INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
CHARACTERIZATION OF TRITHORAX, A PROTEIN REQUIRED FOR LONG-TERM MAINTENANCE OF HOMEOTIC GENE EXPRESSION

by

VANDANA CHINWALLA

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Thesis Adviser: Dr. Peter J. Harte

Department of Genetics
CASE WESTERN RESERVE UNIVERSITY
August, 1995
CASE WESTERN RESERVE UNIVERSITY

GRADUATE STUDIES

We hereby approve the thesis of

_________________________
Vandana Chinwalla

candidate for the ______Ph.D.__________
degree.*

(signed) _______________________

chair

_________________________
Russell Mann

_________________________
J. Hart

date __July, 1995__

*We also certify that written approval has been obtained for any proprietary material contained therein.
I grant to Case Western Reserve University the right to use this work, irrespective of any copyright, for the University's own purposes without cost to the University or to its students, agents and employees. I further agree that the University may reproduce and provide single copies of the work, in any format other than in or from microforms, to the public for the cost of reproduction.

VandanaChirsali
CHARACTERIZATION OF *TRITHORAX*, A PROTEIN REQUIRED FOR LONG-TERM MAINTENANCE OF HOMEOTIC GENE EXPRESSION

Abstract
by
VANDANA CHINWALLA

Homeotic genes of *Drosophila* determine the unique development of each segment and must be precisely regulated to ensure normal development. Long-term maintenance of homeotic gene expression requires the presence of positive (trithorax-Group) and negative (Polycomb-Group) regulators.

*trithorax* (*trx*) is the prototypic member of the trithorax-Group and is required for the positive maintenance of homeotic gene expression patterns. We developed two different polyclonal antibodies against parts of *trx* proteins. Using these antibodies we showed that the *trx* proteins bind to 63 specific sites on the polytene chromosomes of larval salivary glands. These sites include sites of their known targets, the Bithorax and Antennapedia complexes. Binding of *trx* to the cytological location of the Bithorax complex, which is not expressed in salivary glands, suggests that the *trx* binding is constitutive. We localized one *trx* binding site within a 670 bp fragment of the 5' regulatory region of *Ultrabithorax* (*Ubx*). These results
suggest that the $trx$ proteins exert their effect by binding directly or indirectly to specific DNA sequences in their target genes.

We showed that the $trx$ proteins co-localize with Polycomb-Group proteins at many sites on the polytene chromosomes. The same DNA fragment from the $Ub$x regulatory region also contains binding sites for Polycomb-Group proteins. Our results also indicate that the $trx$ proteins are co-localized at almost all sites with another member of the trithorax-Group proteins, the $absent$, $small$, $homeotic$ $l$ protein. These results suggest that the interaction between these positive and negative regulators of the homeotic genes may be important for their mode of action.

The $trx$ proteins contain a DNA-binding motif which is a novel variant of the DNA-binding domains of the nuclear receptor superfamily. We expressed this domain as a fusion protein and used an electrophoretic mobility shift assay to investigate and characterize its DNA-binding activity.
DEDICATED TO MY DEAR PARENTS
AND MY BELOVED HUSBAND
ACKNOWLEDGMENTS

I wish to express my feelings of gratitude to everyone who helped me in my work. I would first like to thank my thesis advisor, Dr. Peter Harte for his guidance and support. I want to thank the members of my committee Drs. Marcello Jacobs-Lorena, Helen Salz, Russ Maurer and David Setzer for their advice and interest in my thesis project.

I want to thank the members of the Harte lab David Bailey, Esther P. Jane, Janna Stassen and Swati Sathe. David helped me survive graduate school in more ways than one. I am especially thankful to Esther for doing all those chromosome preps and for her delicious samosas. Swati's encouragement and friendship have always been greatly valued.

I wish to thank Karen Gustashaw for her time and effort in teaching me confocal microscopy. I would also like to thank Jean Davis for all her help.

I also wish to thank my friends Shanta, Paresh, Raj and Shyam for making it easier for me to initially adjust to this country.

I am grateful to each and every member of my family especially my parents for their love, support and encouragement. I wish to thank my husband Juzer for his love and patience.
TABLE OF CONTENT

CHAPTER I: Background and significance------------------------ 1
The homeotic genes of Drosophila melanogaster--------------- 2
The trithorax-G------------------------------------------ 6
Genetic interactions among the trx-G genes----------------- 9
The trx protein isoforms contain sequences homologous to proteins from other organisms------------------- 14
A member of the trx-G, brahma, encodes a protein which shares homology with ATP-dependent helicases--------- 15
Trithorax-like encodes the GAGA factor---------------------- 16
The ashl protein binds polytene chromosomes at specific sites----------------------------------------------- 18
The trx-G proteins and modifiers of PEV may share some features of gene regulation---------------------- 19
The Polycomb-Group-------------------------------------- 22
Genetic interaction of trx with the Pc-G genes---------- 25
Summary and thesis outline----------------------------- 26
Aims of the thesis------------------------------------- 29

CHAPTER II: Immunodetection of trx proteins

INTRODUCTION------------------------------------------ 30
Immunostaining of polytene chromosomes------------------- 31
MATERIALS AND METHODS------------------------------- 33
Preparation of immunogens for induction of antibodies in rabbits------------------------------- 33
Western Analysis------------------------------------- 38
Immunostaining of polytene chromosomes------------------ 42
Co-immunolocalization of the *trx* and *Pc* proteins
on polytene chromosomes----------------------------- 45
Co-immunolocalization of the *trx* and *ashl* proteins
on polytene chromosomes----------------------------- 47
RESULTS-------------------------------------------------- 48
Immunodetection of *trx* proteins by Western analysis----- 48
Chromosomal binding of the *trx* proteins--------------- 52
The *trx* and *Pc* proteins co-localize at many sites
on polytene chromosomes------------------------------- 61
The *trx* and *ashl* proteins co-localize at
most of their chromosomal binding sites-------------- 64
DISCUSSION-------------------------------------------- 67
The *trx* proteins bind chromosomes-------------------- 67
The binding of *trx* and the transcriptional status
of target genes---------------------------------------- 70
The *trx* and *Pc* proteins are co-localized at several
chromosomal sites------------------------------------- 73
The *trx* and *ashl* proteins are co-localized at
many sites on the polytene chromosomes--------------- 74
The role of *trx* in the homeotic gene activation------- 75
Chapter III: *In vivo* mapping of *trx* binding sites
to smaller regions of *Ubx* UCR
INTRODUCTION------------------------------------------ 81
The P element mediated transformation------------------ 83
The enhancer elements of $Ubx$----------------------------- 85
Analysis of the $Ubx$ regulatory region--------------------- 88
The PRE is contained in a 670 bp fragment------------------- 97
MATERIALS AND METHODS---------------------------------- 99
The $in situ$ hybridization of polytene chromosomes-------- 99
RESULTS---------------------------------------------------- 103
The $trx$ proteins bind to the upstream control
region of $Ubx$------------------------------------------ 103
Delimiting the $trx$ binding sites in the UCR------------- 109
DISCUSSION--------------------------------------------- 113
The 670 bp fragment of the $Ubx$ UCR contains binding
site(s) for both $trx$ and Pc-G proteins--------------------- 113
The $trx$ and Pc-G proteins are co-localized at
the $Ubx$ promoter---------------------------------------- 114
The $trx$ proteins bind constitutively to its target genes-- 115
The mode of action of the $trx$ and Pc-G proteins------------ 115
Determination of the expression state of a promoter------- 117
CHAPTER IV: The $trx$ DNA binding domain binds DNA
with high affinity but low specificity
INTRODUCTION--------------------------------------------- 121
Structure of the nuclear receptor DBDs-------------------- 122
Residues important for the structural stability of
nuclear receptor DBDs-------------------------------------- 129
Does the putative $trx$ DBD bind DNA?-------------------- 131
MATERIALS AND METHODS---------------------------------- 133
Isolation of a cDNA encoding the $trx$ DBD------------------ 133
Preparation of the GST-DBD fusion protein------------------ 134
Electrophoretic Mobility Shift Assay------------------------ 136
RESULTS----------------------------------------------------- 138
The $trx$ DNA binding domain ($trx$-DBD) is encoded
by two exons-------------------------------------------- 138
The expression and purification of the GST-DBD
fusion protein------------------------------------------ 138
The GST-DBD I fusion protein binds DNA with high
affinity but low specificity---------------------------- 140
DISCUSSION----------------------------------------------- 149
The $trx$-DBD binds DNA---------------------------------- 149
The GST-DBD fusion protein exhibits sequence-specific
DNA-binding along with some non-specific binding------ 150
The $trx$ proteins exhibit sequence specificity in vivo----- 151
Chapter V: Future Directions------------------------------- 153
LITERATURE CITED---------------------------------------- 156
LIST OF FIGURES

Figure 1A: Structure of alternatively spliced \textit{trx} mRNA----- 11
Figure 1B: The predicted \textit{trx} protein isoforms------------------ 12
Figure 2: Characterization of the "common" anti-\textit{trx}
antibodies by Western analysis----------------------------- 50
Figure 3: Characterization of the isoform-specific anti-\textit{trx}
antibodies by Western analysis----------------------------- 51
Figure 4: The \textit{trx} proteins are associated with polytene
chromosomes-------------------------------------------------- 54
Figure 5: The \textit{trx} \textsuperscript{l} nuclei have greatly reduced number of
binding sites-------------------------------------------------- 55
Figure 6: Distributions of the number of chromosome
binding sites per nucleus in large matched samples of \textit{trx} \textsuperscript{l}
mutant and wild-type nuclei---------------------------------- 57
Figure 7: Co-localization of \textit{trx} and \textit{Pc} chromosome binding
sites by confocal microscopy------------------------------- 62
Figure 8: Co-localization of \textit{trx} and \textit{ash} \textsuperscript{l} chromosome
binding sites by confocal microscopy------------------------ 65
Figure 9: Schematic representation of the body segments,
compartments and parasegments----------------------------- 86
Figure 10: Map of \textit{Ubx} constructs used for transformation-- 91
Figure 11: Immunostaining of polytene chromosomes
from the 35UZ transformant---------------------------------- 105
Figure 12: Immunostaining of polytene chromosomes
from the 22UZ transformant---------------------------------- 106
Figure 13: Immunostaining of polytene chromosomes from the 14 kb UCR fragment containing transformant 108
Figure 14: Immunostaining of polytene chromosomes from the 2.2 kb UCR fragment containing transformant 110
Figure 15: Immunostaining of polytene chromosomes from the 14 kb UCR fragment containing transformant 111
Figure 16: Alignment of the residues of the putative trx DBD with the other members of the nuclear receptor family 123
Figure 17: Structure of DBD 126
Figure 18: The GST DBD fusion protein binds DNA with high affinity 141
Figure 19: The trx DBD binds DNA 142
Figure 20: Determination of specificity of the GST-DBD I protein by using linear fragments 143
Figure 21: Determination of specificity of the GST-DBD I protein by using Poly dl-dC and salmon sperm DNA 145
Figure 22: The GST-DBD II fusion protein binds DNA with high affinity 146
Figure 23: Determination of specificity of the GST-DBD II protein by using unlabeled supercoiled DNAs 147
LIST OF TABLES

Table I: Chromosomal binding sites of the trx proteins--- 59
CHAPTER I: Background and Significance

Development of a single cell embryo into a multicellular organism has intrigued biologists for over a century. It involves the activation of different sets of genes which are active at particular times and in a subset of cells of the developing embryo. Cellular differentiation requires both positive and negative regulators to activate the correct set of genes in a given cell and to prevent expression of inappropriate ones in that cell. Accumulating experimental evidence in a number of species suggests that the developmental decision of a cell made during early development is maintained by active continuous regulation, both by positive and negative regulators (Blau, 1992).

The determination of body segment identity in *Drosophila melanogaster* is an excellent system for studying the genetic and molecular mechanisms underlying differentiation and the control of cell fate during development. *Drosophila* has a segmented body and the identity of each segment is determined very early during insect development. The progenitor cells of each segment follow an unique developmental pathway which governs the cell types and anatomical structures which distinguish each segment in adults. For example, a pair of wings is formed on the second thoracic segment whereas a pair of antennae is present on the head segment.
The homeotic genes of *Drosophila melanogaster*

Homeotic genes are a set of genes which are required for the unique development of each segment (García-Bellido, 1975; Lewis, 1978). They are also referred to as selector genes because they control segment identity by selecting which target genes are active in a particular segment (García-Bellido, 1975). Mutations in any of the homeotic genes result in homeotic transformations. Homeotic transformation is defined as a transformation of a structure into another normally occurring body structure found in a different part of the body.

Most of the homeotic genes of *Drosophila* are physically clustered in two complexes: the Antennapedia Complex (ANT-C) (Kaufman *et al.*, 1990; Wakimoto and Kaufman, 1981) and the Bithorax Complex (BX-C) (Lewis, 1978). The ANT-C consists of *labial* (*lab*), *proboscipedia* (*pb*), *Sex combs reduced* (*Scr*), *Deformed* (*Dfd*) and *Antennapedia* (*Antp*) which specify head, first and second thoracic segmental identities (Kaufman *et al.*, 1990; Wakimoto and Kaufman, 1981). The BX-C includes the *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*) genes. These genes specify the third thoracic and abdominal segments (Duncan, 1987; Lewis, 1978; Sanchez-Herrero *et al.*, 1985). The ANT-C and BX-C genes encode proteins containing a conserved 60 amino acid domain called the "homeobox" (Gehring and Hiromi, 1986; Scott *et al.*, 1989). These
gene products bind DNA in vitro and act as transcription factors (Levine and Hoey, 1988).

Each homeotic gene is expressed in a spatially restricted pattern in different regions of the embryo and governs the fate of the cells in which it is expressed (Akam, 1987). Long-term stability of the spatial patterns and levels of homeotic gene expression is essential to their function in implementing stable determination of cell fates and differentiation of the normal body plan (Diederich et al., 1989; Merrill et al., 1987; Morata and García-Bellido, 1976; Struhl, 1982).

Both lack of expression of the homeotic genes and their ectopic expression outside of their domains causes homeotic transformations (González-Reyes and Morata, 1990; Kuziora and McGinnis, 1988). For example, a loss of function mutation in Ubx transforms halteres of the third thoracic segment into wings of the second thoracic segment (Lewis, 1978; Morata and Kerridge, 1981). Ectopic expression of Ubx outside of its normal expression domain transforms the structures of head and thoracic segments into homologous structures of first abdominal segment (González-Reyes and Morata, 1990). This indicates that precise regulation of the spatially restricted pattern and levels of homeotic gene expression is very important for normal development.
Determination and maintenance of the homeotic gene expression patterns

Expression of the homeotic genes is initiated at the blastoderm stage and is maintained throughout development (Ceñiker et al., 1989; Kornfeld et al., 1989; Martinez-Arias et al., 1987; Scott et al., 1983). The expression domains of different homeotic genes are initially determined by the transiently expressed segmentation genes (Ingham and Martinez-Arias, 1986; Irish et al., 1989; Tremml and Bienz, 1989). For example, the expression boundaries of \textit{Ubx} are determined by some of the gap and pair-rule genes. The products of the gap genes \textit{hunchback} and \textit{tailless} repress \textit{Ubx} transcription. The \textit{hunchback} protein establishes the anterior boundary of \textit{Ubx} expression and \textit{tailless} sets its posterior boundary (Reinitz and Levine, 1990; White and Lehmann, 1986). Conversely, \textit{fushi tarazu, even skipped} and \textit{Krüpple}, members of the pair-rule and gap gene families, activate \textit{Ubx} expression in its specified expression domain (Ingham et al., 1986; Ingham and Martinez-Arias, 1986; Tremml and Bienz, 1989).

The segmentation genes are active only during early development. Their expression ceases after the first several hours of development, whereas the patterns of homeotic gene expression they initiated are faithfully maintained throughout development. Therefore the products of the segmentation genes
cannot account for the long-term maintenance of the expression patterns of homeotic genes throughout the rest of development (Morata and Kerridge, 1981; Wakimoto and Kaufman, 1981).

Some of the homeotic genes autoregulate their own expression. For example, *Ubx* autoregulates itself in the visceral mesoderm (Bienz and Tremml, 1988). The *Deformed (Dfd)* gene shows autoregulation throughout development and this autoregulation can account at least partly for the stability of the determined state controlled by *Dfd* (Kuziora and McGinnis, 1988). But this mode of regulation is not seen for all homeotic genes. Even in those cases where autoregulation has been shown to exist, it is not present in all tissues and at all times in development. Hence it cannot be sufficient for maintaining stable long-term expression of homeotic genes.

The homeotic genes also exhibit "cross-regulatory" relationships amongst themselves. The best known among them are the down-regulation of *Antp* by the *Ubx* protein and of *Ubx* by the *abd-A* and *Abd-B* proteins (Carroll et al., 1986; Hafen et al., 1984; Struhl, 1982; Struhl and White, 1985). For example, the *Ubx* protein represses *Antp* in the third thoracic and first few abdominal segments (Hafen et al., 1984). Experiments in which *Antp* or *Ubx* was expressed ubiquitously under the control of a heat shock promoter demonstrated that these proteins, when expressed in segments posterior to their normal expression domains, did not transform the posterior segments
into anterior ones (González-Reyes et al., 1990). These surprising results suggested that the cross-regulatory interactions observed among the homeotic genes may not be functionally important. Certainly, some other mechanism must exist to sustain the spatially restricted patterns and levels of homeotic gene expression.

