-10**	
252	C>H	MPS-7, 13, 17	
253	C>H	**MPS-10**	
255	E>H	MPS-2, 7, 13, 19	
256	E>H	**MPS-10**	
263	H>C	**MPS-10**	
274	C>E	PS-4, 9, **10**, 13, 17	
349-353	C>H	MPS-14	MPS-14: 363 Lue>Phe
356	E>C	MPS-14	
375	C>H	MPS-14	
386	E>C	MPS-14	
395	H>C	MPS-14	

Predicted changes by SABLE 2			
Change in 2ndry Structure		**Affected Mutants**	**Note**
455	C>H	MPS-4	
460	C>H	MPS-16	
461	H>C	MPS-17	
481	H>C	MPS-17	
491	E>C	MPS-16, 17	
494	C>E	MPS-2	
495	C>H	MPS-9, 12, 14, 16	
496-498	H>E	MPS-17	
513-514	H>C	MPS-17	MPS-16: 461 Pro>Leu MPS-17: 496 Gly>Val MPS-18: 522 Thr>Ile Leucine zipper (497-525)
514	H>C	MPS-16	
523	C>E	MPS-18	
523-524	C>E	MPS-16, 17	
540	H>C	MPS-16, 17, 18	
562	C>H	MPS-16, 18	

MPS-8 and MPS-10 have nonsense mutation in the P element.

SIFT: predicting amino acid changes that affect protein function

http://blocks.fhcrc.org/sift/SIFT.html

SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT can be applied to naturally occurring nonsynonymous polymorphisms and laboratory-induced missense mutations.

Given a protein sequence, SIFT will return predictions for what amino acid substitutions will affect protein function.

SIFT is a multi-step procedure that:

(1) searches for and chooses similar sequences

(2) makes an alignment of these sequences

(3) calculates scores based on the amino acids appearing at each position in the alignment.

Mutant	Amino acid Change	SIFT predicted effect of mutation on protein structure
MPS-1	29 Trp>**STOP**	**N/A**
MPS-2	30 Glu>Lys	Affect structure
MPS-3	44 Cys>Tyr	Affect structure
MPS-4	65 Arg>Lys	**Mutation will be tolerated**
MPS-5	**SPLICING**	**N/A**
MPS-6	152 Gly>Glu	**Mutation will be tolerated**
MPS-7	168 Leu>Phe	Affect structure
MPS-8	195 Trp>**STOP**	Affect structure
MPS-9	204 Gly>Glu	Affect structure
MPS-10	296 Leu>STOP	Affect structure
MPS-11	297 Val>Glu	Affect structure
MPS-12	299 Ala>Val	Affect structure
MPS-13	322 Thr>Ile	**Mutation will be tolerated**
MPS-14	363 Leu>Phe	Affect structure
MPS-15	404 Ala>Thr	Affect structure
MPS-16	461 Pro>Leu	Affect structure
MPS-17	496 Gly>Val	Affect structure
MPS-18	522 Thr>Ile	Affect structure
MPS-19	567 Lys>Ile	**Mutation will be tolerated**
MPS-20	**None**	**N/A**

Appendix III: Part 2: Results of Protein Prediction programs for the Mutant P-Sal* Proteins (sorted based on location of mutation)

Note: Since MPS-1 encodes a 28 aa polypeptide that is too short for protein prediction programs, and MPS-5 and MPS-20 do not have any mutation in their coding sequence, I did not use them for protein prediction programs.