Mutations in two other sets of genes exhibit homeotic transformations similar to those resulting from either the down-regulation or ectopic misexpression of multiple homeotic genes. These mutations identified genes which were grouped into either the trithorax-group (trx-G) (Capdevila and García-Bellido, 1981; Kennison and Tamkun, 1988; Shearn, 1989; Shearn et al., 1987) or the Polycomb-group (Pc-G) (Jürgens, 1985; Lewis, 1978). The phenotypes of mutations in the Pc-G and the trx-G genes suggest that they are required throughout development to maintain the expression patterns of the homeotic genes which were set by the transiently acting segmentation and gap genes.

The trithorax-G

Genes of the trx-G are required for the maintenance of stable long term expression of each homeotic gene within its expression domain (Duncan, 1987; Shearn, 1989; Shearn et al., 1987). The role of the trithorax (trx) gene as a positive regulator
of the homeotic genes was initially inferred from phenotypes observed in *trx* mutants (Capdevila and García-Bellido, 1981; Ingham, 1985; Ingham and Whittle, 1980). The *trx* mutant embryos and adults exhibit transformations similar to the loss of function phenotypes of multiple homeotic genes. The strongest transformations observed are those of the structures of the third thoracic segment into the structures of the second thoracic segment. For example, transformation of the haltere, a third thoracic structure into a wing, which is a second thoracic structure. These are similar to *Ubx* mutant phenotypes. Posterior abdominal segments show transformations towards the more anterior ones similar to phenotypes caused by mutations in the *abd-A* and *Abd-B* genes (Breen and Harte, 1991; Capdevila and García-Bellido, 1981; Ingham and Whittle, 1980).

**Expression of homeotic genes in *trx* mutants**

Direct analysis of homeotic gene expression patterns in *trx* mutant embryos further confirms that the *trx* gene is involved in maintaining the expression of homeotic genes in their normal expression domains (Breen and Harte, 1991; Breen and Harte, 1993). Staining of the *trx* embryos with antibodies recognizing different homeotic gene products shows reduced expression of these genes (Breen and Harte, 1993). For example, *Ubx* expression is reduced dramatically in the ectoderm and in the
nervous system. *Antennapedia* expression is reduced in the ectoderm, nervous system and visceral mesoderm. The expression patterns of different homeotic genes in *trx* null mutants indicates that their expression is affected differently in various tissues (Ingham, 1985; Breen and Harte, 1993). This suggests that the homeotic genes have tissue-specific requirements for *trx*, implying that the mechanism underlying the regulation of homeotic gene expression by *trx* is very complex.

The *trx* gene product is required throughout development

To test if the requirement for *trx* is limited to a certain time period during development, mosaic animals containing clones of cells homozygous for various *trx* alleles were created (Capdevila and García-Bellido, 1981; Ingham, 1985). Such clones were identified by a cuticular marker and were induced at different stages of development by inducing mitotic recombination with X-ray irradiation. The time of X-ray irradiation determines the time of development at which cells are made homozygous for *trx*. This type of clonal analysis indicated that *trx* is required not only during embryogenesis, but throughout development. This expression pattern is consistent with its role in maintaining homeotic gene expression throughout development.
Genetic interactions among the trx-G genes

The trx-G genes show genetic interaction among themselves and cause similar homeotic transformations. The *absent small homeotic-1* (*ash1*) and *absent small homeotic-2* (*ash2*) genes belong to the trx-G as they cause homeotic transformations similar to those seen in *trx* mutants (Shearn *et al.*, 1987). Double heterozygotes of *ash1*, *ash2* and *trx* show more severe phenotypes than heterozygotes for a mutation in any single gene. This suggests that these genes affect a common process and may be functionally related (Shearn, 1989). Mutations in *trx* suppress the homeotic transformations caused by the Pc-G genes. Like *trx*, mutations in *ash1* and *ash2* also suppress the extra sex combs phenotype of *Pc*\(^{3}/+\) mutants (Shearn, 1989).

Kennison and Tamkun (1988) isolated genes which act as dominant suppressors of *Pc*. They isolated 13 previously unidentified genes as well as new alleles of *trx*, thus defining the trx-G. Despite the large number of mutations recovered in this screen, it is unlikely that all the genes of the trx-G have been identified. For example, they did not recover any alleles of *ash1* and *ash2*, both of which suppress the *Pc* phenotype (Shearn, 1989). Based on the genetic interaction amongst the genes of the trx-G and their interaction with the Pc-G genes, Shearn (1989) proposed that the products of the trx-G genes might form a multimeric protein complex.
Temporal and spatial expression patterns of the $trx$ transcripts

The $trithorax$ locus encodes a 23 kb primary transcript which is alternatively spliced to give 5 different mRNA species (Sedkov et al., 1994; Stassen et. al., 1995; Figure 1A). Differences in the transcripts arise from the alternatively spliced second and third exons and in the use of alternate polyadenylation sites. The alternatively spliced $trx$ mRNAs show specific temporal and spatial expression patterns. The M (10,941 nt) species is maternal and is not expressed zygotically. ME (12,519 nt) is maternal and is also present in early embryos from 2-4 hr. The expression of the embryonic mRNAs E1 (13,756 nt) and E2 (12,968 nt) peak during early embryogenesis at 2-4 hr. Weak expression of E1 is observed during late embryogenesis. The L (14,205 nt) product is abundant in 11-15 hr embryos and is detectable at low levels in larvae. The same transcript again becomes abundant in early pupae and is the major $trx$ mRNA in adults.
Figure 1A: The five alternatively spliced *trx* mRNAs: The *trx* mRNAs show temporal and spatial expression pattern. The 14.2 kb mRNA (L) encoding the TRXII isoform is expressed ubiquitously throughout development.
Figure 1B: The predicted \textit{trx} protein isoforms. The 5 \textit{trx} mRNAs encode two protein isoforms with estimated molecular weights of 365 (TRXI) and 400 kD (TRXII). Both protein isoforms contain several novel motifs which are conserved in the human \textit{ALL-1} gene and its mouse homologue, \textit{All-1}. The putative DNA binding domain is not shared by \textit{ALL-1} and \textit{All-1} proteins and is homologous to the DNA binding domain of the nuclear receptor superfamily.
The *in situ* hybridization of digoxigenin-labeled RNA probes in wild type embryos showed restricted spatial patterns of different *trx* mRNA species (Sedkov *et al.*, 1994; Stassen *et al.*, 1995). Early zygotic expression of *trx* is first apparent at syncytial blastoderm. This expression is limited to the posterior half of the embryo between 50-20% egg length and is present as 4 stripes. These stripes overlap with the expression domains of *Ubx*, *abd-A* and the anterior part of *Abd-B* expression domain. The probe specifically detecting E2 at the early stages shows a uniform expression pattern throughout the embryo. The probe detecting L mRNA showed expression in late embryos and was distributed throughout the embryo with higher expression in the central nervous system.

All the different *trx* mRNAs are predicted to encode only two unusually large protein isoforms with estimated molecular weights of 365 and 405 kD (Figure 1B). The smaller protein isoform, TRXI starts in exon 4 and the larger protein isoform, TRXII, starts in exon 3. Thus TRXI and TRXII differ only at their amino termini with TRXI being a complete subset of the TRXII isoform. The first and second exons are noncoding.
The *trx* protein isoforms contain sequences homologous to proteins from other organisms

The predicted sequence of the *trx* proteins contains several novel motifs which are conserved in the human *ALL-l* gene and its mouse homologue *All-l* (Gu et al., 1992; Ma et al., 1993; Tkachuk et al., 1992). Disruptions of the human *ALL-l* gene are associated with a number of acute leukemias. One of the novel motifs shared by *trx*, *ALL-l* and *All-l* involves a region of approximately 700 residues which are rich in cysteines (Mazo et al., 1990; Stassen et al., 1995). This novel C4HC3 zinc finger motif occurs four times in *trx* and is also present two times in *Polycomblike* (*Pcl*), a Pc-G protein (Stassen et al., 1995).

The other motif shared by *trx*, *ALL-l* and *All-l* proteins comprises the 150 carboxy-terminal residues of each protein. This region is highly conserved across plant and animal kingdoms, indicating that it has an ancient origin and very likely is a functionally important domain. This carboxy terminal domain is also present in *E(z)*, a Pc-G gene (Jones and Gelbart, 1993) and *Su(var)3-9*, a modulator of Position Effect Variegation (PEV) (Tschiersch et al., 1994). This suggests that the mechanism of action of *trx* may share some common features with these other proteins.

Another motif present in *trx*, which is not shared by the *ALL-l* and *All-l* proteins is homologous to the DNA binding domain
of the nuclear receptor family (Stassen et al., 1995). The trithorax protein does not show any other homology to nuclear receptors outside of this DNA binding domain.

The trx proteins also contain several PEST sequences in their amino-terminal half. PEST sequences are found in proteins with short lives and are believed to act as a signal for rapid proteolytic degradation (Stassen et al., 1995).

A member of the trx-G, brahma, encodes a protein which shares homology with ATP-dependent helicases

Study of the other trx-G genes may give us some insight into how trx might function since they all have similar phenotypes and show genetic interactions. The brahma (brm) gene was isolated as a suppressor of the Pc phenotype (Kennison and Tamkun, 1988). It shows genetic interaction with trx similar to ash1 and ash2 (Tamkun et al., 1992).

The brm protein shares homology with ATP-dependent DNA helicases and has homologues in yeast and in humans (Muchardt and Yaniv, 1993; Tamkun et al., 1992). The yeast homologue of brm, SWI2, functions as a part of a multiprotein complex involved in activation of several yeast genes including HO (Peterson and Herskowitz, 1992). Phenotypes of mutations in SWI2 or any other member of the SWI/SNF complex are suppressed by mutations in histone genes. This suggests that
SWI/SNF complex may be acting by weakening DNA:histone contacts and changing the local chromatin structure. This may disrupt the nucleosomes and make binding sites available for the dedicated activators. None of these proteins bind DNA directly but they facilitate binding of the dedicated sequence-specific DNA binding activator proteins to their target binding sites.

Recently, the human SWI/SNF complex was isolated from HeLa cell extract and was shown to exhibit ATP-dependent helicase activity indicative of presence of hSWI2 (Kwon et al., 1994). This complex disrupts nucleosomes and assists a GAL4-VP16 activator and the TATA binding protein (TBP) to bind their respective binding sites (Imbalzano et al., 1994; Kwon et al., 1994). Similar to SWI2, brm is required for maintaining activation of multiple homeotic genes in Drosophila and could be a functional homologue of SWI2 (Winston and Carlson, 1992). It may be acting by changing the local chromatin structure at the homeotic gene promoters, making them accessible to other factors.

*Trithorax-like* encodes the GAGA factor

*Trithorax-like* (*Trl*), which is another member of the trx-G, encodes the GAGA factor (Farkas et al., 1994). The *Trl* gene is required for the normal expression of homeotic genes. It binds
to many discrete loci on polytene chromosomes including the heat shock loci (Tsukiyama et al., 1994). The GAGA factor is a sequence-specific DNA-binding protein and stimulates transcription of Ubx in vitro (Biggin and Tjian, 1988). It acts as an "antirepressor" by counteracting histone-mediated repression of RNA pol II transcription (Croston et al., 1991). Binding of the GAGA factor to its sites on the hsp70 promoter creates a constitutive DNaseI hypersensitive site in vitro and in vivo, suggesting a structural change in the chromatin at this site (Tsukiyama et al., 1994). Deletion of GAGA binding sites in the hsp26 gene greatly reduces its inducibility and abolishes constitutive hypersensitive sites (Lu et al., 1993). All these data suggests that the GAGA factor acts by disrupting nucleosomes and thereby making DNA accessible to other activators. Its binding to the hsp70 promoter even in the absence of a heat shock indicates that the binding of GAGA to its sequences is constitutive.

The GAGA protein contains a zinc-finger DNA binding motif and shares a region of significant homology, the POZ domain, with other Drosophila proteins such as tramtrack, kelch and Broad-complex (DiBello et al., 1991; Harrison and Travers, 1990; Xue and Cooley, 1993). The POZ domain is shown to mediate homodimeric protein-protein interactions (Bardwell and Treisman, 1994).
The nucleosome disruption by the GAGA factor is ATP-dependent, but it does not contain any known ATP-binding motif (Tsukiyama et al., 1994). It is therefore possible that it may be complexing with another protein(s) such as brahma which might assist GAGA by fulfilling the ATP-dependent function.

**The ash1 protein binds polytene chromosomes at specific sites**

Another trx-G protein, ash1, also shares homology to the 150 carboxy-terminal residues of trx. It has been recently shown to bind polytene chromosomes at approximately 100 sites including the BX-C and the ANT-C loci. This indicates that it is also a chromosomally associated protein and acts directly on its target genes (Allan Shearn, personal comm).

The molecular characterization of these other trx-G genes suggests that these proteins may act at the level of chromatin. They may form multiprotein complexes which are targeted to specific sites on chromosomes. The presence of domains which might be involved in protein-protein interactions together with the genetic interactions observed among these genes lends support to the hypothesis that they form a complex.
The trx-G proteins and modifiers of PEV may share some features of gene regulation

Some of the genes of trx-G show an additional phenotype of suppression or enhancement of Position Effect Variegation (PEV) (Farkas et al., 1994). PEV is a result of a chromosomal rearrangement in which a euchromatic region is placed in the vicinity of the centromeric heterochromatin. When genes that lie within euchromatin are brought close to heterochromatin by a chromosomal rearrangement, they are subjected to an inactivating influence of the heterochromatin (Spofford, 1976). The heterochromatinization of the euchromatic region is cytologically visible in the salivary gland polytene chromosomes (Hartmann-Goldstein, 1967). This heterochromatinization which is thought to repress transcription occurs only in some cells and is clonally inherited, resulting in a mosaic pattern of expression/repression.

A number of unlinked genes have been identified, mutations in which either suppress or enhance PEV (Locke et al., 1988; Spofford, 1976). The suppressor genes are thought to encode proteins normally involved in chromatin condensation. Consistent with this idea, it was proposed that a decrease in the amount of suppressor protein will reduce the condensation of chromatin and increase in it will enhance the condensation process. A similar but opposite dosage effect is expected for the
enhancers of variegation which are proposed to encode proteins involved in decondensation of chromatin (Reuter and Spierer, 1992). Genetic analysis of several of these modifier genes showed that they do exert dosage-dependent effects on PEV (Locke et al., 1988; Wustmann et al., 1989). Some of the suppressor genes suppress variegation when present in a single copy and enhance it when present in three copies.

Based on this genetic data and observed cytological changes associated with PEV, it was proposed that the modifier proteins are the structural components of chromatin. The suppressor proteins might form an ordered assembly of modules which spread over large chromosomal regions forming heterochromatin. Enhancer proteins, in contrast, are required for maintenance of the active chromatin conformations (Locke et al., 1988).

Cloning and molecular characterization of several of these genes demonstrated that they encode proteins which are associated with the chromosomes in heterochromatic regions and/or show homology to other DNA binding proteins or transcription factors (Dorn et al., 1993; Eissenberg et al., 1990; Reuter et al., 1990; Tschiersch et al., 1994). For example the product of the gene Su(var)205, the Heterochromatin protein (HP1), is associated with polytene chromosomes within the centric \( \beta \)-heterochromatin and throughout chromosome 4 which is highly condensed and heterochromatic (James et al., 1989).
An enhancer of PEV, \(E(var)3-93D\), is bound to polytene chromosomes in euchromatic regions (Dorn \textit{et al.}, 1993). It shares homology with the POZ domain which raises the possibility that it may recruit other factors, possibly through the POZ domain, to form complexes involved in chromatin modeling (Dorn \textit{et al.}, 1993). Like the \(\text{trx-G}\) genes, mutations in \(E(var)3-93D\) display homeotic transformations suggesting its role in positive regulation of the BX-C and the ANT-C genes (Dorn \textit{et al.}, 1993). It exhibits genetic interactions with other \(\text{trx-G}\) genes suggesting that the enhancers of PEV and the \(\text{trx-G}\) genes may be functionally similar.

Molecular characterization of some of the modifier genes has also revealed that they share sequence homology with some of the \(\text{trx-G}\) and \(\text{Pc-G}\) proteins. The HP-1 protein shares a domain called the "chromo domain" with the \textit{Polycomb} (\(\text{Pc}\)) protein (Paro and Hogness, 1991). The chromo domain is an evolutionarily conserved protein motif found in species ranging from plants to man suggesting the possible functional importance of this domain (Singh \textit{et al.}, 1991). The \(\text{Pc}\) chromo domain is required for binding of \(\text{Pc}\) protein to specific sites on polytene chromosomes (Messmer \textit{et al.}, 1992). The homology suggests that the modifiers of PEV and \(\text{Pc-G}\) proteins may share a common mechanism of transcriptional repression.

As mentioned earlier, the predicted \(\text{trx}\) proteins contain a novel motif which is present in \textit{Enhancer of zeste} (\(E(z)\)), a \(\text{Pc-G}\)
protein and Su(var)3-9, a modifier of PEV (Jones and Gelbart, 1993; Stassen et al., 1995; Tschiersch et al., 1994). The \(E(z)\) protein is associated with polytene chromosomes and is likely to be a part of multiprotein complexes formed by the Pc-G proteins (R. Jones, personal communication; Rastelli et al., 1993). Thus based on the homology between \(tr\), \(E(z)\) and \(Su(var)3-9\) we predict that the \(tr\) proteins may act by associating with chromosomes.

Based on the similarity in function and the sequence homologies, it can be predicted that the \(tr\)-G and enhancers of PEV may behave similarly and the Pc-G and the suppressors of PEV may share some common mechanisms of gene repression. Indeed some of the \(tr\)-G gene mutations show enhancer of PEV phenotypes and some Pc-G mutations show suppressor of PEV phenotypes (Farkas et al., 1994; unpublished results cited in DeCamillis et al., 1992).

**The Polycomb-Group**

The Pc-G is predicted to contain 30 to 40 genes. Mutations in the Pc-G genes cause related phenotypes indicative of global derepression of multiple homeotic genes (Jürgens, 1985). Thus, the Pc-G genes are required to maintain stable long-term repression of each homeotic gene outside of its normal expression domain (Ingham, 1984; Struhl, 1981). Polycomb (Pc)
is the prototypical member of the Pc-G. In Pc mutants, the initial establishment of the homeotic gene expression domains occurs normally (Struhl and Akam, 1985). But after germ band extension, when expression of the segmentation genes begins to decay, the spatially restricted domains of homeotic gene expression break down. At this time all the homeotic genes are expressed in cells throughout germ band (Wedeen et al., 1986).

Six of the Pc-G genes have been isolated and characterized. The proteins encoded by these genes bind to the polytene chromosomes of the larval salivary glands, where they are colocalized at approximately 100 specific chromosomal sites (DeCamillis et al., 1992; Martin and Adler, 1993; Rastelli et al., 1993; Zink and Paro, 1989). These sites include their known target genes in the BX-C and ANT-C indicating that the Pc-G products act directly on their target loci. They may be associated with chromosomes by binding DNA either directly or indirectly through association with other DNA binding proteins. None of these proteins have demonstrable sequence-specific DNA binding activity in vitro (Paro, 1993; V. Pirrotta personal communication).

The Pc-G genes share some common features with the modifiers of PEV. In both cases large number of genes are involved which are predicted to form multiprotein complexes (Locke et al., 1988). Mutations in different Pc-G genes show synergistic interactions suggesting that they may be part of a
multiprotein complex (Jürgens, 1985). The *posterior sex combs* (Psc) and *Suppressor of zeste 2* (Su(z)2) proteins contain a novel Zn finger motif called the "RING" finger which has been shown to mediate protein-protein interactions in a yeast two hybrid system (H. Brock personal comm.). This domain is suggested to be a DNA-binding domain in other proteins but DNA binding has yet to be established (Lovering, et al., 1993; Schwabe and Klug, 1994). The *Polycomb-like* (Pcl) protein shares homology with the *trx* Cys-rich region, which contains several novel double C₄H₃C₃ zinc finger motifs suggesting that the positive and negative regulators of homeotic gene expression may share some common features of their mode of action (Stassen *et al.*, 1995). Co-immunoprecipitation studies and immunostaining of polytene chromosomes have revealed that the Pc-G proteins form a multiprotein complex (Franke *et al.*, 1992; Rastelli *et al.*, 1993; H. Brock, pers. comm.).

Several modifiers of PEV show homeotic phenotypes and mutations in some of the Pc-G genes modify PEV (Reuter and Spierer, 1992; unpublished results cited in DeCamillis *et al.* 1992) These findings have led to the speculation that these two groups of proteins may function analogously.

A direct link between the modifiers of PEV and the Pc-G proteins was established when the *Pc* protein was shown to contain the "chromo domain" homologous to that present in the HP-1 protein. The chromo domain is required for the
chromosomal binding of the Pc protein (Messmer et al., 1992). The Su(var)3-9 protein contains the chromo domain and the carboxy terminus homology present in E(z), the Pc-G protein and trx (Tschiersch et al., 1994). These sequence similarities may represent the functional similarities between these proteins.

Orlando and Paro (1993) demonstrated that in Drosophila Schneider line 2 (S-2) tissue culture cells, the Pc protein binds extensively throughout the Ubx and abd-A, genes but not Abd-B. Indeed the Ubx and abd-A genes, but not Abd-B, are repressed in this cell line. This suggests that the Pc-G protein complexes are spread over a large chromosomal region. All these data suggest that the Pc-G proteins act at the chromosomal level, possibly by altering the local chromatin structure.

**Genetic interaction of trx with the Pc-G genes**

Mutations in trx show antagonistic genetic interactions with mutations in the Pc-G genes, repressors of homeotic gene expression. The dominant derepression phenotypes of Polycomb (Pc\(^3/\)) heterozygotes are enhanced in the presence of three wild type copies of trx and are suppressed when only one copy of trx is present (Capdevila and García-Bellido, 1981). Similarly, the abdominal transformations caused by the complete loss of maternal and zygotic expression of extra sex combs (esc), a member of the Pc-G, is suppressed in the absence of wild type
tron function (Ingham, 1983). Another Pc-G gene, Enhancer of zeste (E(z)), shows similar suppression of derepression phenotypes in E(z) and trx double mutants (Jones and Gelbart, 1993).

Such genetic interactions between these positive and negative regulators of homeotic genes have led to a hypothesis that they may compete with each other to maintain the correct expression pattern of the homeotic genes. The genetic interactions also imply that the maintenance of homeotic gene expression may require a proper balance of these positive and negative regulatory factors. Detailed study of the protein products of these genes will be helpful in elucidating whether these genetic interactions reflect functional interactions between the trx and Pc-G proteins.

**Summary and thesis outline**

In summary, study of some of the trx-G genes has suggested their possible role in chromatin organization. In view of their role as positive regulators they might assist in formation of the "open" chromatin domains. The genetic interactions between trx and other members of the trx-G suggests that the proteins encoded by these genes may act by forming multiprotein complexes directly bound to their target genes.
Presence of a homologous domain in \textit{trx} and \textit{E(z)} and \textit{Pcl}, the Pc-G proteins, suggests that the positive and negative regulators may share some features of their action in maintaining homeotic gene expression. The presence of a domain homologous to the \textit{trx} C-terminal domain in \textit{Su(var)3-9} and the observation that \textit{E(var)3-93D} mutations show \textit{trx} like phenotypes suggests that \textit{trx} may behave similar to the modifiers of PEV in regulating the homeotic gene expression.

The \textit{trx} mutations show antagonistic genetic interaction with the Pc-G mutations. It has been proposed that the \textit{trx-G} and Pc-G proteins compete with each other for the maintenance of homeotic gene expression. Detailed study of the \textit{trx} proteins should help to determine whether this is so and at which level competition between these regulatory factors occurs.

In the course of the research described in this thesis, we developed anti-\textit{trx} antibodies and used them to characterize the \textit{trx} proteins. We tested if the \textit{trx} proteins, like the other \textit{trx-G} and Pc-G proteins, bind polytene chromosomes at discrete sites. This polytene chromosome assay was used to answer several questions: Does \textit{trx} act directly on its targets? Do \textit{trx} proteins co-localize with the Pc-G proteins or other \textit{trx-G} proteins on polytene chromosomes? Do they bind to the regulatory regions of homeotic genes?

To test if the \textit{trx} proteins bind to the regulatory regions of homeotic genes we initially used the polytene chromosome
assay rather than \textit{in vitro} DNA binding assays. The large size of the \textit{trx} proteins makes it difficult to purify them or express them in \textit{E.coli}. Also the regulatory regions of the homeotic genes are very large and hence difficult to analyze by \textit{in vitro} DNA binding assays until \textit{trx} response elements are better characterized. Transformants containing \textit{P} element constructs carrying various DNA fragments from the \textit{Ubx} regulatory region inserted at random locations in the genome are available. We immunostained polytene chromosomes from a selection of these transformants and localized a \textit{trx} binding site to a 670 bp fragment within the \textit{Ubx} gene. This fragment was then used in \textit{in vitro} DNA binding experiments described in chapter IV. This \textit{trx} protein characterization will help in understanding its mode of action in regulating homeotic gene expression.
Aims of the thesis:

1) To develop anti-trx antibodies to analyze the trx proteins in vivo and in vitro.

2) To determine whether the trx proteins are associated with polytene chromosomes and if so, whether they co-localize with the Pc-G proteins and other trx-G proteins.

3) To map trx binding sites within one of its well characterized target genes, the Ubx gene. This was done by immunostaining polytene chromosomes from transformants carrying P element constructs containing DNA fragments from the Ubx regulatory region integrated into their genome.

4) Finally, to determine whether trx is a sequence-specific DNA binding protein and, if so, to determine the sequence of its binding site.

The above mentioned experiments were designed to give some clues to the mechanism of action of trx and are explained in detail in the following chapters.
CHAPTER II: Immuno detection of *trx* proteins

INTRODUCTION:

The *trx* gene is required for the long-term maintenance of homeotic gene expression within specified domains. As mentioned in Chapter I, the two *trx* protein isoforms contain several novel motifs whose function is not yet determined. To characterize the *trx* proteins further we developed antibodies against two different parts of the large *trx* proteins. One of them was raised against the amino-terminal residues present only in the TRXII isoform and hence will recognize only the TRXII protein (isoform-specific antibodies). The other antibody was raised against *trx* residues which are present in both protein isoforms and will therefore recognize both TRXI and II isoforms (common antibodies). The isoform-specific antibodies were prepared to test whether this isoform has an expression pattern which is a subset of the expression pattern seen with the common antibodies.

The respective DNA fragments were cloned into bacterial expression systems, the fusion proteins were purified and injected into rabbits. These antibodies were used to test if the *trx* proteins are bound to chromosomes like the other regulators of homeotic gene expression.
Immunostaining of polytene chromosomes

Immunostaining of polytene chromosomes has been used as an assay to determine if a given protein is associated with the chromosomes and to determine its target genes by cytological mapping of binding sites (DeCamillis et al., 1992; Rastelli et al., 1993; Zink and Paro, 1989). The salivary gland chromosomes of Drosophila larvae undergo polytenization. At the end of third instar larval stage, there are 500-1000 copies of chromosomes which have not been separated after DNA synthesis. These are in perfect registration with each other and appear as a single giant chromosome. The polytene chromosomes have a distinct morphology and a stereotypical banding pattern which helps in identification of each chromosome arm and mapping of loci on the chromosomes (Leferve, 1976). A typical chromosome spread shows five chromosome arms, X, 2L, 2R, 3L and 3R, all attached at their centromeres forming a structure called chromocenter. The small fourth chromosome is usually attached to the base of 3R near the chromocenter. The Y chromosome is difficult to visualize in such preparations.

Immunostaining of polytene chromosomes was developed for in situ determination of chromosomal proteins (Silver and Elgin, 1976). It has been since used for many different proteins, including the proteins of the Pc-G and the trx-G. The formaldehyde fixation preserves the protein distribution along
the chromosome arms and hence the immunostaining indeed represents the in vivo distribution of a given protein. The polytene chromosome assay is useful in asking a variety of questions. The cytological mapping helps in determination of the target genes.

Immunostaining of the polytene chromosomes using anti-trx antibodies demonstrated that the trx proteins are associated with chromosomes at discrete sites. A temperature-sensitive trx mutant shows a reduced number of binding sites indicating the specificity of the antiserum used. Simultaneous localization of trx with either Pc or ash1 showed that the trx proteins colocalize with these proteins at many sites.
MATERIALS AND METHODS:

Preparation of immunogens for induction of antibodies in rabbits:

Two different polyclonal antibodies were used in this study. One of them was isoform-specific, recognizing the larger protein isoform (TRXII) and the other recognizing both TRXI and II isoforms.

"Common" region immunogen: The "common" immunogen consisted of trx residues 2355-2739, present in both trx isoforms. An 800 bp EcoRI-BamHI fragment encoding these residues was subcloned into pATH III, a trpE fusion protein vector (Koerner et al., 1991). The trx-trpE fusion protein was prepared as follows: 50 ml of RR1 cells containing the trx-trpE construct were grown overnight at 37°C in M9+W+CAA+amp medium (1x M9 salts, 0.5% Casa amino acid (CAA), 1 mM MgSO4, 0.1 mM CaCl2, 0.2% glucose, 0.1 mg/ml thiamine, 0.02 mg/ml tryptophan, 100 mg/ml ampicillin; 10x M9 salts: 60 g Na2HPO4, 30 g KH2PO4, 5 g NaCl and 10 g NH4Cl in 1 liter of distilled water). The overnight culture was diluted in 500 ml (1:10) of fresh M9+CAA+amp medium, without tryptophan (W).

Depletion of the co-repressor, tryptophan, induces the operon by preventing binding of the trp repressor to the operator. The cells were allowed to grow for 1 hour at which point Indole
Acrylic Acid (IAA) was added to the culture at the final concentration of 10 μg/ml. IAA is a tryptophan analog which competes with tryptophan for the trp repressor, but does not activate the repressor to recognize the trp operator. It also interferes with the charging of tRNA_{trp} by inhibiting tryptophenyl-tRNA synthetase. Addition of IAA to the culture results in limitation of tryptophan availability and a rapid derepression of the trp operon by relief of both repression and attenuation.

The cells were allowed to grow further for 3 hours at 37°C and then harvested by centrifugation in a GSA rotor for 10 min., at 5K rpm, 4°C. The pellet was resuspended in 5 ml of either TE, pH 7.5 (10 mM Tris-Cl, 1 mM EDTA) or in a urea buffer (0.01 M Sodium Phosphate pH 7.2, 1% β mercaptoethanol, 1% SDS and 6 M urea). The suspended cells were dispensed in 1.5 ml microfuge tubes. A 3X sample buffer (150 mM Tris-Cl pH 6.8, 6 mM EDTA, 3% β mercaptoethanol, 24% glycerol and 0.075% bromophenol blue) was added to each tube to make the final concentration to 1X. Tubes were stored at -20°C. The fusion protein was separated by electrophoresis through an 8% SDS-PAGE gel (Sambrook et al., 1989). Samples were boiled for 3 min in a water bath before loading onto the gel.

To purify the immunogen, the \textit{trx-trpE} fusion protein was separated on 8% SDS-PAGE preparatory gels (Biorad mini protean II). The gels were stained with coomassie blue (0.25%
coomassie blue, 10% acetic acid, 50% methanol) for 1-2 hr and
destained (methanol: acetic acid: double distilled water, 1:1:8)
for 2-3 hr. The fusion protein band was cut out of the gel.

Several such bands were used to electroelute the fusion
protein by using a Biorad electroleluter. A sample of the
electroeluted fusion protein was again run on an 8% SDS-PAGE
gel to check the purity of the sample and to estimate the protein
concentration. BSA standards were run on the same gel to
estimate the concentration of the fusion protein.

The purified \textit{trx-trpE} fusion protein was injected into two
young New Zealand female rabbits. The rabbits were boosted
after first 15 days and then every month for almost a year. The
first bleed was collected after a month and subsequent bleeds
were collected weekly.

\textbf{Partial purification of the total serum:} The total serum was
depleted of anti-\textit{trpE} antibodies by preabsorbing the serum with
proteins from cells containing the pATHIII vector (Kelly \textit{et al.},
1986). The RR1 cells containing the pATHIII vector were
induced and grown as described for the \textit{trx-trpE} fusion protein.
The total protein extract was run through an 8% SDS-PAGE
preparatory gel (0.7 ml of extract +3X loading dye) and then
transferred onto a 0.2 \mu nitrocellulose paper by electroblotting
for 2 hr at 80 amps, on ice (Biorad transfer system). The protein
blots were incubated overnight at 40\textdegree C, in serum diluted 1:20 in a
RPB buffer (10 mM Tris-Cl pH 8.0, 0.15 M NaCl, 1 mM EDTA,
0.1% triton X-100). Before incubation, the total sera, immune and preimmune, were first allowed to preabsorb to *Drosophila* embryos at 1:1 dilution in phosphate buffered saline (PBS: 0.13 M NaCl, 7 mM NaH$_2$PO$_4$ and 3 mM NaH$_2$PO$_4$, 5% BSA and 0.1% tritonX-100). This preincubation eliminates nonspecific antibodies present in animal serum. The supernatant serum after incubation with blots was stored at -20°C.

**Isoform specific immunogen**: The specific immunogen was made from the *trx* residues 1-172 which were cloned as a 500 bp PCR fragment into a pQE-9 vector (Qiagen). The pQE-9 vector adds six histidine residues to the N-terminus of the protein. The histidine residues bind Ni++ ions and thus help in purification of the fusion protein by Ni-NTA sepharose column chromatography (Hochuli and Piesecki, 1992; LeGrice and Grüninger-Leitch, 1990). In the pQE-9 vector, expression of the fusion protein is under the control of the lac operator. The host cells, PDM1.1, contain a plasmid which makes the lac repressor, thus ensuring that there is no basal expression of the fusion protein.