Mutant	Amino acid Change	predicted changes from original P protein						
		Predicted change in 2ndary structure				Affected enzyme motif predicted by "PROSITE motif search"	SIFT predicted effect of mutation on protein structure	Change in NetPhos Predicted phosphorylation sites of original P protein
		Porter	PSIPRED 2.0	SABLE-2	DomPred			
MPS-1	29 Trp>**STOP**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
MPS-2	30 Glu>Lys	**None**	21C>H 63-64C>H 99H>C 427C>H	7H>C 10H>C 45-47C>H 170H>C 205C>H 220-221H>C 231-233H>C 255E>H 494C>E	**None**	**None**	Affect structure	**None**
MPS-3	44 Cys>Tyr	524C>H	21C>H 62-64C>H 84-85C>E 99H>C 103-105H>C 251E>H	21-23C>H 45-47C>H 63-64E>C 65E>H 68E>C 82-86C>H 102H>C 170H>C 274C>E	**None**	**None**	Affect structure	**None**
MPS-4	65 Arg>Lys	**None**	63-64C>H 99H>C	64E>C 65E>H 82-86C>H 170H>C 205C>H 220-221H>C 231-232H>C 455C>H	**None**	**None**	**Mutation will be tolerated**	**None**

Mutant	Amino acid Change	predicted changes from original P protein						
		Predicted change in 2ndary structure				Affected enzyme motif predicted by "PROSITE motif search"	SIFT predicted effect of mutation on protein structure	Change in NetPhos Predicted phosphorylation sites of original P protein
		Porter	PSIPRED 2.0	SABLE-2	DomPred			
MPS-5	**SPLICING**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
MPS-6	152 Gly>Glu	524C>H	**None**	**None**	A new domain around aa 140	**MYRISTYL** 151 GGQRAT	**Mutation will be tolerated**	**None**
MPS-7	168 Leu>Phe	**None**	44C>H 47C>H 99-107H>C 427C>H	4E>H 21-23C>H 45-46C>H 62-63C>H 64-66E>H 68E>C 82-87C>H 99-102H>C 170H>C 189H>C 252C>H 255E>H 274C>E	**None**	**None**	Affect structure	**None**

Mutant	Amino acid Change	predicted changes from original P protein						
		Predicted change in 2ndary structure				Affected enzyme motif predicted by "PROSITE motif search"	SIFT predicted effect of mutation on protein structure	Change in NetPhos Predicted phosphorylation sites of original P protein
		Porter	PSIPRED 2.0	SABLE-2	DomPred			
MPS-8	195 Trp>**STOP**	7-8 C>E 85-87C>E 98H>C 159-160H>C 189-194C>H	21C>H 33C>H 42-44C>E 45-46H>C 62-67C>H 159H>C 170H>C 189-194H>C	22-23C>H 41E>C 45-46C>H 64E>C 65E>H 68E>C 99H>C 173C>H 194-195H>C	**N/A**	**ASN_GLYCOSYLATION** 376 NKSD 410 NTTA 428 NATE 501 NASL 514 NFSM **CAMP_PHOSPHO_SITE** 361 KKLT **PKC_PHOSPHO_SITE** 265 SWK 364 TKK 415 SIR 522 TSR 536 SMR **CK2_PHOSPHO_SITE** 197 SDVD 242 SSAD 276 TRMD 367 TIQE **MYRISTYL** 217 GVDDAD 261 GLKKSW 316 GISESK 355 GLTING 496 GIIVNN 535 GSMRSR **AMIDATION** 359 NGKK	Affect structure	**None**

Mutant	Amino acid Change	predicted changes from original P protein						
		Predicted change in 2ndary structure				Affected enzyme motif predicted by "PROSITE motif search"	SIFT predicted effect of mutation on protein structure	Change in NetPhos Predicted phosphorylation sites of original P protein
		Porter	PSIPRED 2.0	SABLE-2	DomPred			
MPS-9	204 Gly>Glu	524C>H	44C>H 47C>H 99-107H>C 427C>H	170H>C 214H>C 220-221H>C 231-232H>C 495C>H	A new domain around aa 140	**None**	Affect structure	**None**
MPS-10	296 Leu>**STOP**	7C>E 255-258E>H	47C>H 50-51C>H 61-64C>H 83-86C>E 103-105H>C 139-141H>C 227E>C 228H>C 246-248C>E 253-255E>H	4E>H 64E>C 68E>C 83-85C>H 99-100H>C 127-128H>C 139H>C 170H>C 174H>C 181H>C 214H>C 220-221H>C 231-232H>C 233-237H>E 238-239C>E 253C>H 256E>H 263H>C	2 new domain side by side around aa 150	**ASN_GLYCOSYLATION** 376 NKSD 410 NTTA 428 NATE 501 NASL 514 NFSM **CAMP_PHOSPHO_SITE** 361 KKLT **PKC_PHOSPHO_SITE** 364 TKK 415 SIR 522 TSR 536 SMR **CK2_PHOSPHO_SITE** 367 TIQE **MYRISTYL** 316 GISESK 355 GLTING 496 GIIVNN 535 GSMRSR **AMIDATION** 359 NGKK	Affect structure	**None**