PDM1.1 cells with the recombinant pQE-9 were grown overnight in 5 ml of LB+100 µg/ml ampicillin (amp)+25 µg/ml kanamycin (kan) at 37°C. The overnight culture was diluted 1:100 (500 ml) in LB+amp+kan and allowed to grow at 37°C to an O.D. of 0.8 at 600 nm. The cells were then induced with 1.5 mM Isopropyl-β-D-thiogalactopyranoside (IPTG) and allowed to grow for another 3 hours. At the end of induction, cells were
harvested by centrifugation at 5K rpm, 40°C for 10 min. in a GSA rotor.

The supernatant was decanted off and the pellet was resuspended in 100 ml (10 ml/50 ml of culture) of 50 mM Tris-
Cl, 75 mM NaCl pH 7.5. Cells were pelleted at 8K rpm, 40°C and for 10 min. in a SS34 rotor. The pellet was weighed and resuspended in 6M guanidium hydrochloride (GnHCl) in 0.1 M sodium phosphate buffer (buffer A) (Sambrook et al., 1989) at the concentration of 5 ml/g of pellet. The solution was stored overnight at -20°C. Next day the cells were thawed slowly on ice.

A column was packed with 1 ml of Ni-NTA sepharose and equilibrated in 10 ml of buffer A, at approximately 3 ml/hr. The supernatant from cell lysate was then loaded onto the column at 3 ml/hr. The column was washed serially with 10 ml of buffer A, 10 ml of buffer B (6M GnHCl in 0.2 M ammonium acetate pH 6.0) and 10 ml of buffer C (6M GnHCl in 0.2 M ammonium acetate pH 4.0). Fractions of 0.5 ml were collected for washes with buffers B and C. The fusion protein starts eluting in the middle of buffer B wash.

From each elution fraction 100 µl was used for trichloric acid (TCA) precipitation of protein by adding an equal volume of ice-cold 100% TCA. The precipitate was pelleted by spinning for 10 min at 13K rpm in a microfuge and the pellet was
resuspended in the 1X sample buffer. Samples were run on 10% SDS-PAGE gel to identify fractions containing the fusion protein.

Fractions with the fusion protein were pooled together and dialyzed overnight against 50 mM sodium phosphate buffer pH 7.8 at 4°C. These fractions were found to contain a few other bacterial proteins. Therefore further purification was carried out by running through 10% SDS-PAGE gels, cutting the fusion protein bands and then performing electroelution as noted above for the "common" immunogen preparation. The column was stored at 4°C with buffer A and 0.05% sodium azide to prevent bacterial growth.

The isoform-specific immunogen was also injected into two young female rabbits and the injection and collection schedule was same as that for the common antibody.

**Western Analysis**

**Western analysis of bacterial protein extract**: Western analysis was carried out using total protein extracts from bacterial cultures which were either uninduced or induced to synthesize fusion proteins.

**Western analysis using "common" anti-trx antibodies**: Induction of the *trx-trpE* fusion protein was carried out as described earlier. At the end of the induction period, 1 ml of culture was spun down and the pellet was resuspended in 100 μl of TE and
50 µl of 3X sample loading buffer. This suspension was further
diluted 1:10 and 1 µl was loaded on an 8% SDS-PAGE gel. As a
control, identically prepared extracts of bacteria containing
pATHIII vector alone was run on the same gel. This extract
should contain over-expressed trpE protein with estimated
molecular weight of 37 kD. The gel was run until the dye front
ran off. Electrophoresed proteins were transferred onto a 0.2 µ
nitrocellulose membrane for 4 hr on ice, at 80 mA in a transfer
buffer (39 mM glycine, 48 mM tris base, 0.03% SDS and 20%
methanol) using Biorad electroblotter (Sambrook et al., 1989).
Further incubations and color developments were carried out
according to Benson and Pirotta (1987) with few modifications.

The membranes were blocked in 5% non fat dry milk and TBT
(10 mM Tris pH 7.4, 0.9% NaCl and 0.1% tween-20) in a "seal-a-
meal" bag for either 1 h at room temperature (RT) or overnight
at 4°C. They were then incubated with primary antibodies in
TBT + 5% milk at final dilution of 1:10,000 for 1 h at RT. The
primary antibodies used here were common anti-trx antiserum
and control preimmune serum from the same animal. Both were
preincubated with the extracts from bacteria expressing only
the trpE protein, as mentioned before. At the end of incubation,
the membranes were washed twice for 15 min. each in TBT + 5%
milk and were then incubated with 1:1 preincubated
biotinylated goat anti-rabbit secondary antibodies at 1:2000
dilution for 1 h at RT (see material and methods for
immunostaining of chromosomes). The membranes were washed once for 15 min. in TBT + 5% milk and then in TBT alone for 15 min. The membranes were further incubated for 1 h at RT with premixed solutions A and B for alkaline phosphatase (Vectastain alkaline phosphatase kit). Solutions A and B contain biotin and streptavidin-alkaline phosphatase which binds to the biotinylated secondary antibodies. Two drops (100 μl) of solutions A and B each were added to 5 ml of PBT and incubated for 30 min. on ice before adding to the membranes. The membranes were washed twice for 10 min. in PBT at the end of incubation and then color was developed as follows.

The blots were incubated in a coloring solution (10 ml of Alkaline phosphatase buffer, 66 μl of NBT and 33 μl of BCIP) containing NBT and BCIP which are the substrates for alkaline phosphatase (Sambrook et al., 1989). The enzyme reaction gives a blue colored product. The blots were incubated in the coloring solution until the blue protein bands were developed. The reaction was stopped by washing the membranes in distilled water and dried. The blots were stored at room temperature.

Western analysis using isoform-specific antibodies: The bacteria containing the recombinant pQE-9 were induced as before. Uninduced bacterial cultures were used as a control. At the end of incubation, bacteria were pelleted and resuspended as done for the Western analysis with the common anti-trx antibodies. Extracts were loaded on a 12% gel since the molecular weight of
the fusion protein is estimated to be 19 kD. Blotting and rest of the steps were carried out exactly as those mentioned above for the common antibodies.

**Western analysis of embryonic nuclear protein extracts:**

Embryonic nuclear extracts were used as a protein source for the fly Western analysis. Nuclear extracts from wild type *Drosophila* embryos were prepared as described in Kamakaka *et al.* (1991). Embryos were collected overnight on grape juice plates. Eight to ten grams of embryos were processed and 4 ml of final soluble nuclear fraction was dispensed into 100 μl aliquots, frozen in liquid nitrogen and stored at -80°C.

At the time of electrophoresis, the nuclear extract was thawed on ice and 50 μl of 3X sample loading buffer was added. Sample was boiled for 3 min. and 30 μl of it was loaded on a 6% N-N' Diallyltartardiamide (DATD) gel (Hames and Rickwood, 1981). To ensure the entry of high molecular weight *trx* proteins (estimated molecular weight 360 and 400 kD) into the acrylamide gel, it was run until the 180 kD molecular weight marker reached bottom of the gel. Proteins were then transferred onto a RAD free nitrocellulose membrane (Schleicher & Schuell) in transfer buffer using a Biorad electro-blotter overnight at 4°C and 50 volts. Western analysis was carried out as described in the previous section except that primary antibodies were used at 1:200 dilution.
Immunostaining of polytene chromosomes

Larvae were grown in uncrowded bottles containing standard fly medium and dry yeast pellets. For $trx^l$, mothers were kept at 29°C for two days before the egg collection was started. The $trx^l$ larvae grown continuously at 29°C were used in the study. Pairs of salivary glands from wandering third instar larvae were dissected in solution 1 (0.1% Triton X-100 in PBS pH 7.5). The fat body and other tissue was carefully removed without disrupting the salivary glands.

The glands were transferred to a drop of solution 2 (3.7% formaldehyde, 1% Triton X-100, in PBS pH 7.5) on a siliconized coverslip. The glands were fixed for 20 sec in solution 2 for common antibodies and 25 sec for the isoform-specific antibodies (time needs to be adjusted for the individual antigen). The glands were carefully moved into 40 μl of solution 3 (3.7% formaldehyde, 50% acetic acid) on a non-siliconized 18 x 18 mm coverslip and fixed for 3 min. Solutions 2 and 3 were made fresh every 2-3 hours. The coverslip was picked up by lowering a poly-L-lysine treated slide.

Cells were broken open by tapping the coverslip firmly with the eraser end of a pencil. Holding the coverslip, chromosomes were spread using a pointed tip by dragging it in a zig-zag manner across the coverslip. Excess fixative was removed by pressing slides (coverslip down) onto a blotting paper. The
chromosomes were further spread and flattened by pressing down on the slide and using firm circular motions. The position of the coverslip was marked using a diamond pencil. After freezing the slides in liquid nitrogen, coverslips were flipped off with a razor blade. The slides were washed two times for 15 min. in PBS, slowly shaking the rack. Slides can be stored at this point (up to one week) in 100% methanol or in 50% (w/v) ammonium sulfate at 4°C.

The slides were blocked for 1 hour in a blocking solution (3% BSA, 10% non-fat dry milk, 0.2% NP40, 0.2% Tween 20-80 in PBS) at RT. Forty μl of primary antibodies at a 1:50 dilution was added to each slide, covered with a coverslip and incubated overnight at 4°C in a humid chamber. The next day slides were rinsed in PBS and washed sequentially, for 15 min. in PBS, 300 mM NaCl, 0.2% NP40, 0.2% Tween20-80; and then 15 min. in PBS, 400 mM NaCl, 0.2% NP40, 0.2% Tween 20 or 80, shaking thoroughly. (If background problems persist, the NaCl concentration can be raised to 500 mM). The slides were then rinsed in PBS and were incubated with 40 μl of a 1:1000 dilution of biotinylated goat anti-rabbit secondary antibodies (Chemicon) with 2% normal goat serum in blocking solution for one hour at RT in a humid chamber. The secondary antibodies were preincubated at a 1:1 dilution with wild type Drosophila eggs to eliminate nonspecific antibodies recognizing fly proteins. After
incubation, the slides were rinsed in PBS and washed as done for the primary antibodies.

After rinsing in PBS, the chromosomes were incubated with 40 μl of solution A and solution B from the Vectastain elite kit for one hour at RT. Equal volumes of solutions A and B were mixed and incubated on ice for 30 min. before adding to the slides (Vectastain manufacturers instructions). The slides were washed twice in PBT (PBS, 0.2% tween-20) for 10 min. each and were then incubated in coloring solution for 5 min (0.5 mg/ml DAB in 10 mM tris pH 7.5, 0.01% H2O2). After color development, the slides were washed in PBT twice for 10 min. each. The chromosomes were counter-stained with Giemsa (1:100 dilution of the stock from Polysciences in 10 mM sodium phosphate buffer pH 6.8) for 1 min. The slides were washed thoroughly under running distilled water for 5 min., dried and then mounted with permount or glycerol.

**Preparation of poly-L-lysine coated slides:** Slides were soaked in a corrosive detergent solution for 1-2 h and washed under running tap water for 1 h. The slides were then washed in two changes of distilled water. They were dipped in 95% ethanol with two changes and air dried. Slides were then dipped in poly-L-lysine solution (0.1% w/v in water, Sigma) and air dried.
Co-immunolocalization of the *trx* and *Pc* proteins on polytene chromosomes

Wild type third instar larvae were dissected and the salivary glands were fixed for 15 sec and 2 min. in solutions 1 and 2 respectively. These conditions are ideal for *Pc*. The chromosomes were processed as mentioned above and incubated overnight with a 1:50 dilution of anti-*Pc* antibodies at 4°C. The next day, the slides were washed as before and then incubated for 1 hr with a 1:1000 dilution of biotinylated goat anti-rabbit secondary antibodies at RT. After thorough washing the slides were incubated overnight with a 1:50 dilution of anti-*trx* ("common") antibodies at 4°C. The slides were washed and incubated with 1:200 dilution of Rhodamine (Rh)-labeled anti-rabbit secondary antibodies for 1 hr. The Rh-labeled antibodies will recognize the anti-*trx* primary antibodies and the *trx* proteins will be visualized as red fluorescence. After washing, the slides were incubated with avidin-fluorescein (Oncor) in a dark chamber for 30 min. The avidin conjugated to FITC will recognize the biotinylated secondary antibodies thus giving green fluorescence for the *Pc* protein. The slides were washed three times for 5 min. each in PBT and mounted with antifade (Oncor) to minimize fading of fluorescence. The slides were examined by confocal microscopy. Images were collected with red and green filters separately for Rh and FITC respectively. Red fluorescence
represents \( \text{tr}x \) and green represents \( Pc \) proteins. The images were merged to demonstrate co-localization of the two proteins on chromosomes.

Since both primary antibodies used in this study were made in rabbits, incubation was carried out sequentially as noted in the protocol. To test whether yellow bands representing co-localization of \( Pc \) and \( \text{tr}x \) are really due to superimposition of red and green fluorescence and not due to an excess of the first primary used, the following control was carried out. The slides were incubated with the first primary, the anti-\( Pc \) antibodies, followed by the first secondary antibodies, biotinylated anti-rabbit antibodies. After washes, the slides were incubated with the second secondary antibodies, the Rhodamine conjugated anti-rabbit antibodies. The second primary antibodies, the anti-\( \text{tr}x \) antibodies, were not added to these sequential incubations. The slides were then incubated with avidin-FITC and were visualized under the microscope. If there was an excess of the first primary antibodies after incubation with the biotinylated secondary antibodies, the second secondary antibodies, labeled with Rhodamine would react with the excess anti-\( Pc \) antibodies, giving red signals. No red bands were observed indicating that there was no excess of anti-\( Pc \) antibodies and the yellow bands truly represent co-localization of the \( \text{tr}x \) and \( Pc \) proteins (data not shown). If there was indeed an excess of first primary antibodies in the experimental slides,
all signals would be yellow. Presence of green and red signals along with few yellow bands confirms that the yellow signal is truly due to the co-localization of the two proteins.

Co-immunolocalization of the *trx* and *ash1* proteins on polytene chromosomes

Co-localization of *trx* and *ash1* was done as for *trx* and *Pc*. The first primary antibodies used were *trx* followed by the biotinylated secondary antibodies. Then the chromosomes were incubated with the *ash1* antibodies (A gift from Allen Shearn) followed by the Rhodamine labeled anti-rabbit secondary antibodies. The green signal in this experiment represents the *trx* proteins and red, the *ash1* protein. The chromosomes were analyzed by confocal microscopy.
RESULTS:

**Immunodetection of trx proteins by Western analysis**

Two different polyclonal sera against the *trx* proteins were prepared for this study. To avoid possible crossreactivity with other proteins, the region of *trx* selected for the immunogen does not include any of its highly conserved regions.

Figure 2A shows that this antiserum detects a protein of the expected molecular weight (65 kD) in extracts of bacteria expressing the *trx* fusion protein. Lanes 1 and 2 were incubated with the immune serum and lanes 3 and 4 were incubated with the preimmune serum from the same animal. Lanes 1 and 3 contain extract from bacteria expressing *trpE-trx* fusion protein while lanes 2 and 4 contain extracts from bacteria expressing *trpE*. The 65 kD fusion protein band was detected only by the immune serum. It is absent in lane 3 which was incubated with the preimmune serum, indicating that the common immune serum contains antibodies specifically recognizing the immunogen. Absence of the *trpE* band in lane 2 indicates that the preincubation of the immune serum successfully depleted anti-*trpE* antibodies. Coommassie staining of the same type of gel showed presence of high amounts of *trpE*.

Figure 2B shows that the common antiserum also recognizes two high molecular weight bands in nuclear extracts of
Drosophila embryos. They migrate slower than the highest molecular weight marker used (thyroglobulin large subunit, 330 kD, Sigma) and therefore are in the size range predicted for the \textit{trx} proteins. The exact estimation of the molecular weights of these two bands is difficult since the standard curve is not linear in this size range. N-N' Diallyltartardiamide (DATD) was used as a crosslinker instead of bis-acrylamide (Hames and Rickwood, 1981). DATD gels have larger pore size at the same percentage than the bis-acrylamide gels. This permits the large \textit{trx} protein isoforms to enter the gel.

The total protein extract from \textit{Drosophila} embryos did not give any signal on the Western blots and gave high background staining. This could be because the nuclear \textit{trx} proteins are diluted in the total protein extract and may therefore be under the detection limit for this system.

The \textit{trx} residues 1-172 were used to prepare the isoform specific immunogen. Figure 3A shows that the isoform specific immune serum recognizes a 19 kD fusion protein band (lane 2). This band is absent from lane 4 which is incubated with the preimmune serum. The presence of other bands indicate presence of non-specific anti-bacterial antibodies in the serum. As predicted, the isoform specific antiserum recognizes only one band in the nuclear extract from \textit{Drosophila} embryos which runs above 330 kD (Figure 3B). This antiserum also detects a band with an apparent molecular weight of 300 kD which is
Figure 2: Characterization of the "common" anti-\textit{trx} antibodies by Western analysis.

A) Detection of \textit{trx} fusion protein in bacterial protein extracts. Lanes 1 and 3: extracts containing \textit{trpE-trx} fusion protein (68 kD, predicted); lanes 2 and 4: extracts containing the partial \textit{trpE} protein (36 kD) expressed from pATH vector alone. Lanes 1 and 2 were incubated with immune serum, lanes 3 and 4 with preimmune serum. The expected 68 kD \textit{trpE-trx} fusion protein is detected only on blot incubated with the immune serum. Additional bands in lanes incubated with the preimmune serum (lanes 3 and 4) indicate presence of non-specific anti-bacterial antibodies in this serum.

B) Detection of \textit{trx} proteins in \textit{Drosophila} nuclear extracts. Lane 1 was incubated with immune serum and lane 2 with preimmune serum. Two large proteins are detectable in the size range of the two predicted \textit{trx} proteins (365 and 405 kD), i.e. above the 330 kD thyroglobulin marker.
Figure 3: Characterization of the isoform-specific anti-trx antibodies by Western analysis.

A) Detection of *trx* fusion protein in bacterial protein extracts. Lanes 1 and 3: extracts from bacteria which are not induced to make the *trx* fusion protein. Lanes 2 and 4: extracts containing the induced *trx* fusion protein (predicted molecular weight 19 kD). Lanes 1 and 2 are incubated with the immune serum. Lanes 3 and 4 are incubated with the preimmune serum from the same animal. The 19 kD fusion protein band (indicated by an arrow) is recognized only by the immune serum. Additional bands in lanes 3 and 4 indicates presence of non-specific antibacterial antibodies in this preimmune serum.

B) Detection of *trx* proteins in *Drosophila* nuclear extracts. Lane 1 was incubated with immune serum and lane 2 with preimmune serum from the same rabbit. This antiserum also detects a protein in the size range predicted for the *trx* proteins, above 330 kD. The band running at approximately 300 kD is present only in lane 1 and could be a degradation product of the *trx* proteins. The doublet present in both preimmune and immune lanes below the 180 kD marker is due to antibodies present in the preimmune serum that cross-react with other fly proteins. It also serves to indicate that approximately equal amounts of total protein are present in both lanes.
smaller than the TRXI protein isoform and could be a degradation product.

The independent calibration of the detection limit of the common anti-trx antiserum by Western analysis using a dilution series of \textit{trx-trpE} fusion protein showed that protein quantities as low as 1 pg could be detected. Based on this calibration the proteins detected in the nuclear extracts are present in low abundance as predicted for \textit{trx} based on the results of the Northern analysis (Breen and Harte, 1991). Also both protein isoforms contain PEST sequences which are implicated in the high turnover of proteins (Stassen \textit{et al.}, 1995). Further evidence that these antisera specifically recognize \textit{trx} proteins in \textit{Drosophila} is presented below.

\textbf{Chromosomal binding of the \textit{trx} proteins}

Since the \textit{trx} proteins are detected in the nuclear extracts we wanted to determine if they are associated with specific chromosomal sites. The common antiserum was used to stain polytene chromosomes from third instar larval salivary glands. Figure 4A shows that the antibodies present in this serum recognize proteins associated with specific chromosomal sites. These sites include the BX-C (89E1-2) and ANT-C (84 A1-B2) loci, which are the known genetic targets of \textit{trx} (Figures 4B and C). A similar pattern of sites is seen with the isoform specific
Detection of weaker signals was variable with this antiserum. This variation could be due to technical differences arising during fixation and is discussed in detail in the discussion. Similar preparations incubated with the preimmune sera did not show any signal (data not shown).

To confirm that these sites represent specific binding of \textit{trx} proteins and no others, chromosome binding was analyzed in nuclei from salivary glands of \textit{trx} \textit{l} larvae. The \textit{trx} \textit{l} mutation is a spontaneous homozygous-viable temperature-sensitive mutation (Ingham and Whittle, 1980). Ninety-five percent of \textit{trx} \textit{l} flies raised at the restrictive temperature of 290°C show homeotic transformation phenotypes when the mothers are \textit{trx} \textit{l} (Breen and Harte, 1991; Ingham and Whittle, 1980). The \textit{trx} \textit{l} mutation is associated with an insertion of approximately 9 kb DNA in the first intron of the \textit{trx} gene.

Figure 5 shows a representative chromosome spread from a salivary gland nucleus of a third instar homozygous \textit{trx} \textit{l} larva raised continuously at the restrictive temperature. The number of detectable \textit{trx} binding sites is drastically reduced, some nuclei containing no binding sites at all. The morphology of the \textit{trx} \textit{l} chromosomes is not as good as the wild type chromosomes raised at 290°C. In contrast, chromosomes from wild type larvae cultured continuously at 290°C or from \textit{trx} \textit{l} homozygotes raised at the permissive temperature of 220°C, or from animals with
Figure 4: The \textit{trx} proteins are associated with polytene chromosomes
A) Immunolocalization of the \textit{trx} proteins on salivary gland polytene chromosomes of wild-type third instar larvae.
B) Binding to the BX-C region at 89E1-2.
C) Binding to the ANT-C region at 84A1-2 and 84B1-2.
Figure 5: The *trx*<sup>1</sup> nuclei have greatly reduced number of binding sites. A representative homozygous *trx*<sup>1</sup> nucleus, from an animal raised at 290°C, showing the residual pattern of binding sites. Only a few sites, corresponding to the strongest signals in wild-type chromosomes, remain detectable and residual sites tend to have weaker signals. Some *trx*<sup>1</sup> nuclei show no detectable binding sites.
only one copy of $trx^+ (+/Df(3R)red^A p^52)$ show the wild type binding pattern seen in figure 4A.

To carry out quantitative analysis of the differences in the number of $trx$ binding sites seen in wild type and homozygous $trx^I$ chromosomes, a large number of nuclei were scored for the number of binding sites. The sites were mapped for cytological location wherever possible. Figure 6 graphically summarizes this analysis. The mean number of residual binding sites detectable in $trx^I$ nuclei is 5 compared to 50 sites in wild type nuclei. The number 50 does not reflect a departure from the 63 sites listed in Table I, but is simply the average of scorable sites in each nucleus. The scorable sites depend on the quality of a spread, some sites may be obscured by other chromosome arms and to a lesser extent, some of the weaker sites may not be detected due to the normal variability in signal intensities. No correction was made for these factors since they are expected to affect mutant sample equally and therefore have little effect on the magnitude of the difference observed between the two samples.

In $trx^I$ nuclei the residual sites which were observed most frequently were those exhibiting strong signals in wild-type nuclei. This suggests that $trx^I$ mutation causes the concentration of active $trx$ proteins to be reduced and that the residual sites may have higher affinity for the $trx$ proteins. The BX-C and ANT-C binding sites, which show strong signals, are absent in
Figure 6: Distributions of the number of chromosome binding sites per nucleus in large matched samples of \( \text{trx}^l \) mutant (black) and wild-type (white) nuclei. The distributions are non-overlapping. The \( \text{trx}^l \) nuclei have a greatly reduced number of binding sites per nucleus (mean = 5; mode = 5; median = 5.5) compared with wild-type (mean = 50; mode = 50; median = 50.6). Residual sites tend to have weaker signals. The pattern of binding site loss in \( \text{trx}^l \) nuclei is not random. Weak sites are always lost and some nuclei contain no detectable sites. When residual sites are evident, those with strongest signals in wild type have the highest probability of being retained: 49F, 99B, 98D, 22A, 29F, 7B. Of the \( \text{trx}^l \) nuclei in which ANT-C and BX-C binding sites were scorable, 86% had no detectable signal at the ANT-C and 87% had none at the BX-C.
more than 86-87% of the \( \text{trx}^1 \) nuclei in which these sites were scorably.

The residual sites seen in \( \text{trx}^1 \) nuclei are also detected by the isoform-specific antibodies in wild-type nuclei. Thus detection by two different antibodies confirm that these sites represent binding of the \( \text{trx} \) proteins only and are not "residual" because the serum contains cross-reacting antibodies recognizing other fly proteins. These results clearly establish that the proteins detected at these chromosomal sites are indeed \( \text{trx} \) proteins and that the antibodies present in this serum specifically recognize the \( \text{trx} \) proteins.

Table I summarizes cytological mapping of 63 consistently observable \( \text{trx} \) binding sites in wild-type chromosomes detected with the common region antiserum. They are detectable in euchromatic regions of chromosomes 1, 2 and 3 but not on the tiny fourth chromosome or the chromocenter. The staining intensities observed at individual sites are highly reproducible, but there are considerable differences in the intensities of signals observed at different sites. Ten sites, including the \( \text{ANT-C} \) and \( \text{BX-C} \) always exhibit a strong signal (++++). Twenty one sites show moderate signal (++) and 31 sites consistently show weak signals (+). The region containing the \( \text{en} \) locus (48A3-4), which is regulated by \( \text{trx} \) consistently exhibits a weak to moderate signal (Breen et al., 1995). The variation in signal intensities may reflect the difference in the number of binding sites or in the
Table I: Chromosomal binding sites of the \textit{trx} proteins

<table>
<thead>
<tr>
<th>\textit{trx}</th>
<th>Signal</th>
<th>Pc-G</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chromosome 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2F</td>
<td>++</td>
<td>*</td>
</tr>
<tr>
<td>4C</td>
<td>++</td>
<td>*</td>
</tr>
<tr>
<td>6D</td>
<td>++</td>
<td>*</td>
</tr>
<tr>
<td>7A</td>
<td>+++</td>
<td>*</td>
</tr>
<tr>
<td>7B</td>
<td>+++</td>
<td>*</td>
</tr>
<tr>
<td>8C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11C</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>11F</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>16D</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>17D/E</td>
<td>++</td>
<td>?</td>
</tr>
<tr>
<td><strong>Chromosome 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22A</td>
<td>+++</td>
<td>*</td>
</tr>
<tr>
<td>22B</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>25A</td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>26A</td>
<td>++</td>
<td>*</td>
</tr>
<tr>
<td>26E</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>29F</td>
<td>+++</td>
<td>*</td>
</tr>
<tr>
<td>30C</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>33B</td>
<td>++</td>
<td>*</td>
</tr>
<tr>
<td>34B</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>35B</td>
<td>++</td>
<td>?</td>
</tr>
<tr>
<td>36E</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>37B</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>38B</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>39A/B</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>43E</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>45B/C</td>
<td>++</td>
<td>?</td>
</tr>
<tr>
<td>46C</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>46D</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>48A</td>
<td>++</td>
<td>*</td>
</tr>
<tr>
<td>49F</td>
<td>+++</td>
<td>*</td>
</tr>
<tr>
<td>50D</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>52C</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>53C/D</td>
<td>++</td>
<td>?</td>
</tr>
<tr>
<td>56C</td>
<td>+++</td>
<td>*</td>
</tr>
<tr>
<td>56E/F</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>58E</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

\textit{Pc}, \textit{ph}, \textit{Psc}, \textit{Su(z)2}, \textit{z}
| 61C | + | * | 61C | Pc, ph, Su(z)2 |
| 61F | + | * | 61F | Pc, ph, Psc, Su(z)2 |
| 63E | + |
| 64D | ++ |
| 65E | + | 65E | Su(z)2 |
| 66B | + |
| 66F | + | ? | 66E/F | Pc, ph, Psc |
| 69C/D | ++ | ? | 69C | Pc, ph, Psc, Su(z)2, z |
| 70A | ++ | ? | 70A/B | Pc, ph, Psc, z |
| 70D | ++ | ? | 70D/E | Pc, ph, Psc, Su(z)2, z |
| 84A | +++ | * | 84A/B | Pc, ph, Psc, Su(z)2, z |
| 84B | +++ | * | 84A/B | Pc, ph, Psc, Su(z)2, z |
| 84D | +++ | * | 84D | Pc, ph, Psc, Su(z)2, z |
| 85D | ++ | * | 85D | Su(z)2, z |
| 86C | + | 86C | Pc, ph, Psc, Su(z)2 |
| 88C | ++ |
| 88D | ++ |
| 89B | + | * | 89B | Pc, ph, Psc |
| 89E | +++ | * | 89E | Pc, ph, Psc, Su(z)2, z |
| 91E | + |
| 92B | + |
| 93D | ++ |
| 94F | + |
| 98D | +++ | * | ? | 98C/D | Pc, ph |
| 99B | +++ | * | ? | 99A/D | Pc, ph, Psc, Su(z)2, z |
| 100A | + | * | 100A | Pc, ph, Psc |
| 100F | + | * | 100F | Pc, ph, Psc |

Binding sites are reported to the lettered division. Sites are classified as strong (+++), medium (++), and weak (+) based on the intensity of signals seen in at least 50 nuclei. The sites seen with the isoform specific antibodies are indicated as an asterisk, though exhaustive analysis was not done for this antibody. Sites of likely trx co-localization with previously reported Pc-G sites are indicated in the 'Pc-G' column. A * indicates that the reported sites are identical; a question mark indicates that the reported sites overlap, but differ in the resolution with which they were mapped. This is followed by the corresponding Pc-G sites and the specific proteins which have been shown to bind at that site (Rastelli et al., 1993).
affinity for the \textit{trx} proteins.

As mentioned before, a similar pattern of sites is seen with the isoform-specific antibodies. However the detection of weaker sites is much more variable with this antiserum. Table I lists the strong sites seen with this antiserum, but detailed mapping of binding sites was not done.

The \textit{trx} and \textit{Pc} proteins co-localize at many sites on polytene chromosomes

Even at the relatively low resolution of this type of cytological analysis, it is evident that up to 32 of \textit{trx} binding sites appear to coincide with previously reported binding sites for one or more of the Pc-G and \textit{zeste (z)} proteins (Rastelli \textit{et al.}, 1993). To determine whether these sites truly coincide, dual fluorescence confocal microscopy was used to visualize both \textit{trx} and \textit{Pc} proteins simultaneously on chromosomes using anti-\textit{trx} and anti-\textit{Pc} antibodies. Anti-\textit{Pc} antibodies were a gift from R. Paro (Zink and Paro, 1989). Figure 7C shows that \textit{trx} and \textit{Pc} are indeed co-localized at many sites as indicated here by the presence of yellow bands. The yellow bands result from merging of green and red signals from the different fluorochromes used to detect \textit{Pc} and \textit{trx}, respectively. Sites were not mapped individually, but there are at least 30 overlapping sites, corresponding closely to the number predicted from the
Figure 7: Co-localization of *trx* and *Pc* chromosome binding sites by confocal microscopy: A) Wild type polytene chromosomes stained with anti-*Pc* antibodies and visualized using FITC-conjugated anti-rabbit secondary antibodies (green fluorescence). B) The same polytene chromosome preparation used in A stained with anti-*trx* antibodies and visualized using rhodamine-conjugated anti-rabbit secondary antibodies (red fluorescence). C) Merged red (*trx*) and green (*Pc*) fluorescence images in A and B (to yield yellow). The yellow bands identify sites at which *Pc* and *trx* are co-localized. Green and red bands identify sites at which only one of the proteins is bound.
independent mapping (table I). Figures 7A and 7B show trx and 
Pc binding patterns, respectively, in the same nucleus shown in 
figure 7C.

The *trithorax* and *ashl* proteins co-localize at most of 
their chromosomal binding sites

To determine if *trx* co-localizes with any other trx-G protein 
we decided to carry out co-immunolocalization assay using anti-
trx and anti-ashl antibodies. We chose *ashl* for several reasons. 
The *ashl* protein has been shown to be associated with the 
polytene chromosomes (A. Shearn personal communication). As 
described earlier, *ashl* mutations exhibit synergistic genetic 
interactions with *trx* mutations. They have similar mutant 
phenotypes and belong to the trx-G proteins. The assay was 
carried out exactly as it was done for the co-localization of *trx* 
and *Pc*, except that *trx* antibodies were used first. Figure 8C 
shows a nucleus in which the red (*trx*) and green (*ashl*) signals 
are superimposed. Most of the bands seen are yellow indicating 
that the *trx* and *ashl* proteins co-localize at most of their target 
sites. Figures 8A and B show the same nucleus stained with *trx* 
and *ashl* respectively.
Figure 8: Co-localization of \textit{trx} and \textit{ashl} chromosome binding sites by confocal microscopy: A) Wild type polytene chromosomes stained with anti-\textit{trx} antibodies and visualized using FITC-conjugated anti-rabbit secondary antibodies (green fluorescence). B) The same polytene chromosome preparation used in A stained with anti-\textit{ashl} antibodies and visualized using rhodamine-conjugated anti-rabbit secondary antibodies (red fluorescence). C) Merged red (\textit{ashl}) and green (\textit{trx}) fluorescence images in A and B (to yield yellow). The yellow bands identify sites at which \textit{ashl} and \textit{trx} are co-localized.
The *trx* proteins bind chromosomes

Presence of the *trx* protein at 63 discrete sites on chromosomes strongly suggests that it is either a sequence-specific DNA-binding protein or is associated with at least one such protein. Binding of *trx* to the BX-C and ANT-C loci indicates that *trx* acts directly on its target genes. The relatively small number of *trx* binding sites suggests that it is not a global transcription factor. However the phenotype of *trx* null mutations, which appears to be limited primarily to homeotic transformations, raises questions about the identity of these other putative targets for *trx*.