Mutant	Amino acid Change	predicted changes from original P protein						
		Predicted change in 2ndary structure				Affected enzyme motif predicted by "PROSITE motif search"	SIFT predicted effect of mutation on protein structure	Change in NetPhos Predicted phosphorylation sites of original P protein
		Porter	PSIPRED 2.0	SABLE-2	DomPred			
MPS-11	297 Val>Glu	518C>E	44C>H 47C>H 99-107H>C 427C>H	**None**	**None**	**None**	Affect structure	**None**
MPS-12	299 Ala>Val	518C>E	**None**	99-102H>C 170H>C 205C>H 220-221H>C 231-232H>C 495C>H	**None**	**None**	Affect structure	**None**
MPS-13	322 Thr>Ile	**None**	427C>H	99-102 H>C 170H>C 189H>C 205C>H 252C>H 255E>H 495C>H	**None**	**None**	**Mutation will be tolerated**	**None**
MPS-14	363 Leu>Phe	518C>E 524C>H	356C>H 357-358E>H 375-377C>H 381C>H 426-427C>H	170H>C 205C>H 220-221H>C 231-232H>C 349-353C>H 356E>C 375C>H 386E>C 395H>C 495C>H	**None**	**None**	Affect structure	**None**

Mutant	Amino acid Change	predicted changes from original P protein						
		Predicted change in 2ndary structure				Affected enzyme motif predicted by "PROSITE motif search"	SIFT predicted effect of mutation on protein structure	Change in NetPhos Predicted phosphorylation sites of original P protein
		Porter	PSIPRED 2.0	SABLE-2	DomPred			
MPS-15	404 Ala>Thr	**None**	426-427C>H	**None**	**None**	**None**	Affect structure	Gain of 404 Ala>Thr
MPS-16	461 Pro>Leu	518C>E 523-524C>H	427C>H 481H>C	170H>C 274C>E 460C>H 491E>C 495C>H 514H>C 523-524C>E 540H>C 562C>H	**None**	**None**	Affect structure	**None**
MPS-17	496 Gly>Val	518C>E 524C>H	427C>H 481H>C 493H>C 515C>H 528H>C	99-102H>C 189H>C 252C>H 274C>E 461H>C 481H>C 491E>C 496-498H>E 513-514H>C 523-524C>E 540H>C	**None**	**MYRISTYL** 496 GIIVNN	Affect structure	**None**

Mutant	Amino acid Change	predicted changes from original P protein						
		Predicted change in 2ndary structure				Affected enzyme motif predicted by "PROSITE motif search"	SIFT predicted effect of mutation on protein structure	Change in NetPhos Predicted phosphorylation sites of original P protein
		Porter	PSIPRED 2.0	SABLE-2	DomPred			
MPS-18	522 Thr>Ile	**None**	528H>C	99-102H>C 274C>E 523C>E 528H>C 540H>C 562C>H	**None**	**PKC_PHOSPHO_SITE** 522 TSR	Affect structure	Loss of 522 Thr>Ile
MPS-19	567 Lys>Ile	518C>E 524C>H	427C>H 562-567H>C	99H>C 102H>C 173C>H 220-221H>C 231-232H>C 255E>H	**None**	**None**	**Mutation will be tolerated**	**None**
MPS-20	**None**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**