Some of the *trx* binding sites may correspond to "downstream" target genes of the homeotic proteins themselves. It is also possible that some other target genes have tissue-specific requirements for *trx* which may create subtle tissue-specific phenotypes that have so far gone unnoticed, especially in presence of the profound homeotic transformations seen in the *trx* mutants. It is also possible that some of the *trx* binding sites may not be biologically significant. There is evidence from other polytene chromosome associated proteins that this might be the case for some of their binding sites. For example, the *Sex-lethal* protein is bound to the cytological
location of *Notch* (Samuels *et al.* 1994). There is no evidence that *Notch* is regulated by *Sex-lethal* suggesting that all the binding sites of a given protein on polytene chromosomes may not be biologically significant. In any case, whether these other binding sites correspond to genes which are regulated by *trx* will be conclusively proven when it is demonstrated, genetically and molecularly, that the transcription of these genes is dependent on *trx*.

Some of the candidate target genes, based on the close correspondence of their cytogenetic location with the *trx* binding sites, include *cut* (7B3), *esc* (33B1-2), *en* (46A), *eve* (46C3-11), *Psc*, *Su(z)2* (49E2-F1), *Rpl1* (53C-D), *mle3* (66B), *roe* (84D2), *mor* (89B1-4), *fkh* (98D2-3), *ttl* (100A5-B2) (Lindsley and Zimm). Most of these genes are involved in certain developmental decisions. For example, *ttl* is a gap gene involved in determination of the ends of an embryo, *mle3* is required for dosage compensation in males, *eve* and *en* are segmentation genes. *moira* belongs to the *trx*-G genes and *Psc* and *esc* are the *Pc*-G genes. Some of these genes, such as *cut* and *eve*, encode homeobox containing proteins (Lindsley and Zimm). Thus the candidate target genes constitute a variety of genes performing different functions, mostly involved in developmental pathways.

The 63 sites listed in Table I are the sites consistently seen in many nuclei from different animals. Variation in the staining pattern has been reported for other proteins (Rastelli, 1994). We
do not see much variability in our chromosome preparations stained with the common antibodies, though some of the weaker sites are not stained in some nuclei. This variability could be due to variation in fixing the tissue or age of larvae, as has been observed for some of the Pc-G proteins, whose expression decays towards the end of the larval stage, when salivary glands are harvested (Rastelli et al. 1993). The sites which were seen occasionally are not included in the list and these sites usually exhibited very weak signals. It is possible that some sites are less accessible or inaccessible to the antibodies. Also, not all target loci of a given protein are always recognized by the immunostaining of polytene chromosomes. For example, the white locus, a biological and molecular target of zeste, is not recognized by anti-zeste antibodies on polytene chromosomes (Pirrotta et al., 1988; Rastelli et al., 1993). This could be due to failure of zeste to bind to white in the salivary gland or failure to detect its binding.

Reduction in the number of binding sites in trxI nuclei could be due to reduced synthesis of trx protein or or its increased degradation. This mutation is caused by an insertion of a transposable element in the first intron of the trx gene. This insertion may affect splicing, thus reducing the amount of wild type mRNA and hence protein. Nevertheless it constitutes a good control to prove specificity for trx of antibodies used in this study.
The reproducible detection of \textit{trx} proteins at specific chromosomal sites strongly suggests, but does not prove, that \textit{trx} binds DNA either directly or indirectly. There are several examples of proteins involved in RNA processing associated with specific chromosomal sites by binding to the nascent RNA at these sites. The \textit{Sex lethal} protein binds to discrete sites on polytene chromosomes and this binding is to the newly synthesized RNA present at these loci (Samuels \textit{et al.} 1994). Binding of \textit{trx} at the cytological location of the BX-C supports the prediction that it binds DNA, since the genes of the BX-C are never expressed in salivary glands. The repressed state of the BX-C suggests that nascent RNA will be absent at this site. Furthermore the binding of \textit{trx} to a small nontranscribed fragment of the \textit{Ubx} regulatory region (Chapter III) also suggests that its binding is mediated by specific DNA sequences.

\textbf{The binding of \textit{trx} and the transcriptional status of target genes}

The binding of Pc-G proteins to the chromosomal sites of the BX-C and the ANT-C correlates well with the repressor function of these proteins and the fact that none of the BX-C and ANT-C genes are active in salivary glands (Zink and Paro, 1989). On the other hand binding of \textit{trx} proteins to the BX-C and ANT-C in the salivary gland does not reflect the transcriptional status of these
genes, as none of these genes are active in this tissue. This suggests that the binding of \textit{trx} to chromosomes at its target loci is not sufficient for activation.

The observation that \textit{Pc} protein is extensively bound throughout the entire BX-C in S2 cells except at \textit{Abd-B}, which is the only BX-C gene found to be active in S2 cells, supports the view that binding of Pc-G proteins is an important determinant in repressing transcription (Orlando and Paro, 1993). Thus the active vs. repressed state is correlated with the binding of Pc-G proteins, while the binding of \textit{trx} may be "constitutive" \textit{i.e.} \textit{trx} binds to its target regardless of their transcriptional status. It is possible that each target gene of \textit{trx} requires other tissue-specific activator(s) and thus binding of \textit{trx} alone is not sufficient to activate a particular gene in a given tissue. Mutations in \textit{trx} exhibit tissue-specific effects on several homeotic genes (Breen and Harte, 1993).

There are data suggesting that the other \textit{trx-G} proteins bind constitutively to their targets in the salivary glands although these genes are inactive in this tissue. For example, the GAGA protein, encoded by \textit{Trithorax-like}, binds to the hsp70 locus on polytene chromosomes from larval salivary glands in the absence of a heat response. The heat shock factor (HSF), the known activator of the hsp70 gene, binds only after a heat shock (Tsukiyama \textit{et al.}, 1994 and references therein). The hsp70 promoter is active only after a heat shock, indicating that the
binding of HSF causes the actual activation, whereas binding of the GAGA factor may have prepared the promoter to be accessible for the HSF. The ash1 protein also binds to the BX-C and the ANT-C on polytene chromosomes in salivary glands (A. Shearn, personal communication; Chinwalla and Harte, unpublished). Like trx, ash1 is an activator of homeotic genes of the BX-C and ANT-C which are repressed in the salivary glands.

The Fork head (fkh) gene is located at 98D, a site where trx exhibits a strong (++++) signal. It is apparently not regulated by the Pc-G proteins (Jürgens and Weigel, 1988). The fork head gene is required for normal development of the end of Drosophila embryo. Immunostaining of Drosophila embryos using anti-fkh antibodies demonstrates the presence of fkh protein in various different tissues including the embryonic salivary gland placode. Histological analysis of fkh mutant embryos has revealed that the development of midgut and salivary glands are abnormal in these embryos (Jürgens and Weigel, 1988).

Fork head is expressed in salivary glands of the third instar larvae and its expression is reduced in trx mutants (Kuzin et al., 1994). Fork head, a genetically and molecularly established target of trx, constitutes an example of a target gene of trx which is active in salivary glands.

It is possible that the binding of trx to the ANT-C and the BX-C does not activate genes of these complexes because of the
absence of dedicated activators of these genes in salivary glands. The Pc-G protein bind DNA in absence of the dedicated activators. In any case, binding of \textit{trx} is not sufficient to activate transcription.

A number of \textit{trx} binding sites may correspond to suppressor of variegation loci \textit{e.g.} \textit{Su(var)2-8} (25A1-2), \textit{Su(var)2-11} (33B1-2), \textit{Su(var)3-4} (84D13-E2), \textit{Su(var)339} (85D4-7), \textit{Su(var)330} (88C), \textit{Su(var)3-9} (88D10-E1), \textit{Su(var)3-12} (100E3-5). Most of these suppressor genes are not yet characterized. The Pc-G proteins are not bound at most of these loci, thus indicating that they might be active in this tissue.

\textbf{The \textit{trx} and \textit{Pc} proteins are co-localized at several chromosomal sites}

At the resolution of the chromosome assay \textit{trx} and Pc-G proteins are bound at 32 common sites. This suggests that the binding of Pc-G proteins at a locus does not exclude binding of \textit{trx} at the same locus. As described before, \textit{trx} and \textit{Pc} exhibit antagonistic genetic interactions suggesting competition between the two. The fact that \textit{trx} and \textit{Pc} are co-localized at many sites suggests that any competition suggested by the genetic analysis, is not at the level of chromosomal binding but may be at some other level. The correlation between binding of the Pc-G proteins and the transcriptionally repressed state of a gene (Orlando and
Paro, 1993) suggests that binding sites of \textit{trx} where Pc-G are not bound may represent genes which are active in salivary glands. This also suggests that the co-localization pattern could be different in other tissues where different sets of genes will be active or repressed in those tissues.

The \textit{trx} and \textit{ashl} proteins are co-localized at many sites on the polytene chromosomes.

The \textit{trx} proteins are also co-localized with \textit{ashl} another member of the \textit{trx}-G. The number of sites at which these two proteins are co-localized is greater than that seen for the \textit{Pc/trx} co-localized sites. Virtually complete co-localization of the \textit{trx} and \textit{ashl} proteins suggests that they might be part of a multiprotein complex.

The resolution of the polytene chromosome assay used here does not allow us to rigorously conclude if \textit{trx} and \textit{ashl} are binding to the same locus or to adjacent loci. Further analysis is required to determine this. Such detailed analysis is described in Chapter III for the Pc-G proteins and indicates that the \textit{trx} and the Pc-G proteins are co-localized at the \textit{Ubx} regulatory DNA. This suggests that they may be binding to the same genes at the other sites where they are co-localized.
The role of *trx* in homeotic gene activation

The role of *trx* as a chromosomal protein possibly acting by modulating chromatin structure is supported by *trx* binding to discrete sites on polytene chromosomes. The presence of two novel motifs in *trx* which are shared by other proteins, suggested to modulate chromatin structure, such as *E(z)* and *Pcl*, the *Pc-G* proteins and *Su(var)*3-9 suggests that *trx* may act by altering local chromatin structure (Jones and Gelbart, 1993; Tschiersch et al., 1994). *Su(var)*3-9 is a modifier of PEV and also contains the "chromo domain" found in *Pc*. Thus it shares homology with proteins which have antagonistic actions. If this sequence homology represents functional similarity, this may mean that the mechanism of action of all three classes of proteins, the modifiers of PEV, the positive and the negative regulators of homeotic genes, share some common features.

Certain alleles of *Trithorax-like* (*Trl*), another member of the *trx-G*, act as enhancers of PEV (Farkas et al., 1994). Enhancers of PEV are implicated in chromatin decondensation (Hartmann-Goldstein, 1967). *Trl* encodes the Drosophila GAGA factor (Farkas et al., 1994). The GAGA factor has been shown to create a DNase I hypersensitive site at the hsp70 promoter which is essential for subsequent binding of HSF upon heat induction (Tsukiyama et al., 1994). This suggests that GAGA may form an
"open" chromatin domain that facilitates binding of other proteins.

Mutations in other well characterized enhancer of PEV, $E(var)3-93D$, also shows homeotic transformations, suggesting a role in activation of the BX-C and the ANT-C genes (Dorn et al., 1993). It encodes a protein which binds to polytene chromosomes at many sites including the BX-C and the ANT-C. $E(var)3-93D$ contains a POZ domain shown to be involved in protein-protein interactions suggesting that it may be part of a multiprotein complex. Its role in the activation of homeotic genes may be indirect as a result of its involvement in chromatin decondensation. It shows genetic interactions with $trx$ suggesting a possible role of $trx$ in determination of the local chromatin structure.

**Does $trx$ function by changing local chromatin structure?**

The fundamental structural unit of chromatin is the core nucleosome particle, which consists of an octameric assembly of the four core histones, H3, H4, H2A and H2B complexed with 145 bp of DNA. In the presence of a fifth histone, H1, this assembly can condense to form a higher order structure in which each nucleosome is now associated with 168 bp or two full superhelical turns of DNA. Chromain is not a static, stable
structure. Proteins, including histones, continually equilibrate in and out of it. The nucleosomal arrays are further packaged to form irregular rod-like structures with a diameter of about 30 nm. These 30 nm fibers are then compacted into higher order structures, such as irregular loop-like structures, to form chromosomes (Wolffe, 1993; Travers, 1993).

This condensation of DNA into chromosomes creates a problem for access of sequence-specific DNA-binding proteins to bind their target sequences. These structural constraints also necessitates different levels of transcriptional regulation. Proteins involved in gene regulation can act at any level of the chromatin assembly to control gene expression.

Experimental evidence increasingly suggests that the GAGA factor encoded by *Trl*, a trx-G gene, acts at the level of chromatin to disrupt nucleosome assemblies (Croston et. al. 1991; Lu et. al. 1993; Tsukiyama et. al. 1994). Deletion of a GAGA binding site in the upstream region of *hsp26* abolishes its heat response upon heat induction (Lu et. al. 1993). The binding of GAGA creates a DNase I hypersensitive site and also increases accessibility of a restriction enzyme site present in this region. The inductive response of *hsp26* is also dependent on binding of HSF to its recognition sequence, HRE. Mutations in HRE prevent HSF binding and cause a dramatic reduction in the heat shock-induced transcription of *hsp26*. But these mutations do not cause any change in the DNase I hypersensitive sites or accessibility of
a restriction site, suggesting that binding of HSF does not cause change in nucleosome positioning (Lu et al. 1993). These results suggest that the GAGA factor and HSF have distinct roles in gene activation: the GAGA factor prepares or "opens" chromatin for binding of HSF and the HSF binding mediates the actual heat shock response.

In vitro studies done with GAGA factor activity at the hsp70 promoter also suggest its role in nucleosome repositioning. The DNase I digestion pattern at the hsp70 promoter changes after addition of purified GAGA protein at any time during or after nucleosomal assembly on naked DNA (Tsukiyama et al. 1994).

How may GAGA mediate its function in chromatin? The structure of chromatin is dynamic, with histones constantly coming on and off DNA and being in equilibrium. GAGA may access its sites during such event while competing with histones. Once bound it is stabilized by protein-protein interactions through its POZ domain. GAGA also acts as an antirepressor of H1 mediated repression, supporting the competition with histones during dynamic assembly of nucleosomes.

The GAGA-mediated nucleosome disruption requires ATP. GAGA does not contain any known motif involved in ATP hydrolysis and the ATP hydrolysis function may be provided by complexing with another trx-G protein, brahma. The brahma protein is homologous to ATP-dependent helicases and may function in providing the required energy to GAGA by
complexing with it (Tamkun et al., 1992). The human homologue of *brhama*, hSWI2, forms a complex with other hSWI proteins. This complex has been shown to disrupt nucleosomes and the presence of hSWI2 is essential for this function (Imbalzano et al., 1994; Kwon et al., 1994). The hSWI2 protein is present in a large multiprotein complex of $2 \times 10^6$ kb and *brahma* has recently been found to be in a similar large complex, further suggesting that it has a similar role to hSWI2 (Dingwall et al. 1995). Thus, by analogy *brhama* may also work in a similar fashion.

How these proteins like GAGA and SWI act to disrupt nucleosomes is not yet fully understood. But their action is required for binding of other factors to their respective sites and activation of a promoter.

The binding of the *trx* proteins to specific sites on polytene chromosomes, the presence of novel motifs in *trx* which are shared by other chromatin modifying proteins such as *E(z)*, *Pcl* and *Su(var)3-9* and the possibility that the other *trx*-G proteins act by changing local chromatin structure strongly suggests that *trx* may also act by disrupting nucleosomes, similar to *Trl*. It is also possible that *trx* may act at higher levels of chromatin packaging which causes decondensation and thus helps in binding of other factors. The *trx* proteins may act by complexing with other proteins such as *brhama* which may help in forming chromatin domains accessible for binding of other transcription
factors. It may facilitate binding of tissue-specific activators, thus exhibiting tissue-specific effects of homeotic gene expression (Breen and Harte, 1993). If so, it might provide a basis for an explanation for the proposed "constitutive" binding of *trx* at sites which also bind Pc-G proteins. Binding of *trx* to a target gene may be of no consequence unless that gene is destined to be transcriptionally activated. In those cells in which the target gene is not active, Pc-G binding would establish a stable repression domain and override or prohibit any subsequent effects of the *trx* proteins which remain bound.
Chapter III: In vivo mapping of trx binding sites to smaller regions of Ubx UCR.

INTRODUCTION:

The spatially restricted expression domains of the homeotic genes ensure correct development of an embryo into an adult. The expression patterns of the homeotic genes are determined by the products of the transiently acting segmentation genes, which disappear shortly after gastrulation (Ingham and Martinez-Arias, 1986; Irish et al., 1989; Tremml and Bienz, 1989). The homeotic genes are expressed and required throughout development to ensure stable maintenance of cell fates (Diederich et al., 1989; Morata and García-Bellido, 1976).

The trx-G and the Pc-G proteins maintain the established expression domains of the homeotic genes throughout development. The trx-G genes maintain expression of the homeotic genes within their normal expression domains (Capdevila and García-Bellido, 1981; Shearn, 1989; Shearn et al., 1987). The Pc-G genes repress each homeotic gene outside of its normal expression domain (Ingham, 1984; Struhl, 1981). The establishment and maintenance of the restricted expression patterns of the homeotic genes is important for normal development of the embryo.
The homeotic gene *Ultrabithorax* (*Ubx*) is required for correct development of the third thoracic and the first abdominal segments (Lewis, 1978). It is a member of the BX-C, which is located at 89E, on the right arm of the third chromosome. The *Ubx* gene is expressed in parasegments 5-13 of embryo, with maximum expression in parasegment 6 (Beachy *et al*., 1985). In *trx* embryos, *Ubx* expression is dramatically reduced in all tissues throughout its normal domain except in the visceral mesoderm (Breen and Harte, 1991; Mazo *et al*., 1990). The reduction in *Ubx* expression is greatest in parasegment 6, where it is normally expressed at its highest level.

We examined whether the *trx* proteins bind to the *Ubx* gene. The regulatory region of *Ubx* has been studied extensively. Transformant flies carrying P element constructs containing various *Ubx* regulatory DNA fragments are available (Chan *et al*., 1994; Irvine *et al*., 1991; Simon *et al*., 1993; Simon *et al*., 1990). Some of the P element constructs contain a *lacZ* reporter gene, under the regulation of a *Ubx* minimal promoter. Various DNA fragments from the *Ubx* regulatory regions have been cloned upstream of the promoter to identify elements required for the restricted expression pattern of *Ubx* (Chan *et al*., 1994; Irvine *et al*., 1991; Simon *et al*., 1993; Simon *et al*., 1990).

The polytene chromosomes from the salivary glands of transformant larvae carrying P element constructs containing *Ubx* regulatory regions were stained with anti-*trx* antibodies.
They were screened for the presence of ectopic *trx* signal at the site of integration of the transposon. This provides an *in vivo* test to determine whether *trx* binds to the regulatory DNA of its target genes. Since it is an *in vivo* test, it does not require that the *trx* proteins bind DNA directly. Any other factors which could be required by *trx* to bind DNA are presumably present in the salivary gland tissue. By testing different DNA fragments from the *Ubx* gene, we have mapped the *trx* binding sites to a small region.

The P element mediated transformation

P element mediated transformation has become one of the key methods of *Drosophila* molecular genetics. The P element is a naturally occurring transposable element in *Drosophila melanogaster*. It contains 31 base pair terminal repeats which enables the P element to integrate stably into the *Drosophila* genome. It contains a gene encoding for the enzyme transposase, which is required for its movement from one integration site to another. This movement of the transposon occurs only in the germ cells. These properties of the P element made it suitable for its use as a vector carrying foreign DNA into the germ line (Rubin and Spradling, 1982). The vector contains P element ends required for its integration into chromosomes, restriction enzyme sites for cloning the DNA of interest and a marker gene.
which helps in identifying transformants, but is deficient in transposase. The vector DNA is injected along with the helper DNA coding for transposase into the posterior end of young embryos before formation of the pole cells. This makes it highly probable that the pole cells, the future germ cells, will pick up the transposon and will integrate it into its chromosomes.

All surviving adults after the injection (G0) are individually back-crossed and then checked for the expression of the marker gene. The embryos selected for injection are mutant for a marker gene e.g. white. The wild type white gene encodes red pigment of adult eyes. The G1 generation is therefore screened for red eyed flies which are due to the wild type expression of the white gene present in the transposon.

This is a very useful technique and has been used to answer a variety of questions. One of them is discussed here, i.e. to use this as a promoter-reporter construct to identify the enhancers or other elements required for the promoter function. The lacZ gene serves as a reporter gene which follows the expression pattern determined by the attached promoter and its enhancer elements. The transformant embryos are then tested for the expression pattern of lacZ which is compared with the native gene expression pattern of the promoter being tested.
The enhancer elements of *Ubx*

The *Ubx* gene spans 120 kb of genomic DNA with an unusually large upstream regulatory region and intronic DNA (Hogness *et al.*, 1985; Kornfeld *et al.*, 1989; O'Connor *et al.*, 1988). Mutations in *Ubx* result in transformations of parasegments 5 and 6 into parasegment 4 (Beachy *et al.*, 1985; Lewis, 1978; White and Wilcox, 1985). A segment is made up of an anterior and a posterior compartment which have separate cell lineages (Lawrence, 1981). A parasegment is an overlapping metamere consisting of a posterior compartment of one segment and the anterior compartment of the next segment (Figure 9; Martinez-Arias and Lawrence, 1988). Parasegments are of fundamental importance in *Drosophila* embryogenesis as they are the first metameric units to become visible in embryogenesis (Ingham *et al.*, 1985)

Immunostaining of the *Drosophila* embryos using anti-*Ubx* antibodies demonstrated that it is expressed in parasegments 5-13 with a maximum level of expression in parasegment 6 (Beachy *et al.*, 1985). Consistent with its expression pattern, mutations in *Ubx* cause homeotic transformations of parasegments 5 and 6. The *Ubx* and *bxd* mutations cause subtle transformations of the larval cuticle in the anterior portions of abdominal segments 2-7 where *Ubx* is expressed weakly (Beachy *et al.*, 1985). Conversely ectopic expression of *Ubx* in the
Figure 9: Skematic representation of the body segments, compartments and parasegments: Ultrabithorax is expressed in PS 5-13. The Ubx subfunction mutations specifying its expression in subdomains is also indicated.
anterior parasegments causes homeotic transformations of the structures of the head, first and second thoracic segments into structures of the first abdominal segment (González-Reyes and Morata, 1990). This indicates that the spatially restricted expression pattern of *Ubx* is important for normal development.

Genetic analysis of the *Ubx* gene identified *anterobithorax* (*abx*), *bithorax* (*bx*), *bithoraxoid* (*bxd*) and *postbithorax* (*pbx*) mutations which cause a subset of *Ubx* mutant phenotypes (Duncan, 1987; Lewis, 1978). These mutations identify *Ubx* regulatory regions which are required for expression in different subsets of the *Ubx* expression domain. The *abx* and *bx* mutations primarily affect the development of parasegment 5 and map within the large third intron (Bender *et al.*, 1983; Peifer and Bender, 1986). The *pbx* and *bxd* mutations affect the development of parasegment 6 and map upstream of the transcription start site (Bender *et al.*, 1983; Bender *et al.*, 1985; Lipshitz *et al.*, 1987). The region defined by the *abx* and *bx* mutations has been called the downstream control region (DCR) and that defined by the *pbx* and *bxd* mutations is termed the upstream control region (UCR) (Irvine *et al.*, 1991).

Many of the *bxd* mutations are caused by chromosome breakpoints within the UCR and are of particular interest. They form an allelic series in which breakpoints closer to the *Ubx* transcription start site cause strong transformations while breakpoints further away cause weaker transformations.
Breakpoints that cause bxd phenotypes are located from 2 kb to approximately 40 kb upstream of the transcription start site. This, together with the alterations in the Ubx protein expression in bxd mutants and the genetic requirement for bxd in cis to Ubx, established that the UCR is a transcriptional regulatory region of extraordinary length (Beachy et al., 1985; Lewis, 1955).

Analysis of the Ubx regulatory region

Presence of the extraordinarily large regulatory region in Ubx has been suggested to imply complex regulation of this gene by many factors (Duncan, 1987; Hogness et al., 1985). A detailed analysis of the regulatory region of the Ubx gene has revealed the presence of multiple enhancers. These enhancer elements when joined to a minimal Ubx promoter, can drive the expression of a lac Z reporter gene in a pattern which is a subset of the complete Ubx expression pattern (Muller and Bienz, 1991; Qian et al., 1991; Simon et al., 1990). Three of these enhancers ABX, PBX and BRE map to the genetically defined abx, pbx and bx regions respectively. These enhancers have been shown to contain multiple binding sites for transiently expressed activators and repressors encoded by the segmentation genes: hunchback (hb), tailess (tll), krüppel (kr), fushi-tarazu (ftz) and knirps (kni) (Qian et al., 1993; Zhang et al., 1991).
The segmentation genes encode transcription factors which bind to the enhancers in a sequence-specific manner and control the transcriptional status of the homeotic genes. *Hunchback* and *tll* repress *Ubx* expression outside of its normal expression domain. The *hunchback* protein establishes the anterior boundary of *Ubx* expression and *tailless* sets its posterior boundary (Reinitz and Levine, 1990; White and Lehmann, 1986). *fushi tarazu*, *even skipped* and *Krüpple* positively regulate *Ubx* expression in its specified expression domain (Ingham *et al.*, 1986; Ingham and Martinez-Arias, 1986; Tremml and Bienz, 1989).

The segmentation gene products are active only transiently and fade away after gastrulation. The *lacZ* pattern driven by the different *Ubx* enhancers to which the segmentation gene products bind changes at this point (Irvine *et al.*, 1991; Qian *et al.*, 1991; Simon *et al.*, 1990).

Expression of *lacZ* expression is derepressed in cells anterior to parasegment 5 indicating that these enhancers lack sequences needed for the action of the negative regulators such as the Pc-G proteins. This ectopic expression of *lacZ* in parasegments anterior to parasegment 5 is indicative of the loss of maintenance of the expression boundaries by the Pc-G proteins. Ultimately the expression of *lacZ* decays prematurely indicating that these enhancers do not contain sequences required for the action of the trx-G proteins.
To identify the DNA sequences which confer the long term maintenance of the *Ubx* expression pattern, further dissection of the large *Ubx* regulatory regions was carried out using the same *lacZ* expression assay.

**Localization of the Polycomb Response Element (PRE)**

Transformants carrying a P element transposon containing a 35 kb UCR DNA attached to the *Ubx* minimal promoter (35UZ) showed the same expression pattern as that of wild type *Ubx* (Irvine *et al.*, 1991; Figure 10). The 35UZ expression domain is the same as that affected in the *pbx* mutations *i.e.* strongest in PS 6 and the posterior compartments of PS 7-13. This construct contains UCR DNA from map coordinates -31 to +4 (map coordinates as described in Bender *et al.*, 1983). Expression of *lacZ* remained stable throughout embryogenesis and in the imaginal discs. The imaginal discs of larva contain the adult progenitor cells. This indicated that the 35 kb UCR DNA contains the Pc-G and the trx-G response elements along with the DNA sequences required for the action of the segmentation genes for early determination of the expression pattern.
Figure 10: A) The Ubx mRNA: The Ubx mRNA with the map coordinates is shown here. The Ubx upstream regulatory region is large and spreads till the map coordinate +10. The PBX enhancer lies in this upstream regulatory region. The third large intron contains the downstream regulatory region. The enhancers ABX and BX map in this intronic region.

B) The UCR DNAs used in transformation constructs: We tested 35UZ, 22UZ, 14.5SH, 2.2KE and 0.67PN constructs for presence of trx binding sites by immunostaining the polytene chromosomes from the transformants. (+) or (−) indicates presence or absence of the trx binding sites.

SH - Sall-HindIII, H - HindIII, KE - KpnI-EcoRI, PN - PstI-NdeI, StP - Styl-PstI
Another construct 22UZ, containing UCR DNA from -31 to -9 initially shows the wild type expression pattern (Irvine et al., 1991; Figure 10). This expression pattern was not maintained after gastrulation indicating that it is missing at least some sequences required for the action of the trx-G and Pc-G proteins. This suggested that sequences required for the action of the trx-G and Pc-G proteins are located in the 13 kb DNA fragment from -9 to +4 which is deleted in the 22UZ but is present in 35UZ.

Transformants carrying a transposon containing a 14 kb UCR DNA from -18.5 to -4 showed lacZ expression in parasegments 6-13 (Simon et al., 1993; Figure 10). This 14 kb DNA fragment overlaps with the 22 kb fragment of 22UZ from map coordinates -18.5 to -9. This expression is stronger in the posterior compartments of each segment as is true for the pbx expression patterns from the wild type Ubx gene. The anterior boundary in parasegment 6 is maintained throughout embryogenesis and the lacZ expression does not decay prematurely. This indicates that this 14 kb fragment contains sequences necessary for the stable maintenance of long term expression of Ubx by the Pc-G and the trx-G proteins. Maintenance of Ubx expression by the 14 kb DNA fragment but not by 22UZ indicates that the sequences necessary for the action of the trx-G and the Pc-G proteins lie within map coordinates -9 to -4, thus limiting them to approximately 5 kb of UCR DNA.
Immunostaining of polytene chromosomes from transformants carrying the 14 kb fragment showed the presence of new binding sites for \textit{ph} and \textit{Pc} at the site of integration of the transposon (DeCamillis \textit{et al.}, 1992; Ann Chiang personal communication). This suggests that the 14 kb UCR DNA contains sequences required for the action of the Pc-G proteins and that they act directly on their target genes by interacting with the regulatory sequences. The maintenance of the anterior boundary during late embryogenesis was dependent on the presence of the Pc-G proteins (Simon \textit{et al.}, 1993). Thus the 14 kb DNA was proposed to contain the Polycomb Response Element (PRE) (Simon \textit{et al.}, 1993).

\textbf{Narrowing down the region containing the PRE}

A 6.5 kb HindIII fragment encompassing DNA from map coordinates -10 to -4.5 showed maintenance of the \textit{lacZ} expression pattern (Chan \textit{et al.}, 1994). This 6.5 kb fragment contains the 5 kb DNA fragment predicted to contain sequences necessary for the action of the trx-G and the Pc-G proteins based on the 35UZ, 22UZ and 14 kb expression data described earlier.

Several different lines carrying the 6.5 kb HindIII \textit{Ubx-lacZ} transposon showed various degrees of maintenance. This fragment when combined with the PBX enhancer showed more complete maintenance of \textit{lacZ} expression throughout
embryogenesis and in the imaginal discs. The PBX enhancer alone does not maintain the lacZ expression later in the embryogenesis. This implies that the 6.5 kb HindIII fragment contains sequences for the action of the Pc-G and the trx-G proteins.

To test if the maintenance properties were dependent on the Pc-G proteins, this transposon from different lines was introduced into the Pc$^3$ mutant background (Chan et al., 1994). These embryos showed either partial or complete loss of maintenance at late embryonic stages. Embryos from parents carrying the transposon and homozygous for a temperature-sensitive mutation in E(z), a Pc-G gene, kept at non-permissive temperature, also showed loss of maintenance. These results indicate that long term maintenance observed with the 6.5 kb HindIII fragment is dependent on the Pc-G proteins and contains the PRE.

To determine whether the maintenance of lacZ expression by this construct requires trx, trx$^l$, a temperature-sensitive mutation, was used (Chan et al., 1994). Larvae homozygous for trx$^l$ and the 6.5H + PBX transposon showed loss of expression in imaginal discs. Even ectopic expression in the wing and eye imaginal discs, in the lines that showed it, was strongly decreased by the trx$^l$ mutation. In the embryos, the lack of positive maintenance is partially obscured by the long half-life of the β-galactosidase. These results suggest that this construct
also contains the positive maintenance element which responds to \textit{trx}.

\textbf{Maintenance correlates with the variegated expression of a \textit{white} marker gene}

The flies carrying these constructs showed an interesting additional phenotype. The transformant flies should have red eyes due to the presence of the \textit{white} marker gene on the transposon. These transformant flies showed variegated eyes with patches of white cells interspersed between red cells, reminiscent of position effect variegation (PEV) phenotype (Chan \textit{et al}., 1994). The white patches of cells are due to the lack of expression from the \textit{white} gene present on the vector used for transformation. Generally the variegation is seen when the \textit{white} gene responsible for the red eye pigment is placed near a heterochromatic region due to a chromosomal rearrangement. In some cells the heterochromatin spreads into euchromatic region shutting off the \textit{white} gene expression and thus giving rise to the white cell clones (Spofford, 1976).

Flies carrying only the P element vector without the 6.5 kb HindIII fragment do not show the eye variegation phenotype. Presence of the variegated eyes in the transformant flies indicate that the \textit{white} gene in the vector is variably repressed due to the presence of the 6.5 kb HindIII fragment. There is an
excellent correlation between variegation and the ability of the transposon to maintain anterior repression of lacZ. Lines that variegate strongly also maintain lacZ expression strongly while lines that variegate weakly maintain it partially (Chan et al., 1994).

This variegation is dependent on the Pc-G proteins (Chan et al., 1994). Flies homozygous for $E(z)^{S2}$ mutation which are raised at the non-permissive temperature reach the pharate adult stage and die before eclosing. Eye variegation is abolished in these flies indicating that it is dependent on the Pc-G proteins. The variegation might be due to the repressive action of the Pc-G proteins on the Ubx UCR DNA which "spreads" over to the mini-white gene present in the vector. It also indicates that the repression by the Pc-G proteins can act at a distance, in this case at least through the length of the 4.5 kb lacZ gene. This variegation is not due to presence of the PBX enhancer because transposons carrying only the PBX enhancer did not show the variegation indicating that it requires the presence of PRE. Analysis of the entire upstream control region showed the presence of only one PRE which is located within the 6.5 kb HindIII fragment.
The PRE is contained in a 670 bp fragment

Expression of lacZ requires the presence of *Ubx* enhancers necessary for the early establishment of the *Ubx* expression pattern. This affected the ability to isolate smaller fragments carrying PREs by using the lacZ expression assay unless the PRE itself behaves like an enhancer. However the correlation between the white variegation phenotype and the maintenance of lacZ expression made it easier to identify smaller fragments containing the PRE. Variegation is strictly a function of presence of PRE (Chan *et al.*, 1994). Further analysis using transposons containing smaller DNA fragments from the 6.5 kb HindIII region identified a 2.2 kb KpnI-EcoRI fragment containing the PRE (Rastelli, 1994). Transformant flies carrying the transposon containing 2.2 kb KpnI-EcoRI fragment showed strong variegation. This variegation was dependent on the Pc-G function as it decreased in the presence of *E(z)* mutation. The transposon insertion creates a new binding site for the *Psc* protein in polytene chromosomes of the transformant larvae. This demonstrates that the Pc-G proteins act directly on the PRE.

The PRE function was further narrowed down to a small 1.6 kb StyI-EcoRI fragment included in the 2.2 kb KpnI-EcoRI fragment (Rastelli, 1994). The 1.6 kb StyI-EcoRI fragment located within the 2.2 kb KpnI-EcoRI, showed the presence of PRE as determined by several criteria mentioned above. This
DNA fragment was then divided into three fragments: 500 bp StyI-PstI, 670 bp PstI-NdeI and 400 bp NdeI-EcoRI. The 670 bp PstI-NdeI showed the maximum response to the Pc-G and was concluded to contain the core PRE. It contains a site for the binding of the Psc protein in polytene chromosomes. All these fragments do not show any recognizable expression pattern of lacZ indicating that they do not contain any enhancer elements.

We tested whether these transformants create new binding sites for the trx proteins in the polytene chromosomes. We carried out immunostaining of the polytene chromosomes from various transformant larvae using the anti-trx "common" antibodies. This will allow us to map the trithorax binding sites to the smallest available DNA fragment from the UCR and to prove that the effect oftrx on its target genes is direct as in the case of the Pc-G proteins.
MATERIALS AND METHODS:

Immunostaining of the Polytene chromosomes

Immunostaining of the polytene chromosomes was done as described in chapter II. The anti-\textit{Psc} antibodies were used at 1:10 dilution and were a gift from Dr. Paul Adler (Martin and Adler, 1993).

The \textit{in situ} hybridization of polytene chromosomes

A previously described protocol for \textit{in situ} hybridization was followed with few modifications (Ashburner, 1989). The \textbf{Preparation of chromosome squashes:} Salivary glands of third instar larvae were dissected in 45\% acetic acid and were then transferred to a drop of 123 fix on a siliconized coverslip (123 fix: 1:2:3 parts of lactic acid:distilled water:acetic acid). The glands were fixed for 4 min. and the coverslip was picked with a poly-L-lysine coated slide. Cells were broken by tapping the coverslip with a pencil tip. The slide was then inverted and the coverslip was pressed on a tissue paper. The chromosomes were flattened and spread by pressing down with a thumb and moving in a circular motion. The slide was warmed on a slide warmer at 45^\circ\text{C} for 5 min. and was placed on dry ice for 5 min. with coverslip down. Coverslips were then removed with a razor
blade and slides were immediately immersed into 95% ethanol for 30 min. to several hours. Slides were air dried and processed further.

**Preparing chromosomes for hybridization:** Slides were heated for 30 min. at 65°C in 2X SSC. They were placed sequentially in two 70% ethanol tanks and then one 95% ethanol tank for 5 min. each at 65°C. Slides were then air dried and denatured in freshly prepared 70 mM NaOH for 3 min. After denaturation they were washed in 2X SSC for 5 min. and then dehydrated in ethanol as before at room temperature. Slides were allowed to air dry.

**Labeling reaction using biotinylated dUTP:** The DNA used to make probe was isolated from pCaSpeR-AUGβgal vector as a 4.8 kb EcoRI-HindIII fragment containing the mini-white gene (Ashberner, 1989). The following nick translation protocol was used to prepare the probe. 100 µl of a nick translation reaction contained 2 µg of DNA, 10 µl of 10X NT buffer (0.5 M Tris-HCl pH 7.8-8.0, 50 mM MgCl₂, 0.5 mg/ml BSA), 10 µl of dGCA (0.5 mM each dATP, dCTP, dGTP), 5 µl of Biotin-16-dUTP (Boeringer Manheim), 10 µl of 0.1 M β mercaptoethanol and water up to 100 µl. 1 µl of DNase I (Sigma, 1 µg/µl) was added to 1 ml of pre-cooled double distilled water. 6 µl of this diluted DNase I was mixed with 1 µl of Klenow fragment (NEB) and was used in the nick translation reaction. The microfuge tube containing the reaction mix was immediately placed in 15°C water bath for 2
hr. The tube was placed on ice and 5 μl of the reaction was analyzed on a 2% agarose gel for completion of the reaction. The majority of DNA fragments should be less than 500 bp. The reaction was stopped by adding 2.5 μl of 0.5 M EDTA and 1 μl of 10% SDS and incubating tube at 65°C for 15 min.

The nick translated DNA (100 μl reaction) was ethanol precipitated by adding 160 μg of salmon sperm DNA, 30 μl of 3 M NaOAc (10 μl/100 μl solution) water to 300 μl and 750 μl (2.5X) of ethanol stored at -20°C. The tube was incubated at -80°C for 1 hr and then spun at 12k for 30 min. to pellet DNA. The supernatant was poured off and the pellet was air dried. The pellet was resuspended in 100 μl of hybridization buffer (50 mM Sodium phosphate pH 6.8, 600 mM NaCl, 5 mM MgCl₂ and 1X Denhardt's solution) and stored at -20°C.

Hybridization: The biotinylated probe was denatured by boiling at 75°C for 5 min. and chilled on ice. 20 μl of the hybridization mixture (denatured probe) was added per slide and covered with a siliconized coverslip. The coverslips were sealed with rubber cement and the slides were stored in a humid chamber at 55°C overnight. The next day, the rubber cement was peeled off and the slides were washed three times, 20 min. each, in 2X SSC at 55°C.

Peroxidase staining: The slides were washed twice for 5 min., each in PBS at room temperature, followed by a 2 min. wash in PBS-TX (PBS + 0.1% triton X-100). The slides were then rinsed in
PBS and incubated with 40 μl of ABC solution (Vectastain) for 30 min. at room temperature. The slides were washed twice at the end of incubation in PBS for 5 min. each and in PBS-TX for 2 min. They were rinsed in PBS and then incubated with the coloring solution for 10 min. (0.5 mg/ml DAB in PBS, 1/100 volume of 1% H2O2). The slides were washed with distilled water and then rinsed in PBS. The chromosomes were counterstained with 1:50 Giemsa (Polyscience) for 5 min., washed in running distilled water, air-dried and mounted with permount. The slides were examined under bright field using a Zeiss microscope.
RESULTS:

The *trx* proteins bind to the upstream control region of *Ubx*.

The *Ubx* proteins show a restricted expression pattern. They are expressed in PS5 to PS13 and are required for the normal development of the third thoracic and first abdominal segments. The *Ubx* gene has a large regulatory region containing multiple independent enhancer elements which together determine its complete expression pattern. Binding sites for various segmentation gene products have been identified in each of these enhancers. They are involved in the early determination of the expression pattern of *Ubx*. After the products of segmentation genes disappear, the *trx-G* and *Pc-G* proteins maintain the *Ubx* expression patterns established by these transient factors.

Transformant embryos carrying the 35UZ transposon shows a *lacZ* expression pattern similar to the endogenous *Ubx* pattern. This *lacZ* expression is maintained during late embryogenesis and in the imaginal discs (Irvine et al., 1991). The long-term maintenance of the expression pattern indicates that the 35 kb DNA included in the transposon contains sequences required for the action of the *trx-G* and the *Pc-G* proteins. We examined the
35UZ transformant for presence of a new *trx* binding site at the site of insertion of the transposon in polytene chromosomes.

Immunostaining of polytene chromosomes from larvae carrying the 35UZ transposon with anti-*trx* antibodies showed the presence of a new signal at 34E, the site of integration of the transposon (Figure 11B). The wild type chromosome does not show a signal at this site (Figure 11C). The site of integration was confirmed by performing *in situ* hybridization using biotinylated DNA probe from the *white* marker gene present in the transformation vector (Figure 11A). Presence of signal at the site of integration of the 35UZ transposon indicates that the 35 kb UCR DNA contains a *trx* binding site(s).

The presence of a *trx*-binding site in 35UZ suggests that *trx* maintains expression of the *Ubx* gene by direct interaction with its regulatory sequences. The *trx* proteins may be binding directly to *Ubx* DNA or may be complexing with some other sequence-specific DNA-binding protein. It is possible that *trx* may act similarly with its other target genes.

Another transposon, 22UZ, does not create a new binding site for the *trx* proteins (Figure 12A). *In situ* hybridization using a biotinylated *white* DNA probe shows the presence of transposon at 73F (Figure 12B). This construct is missing the 13 kb promoter-distal DNA which is present in the 35UZ. The absence of *trx* binding at the site of the 22UZ insertion indicates that the promoter-proximal 22 kb DNA of *Ubx* does not contain a *trx*
Figure 11: Immunostaining of polytene chromosomes from the 35UZ transformant: A) In situ hybridization using a biotinylated white DNA probe to polytene chromosomes from transformant larvae containing the 35UZ transposon. The arrow indicates the presence of 35UZ at 34E. This transformant line contains another insertion of transposon. The probe also detects the endogenous white gene. Thus one nucleus has three strong signals.

B) Immunostaining of the polytene chromosomes from the 35UZ-containing transformant larvae with anti-trx antibodies. The arrow indicates presence of HRP signal indicative of trx binding at the site of insertion of the transposon, 34E. The line represents wild type trx binding sites.

C) Immunostaining of the polytene chromosomes from wild type larvae. The arrowhead indicates absence of signal at 34E.
Figure 12: Immunostaining of polytene chromosomes from the 22UZ transformant: A) Immunostaining of the polytene chromosomes from the 22UZ-containing transformant larvae with anti-trx antibodies. The arrowhead indicates absence of HRP signal at the site of insertion of the transposon, indicating that 22UZ does not contain a detectable trx binding site. B) In situ hybridization using a biotinylated white DNA probe of polytene chromosomes from transformant larvae containing the 22UZ transposon. The arrow indicates presence of 22UZ at 73F.
binding site. Together with the 35UZ results, this restricts the \(trx\) responsive sequences to the 13 kb DNA from map coordinates -9 to +4.

A transposon carrying a 14 kb fragment from coordinates -18.5 to -4 is predicted to contain the sequences required for the long term maintenance of the \(Ubx\) expression pattern (Simon et al., 1993; Simon et al., 1990). It partially overlaps with the 22UZ construct and is completely contained within the 35UZ.

This transposon creates a new binding site for the \(trx\) proteins at its insertion site at 62A, on the left arm of the third chromosome (Figure 13A). Figure 13B shows a wild type chromosome with the wild type \(trx\) binding sites and absence of signal at 62A.

Similar analysis using anti-\(ph\) and anti-\(Pc\) antibodies show that this transposon contains sequences recognized by the Pc-G proteins (DeCamillis et al., 1992; A. Chiang personal comm.). This demonstrates that the \(Pc\) and \(trx\) proteins are truly co-localized at the \(Ubx\) promoter as was suggested by the co-localization data presented in the chapter II.
Figure 13: **Immunostaining of polytene chromosomes from the 14 kb UCR fragment containing transformant:**

A) Immunostaining of polytene chromosomes from the 14 kb containing transformant larvae with anti-\(trx\) antibodies. The arrow indicates presence of HRP signal indicative of \(trx\) binding at the site of insertion of the transposon, 62A. The lines represents wild type \(trx\) binding sites.

B) Immunostaining of the polytene chromosomes from wild type larvae. The arrowhead indicates absence of signal at 62A.
Delimiting the \textit{trx} binding sites in the UCR

We obtained transformant stocks carrying various smaller fragments of the UCR DNA from Dr. V. Pirrotta's lab. These fragments create new binding sites for the \textit{Psc} protein, a member of the Pc-G. We tested these different transformants for the presence of the \textit{trx} binding sites.

The 2.2 kb KpnI-EcoRI fragment contains the PRE and creates a new \textit{trx} binding site at 21F (Figure 14A) which is absent in the wild type chromosome not carrying the transposon (Figure 14B). The \textit{Psc} protein, a member of the Pc-G proteins, also binds to the same transposon creating a signal at 21F, the site of integration (Rastelli, 1994; Figure 14C). The wild type chromosome stained with anti-\textit{Psc} antibodies does not show signal at 21F (Figure 14D).

The 670 bp PstI-NdeI containing transposon also creates a new binding site for the \textit{trx} proteins at 65B, the site of integration of the transposon (Figure 15A). This site is also seen with the anti-\textit{Psc} antibodies (Figure 15C; Rastelli, 1994). The site is absent in the wild type chromosomes stained with anti-\textit{trx} and anti-\textit{Psc} antibodies separately (Figures 15B and D).

The 0.5 kb StyI-PstI fragment situated next to the PstI-NdeI fragment doesn't create a detectable \textit{trx} binding site (data not shown). The \textit{Psc} protein also does not bind to this transposon (Rastelli, 1994).
Figure 14: Immunostaining of polytene chromosomes from the 2.2 kb UCR fragment containing transformant:
A) Immunostaining of the polytene chromosomes from the 2.2 kb containing transformant larvae with anti-trx antibodies. The arrow indicates presence of HRP signal indicative of trx binding at the site of insertion of the transposon, 21F. The lines represent wild type trx binding sites.
B) Immunostaining of the polytene chromosomes with anti-trx antibodies from wild type larvae. The arrowhead indicates absence of signal at 21F.
C) Immunostaining of the polytene chromosomes from 2.2 kb transformant larvae with anti-Psc antibodies. The arrow indicates presence of signal at 21F.
D) Immunostaining of the polytene chromosomes with anti-Psc antibodies from wild type larvae. The arrowhead indicates absence of signal at 21F.
Figure 15: Immunostaining of polytene chromosomes from the 670 bp UCR fragment containing transformant:
A) Immunostaining of the polytene chromosomes from the 670 bp containing transformant larvae with anti-\textit{trx} antibodies. The arrow indicates presence of HRP signal indicative of \textit{trx} binding at the site of insertion of the transposon, 65B. The lines represent wild type \textit{trx} binding sites. This 670 bp fragment contains the \textit{Ubx} PRE. Binding of \textit{trx} to the PRE suggests that the Pc-G and the \textit{trx} proteins may be physically associated with each other.
B) Immunostaining of the polytene chromosomes from wild type larvae with anti-\textit{trx} antibodies. The arrowhead indicates absence of signal at 65B.
C) Immunostaining of the polytene chromosomes from the 670 bp containing transformant larvae with anti-\textit{Psc} antibodies. The arrow indicates presence of signal at 65B.
D) Immunostaining of the polytene chromosomes from wild type larvae with anti-\textit{Psc} antibodies. The arrowhead indicates absence of signal at 65B.
These results indicate that a $\textit{trx}$ binding site(s) lies within the 670 bp fragment of the UCR which also contains binding sites for the Pc-G proteins. The binding of $\textit{trx}$ to the fragment containing the PRE also suggests a very close proximity of the $\textit{trx}$ and Pc-G proteins at the DNA level, at least within the $\textit{Ubx}$ UCR, but perhaps at other loci where they appear to be co-localized. Thus these experiments identified a $\textit{trx}$ binding site(s) within $\textit{Ubx}$ UCR and demonstrated that $\textit{trx}$ acts directly on its target genes. Similar analysis of the DCR is required to determine if it contains any other $\textit{trx}$ binding sites or if the UCR-binding site is the only $\textit{trx}$ binding site present in the $\textit{Ubx}$ gene.
DISCUSSION:

The 670 bp fragment of the *Ubx* UCR contains binding site(s) for both *trx* and Pc-G proteins.

Immunostaining of the transformant larvae carrying the *Ubx* PRE shows binding of the *trx* proteins detected by the presence of a new site on the polytene chromosomes. This indicates that *trx* interacts directly with the regulatory region of *Ubx* and very likely of its other target genes. It may act as a sequence-specific DNA-binding protein or may bind indirectly to specific DNA sequences by complexing with another DNA-binding protein(s).

The binding of *trx* to the PRE occurs in the absence of the normal UCR DNA context and in a new chromosomal environment, indicating that the binding is sequence-specific.

The 22UZ and 0.5 kb StyI-PstI fragments do not contain *trx* binding sites, indicating that *trx* binding to the other transposons is due to the UCR sequences and not due to the vector DNA.

Presence of the *trx* binding site(s) in PRE suggests that it contains the *trithorax* response element (TRE), although the genetic test for the presence of TRE needs to be done.
The *trx* and Pc-G proteins are co-localized at the *Ubx* promoter

The Pc-G proteins bind to the same fragment as do the *trx* proteins implying that the response elements for the two antagonistic proteins are located very close to each other. Further dissection of this region into smaller DNA fragments and identification of the binding site sequence will determine whether they have overlapping binding sites. This demonstrates that the Pc-G and *trx* proteins are co-localized within a short distance at the *Ubx* promoter. This could be true for the other loci where they seem to be co-localized on polytene chromosomes in the dual fluorescence experiments described earlier (Chapter II).

Do these proteins act independently or are they part of one complex? This question can be answered when the binding sites for these proteins is determined. None of the six Pc-G proteins characterized to date exhibit DNA-binding activity in *in vitro* (Paro, 1993; Pirotta personal communication). The *trx* proteins contain a putative DNA-binding motif, suggesting they might bind DNA directly. DNA-binding experiments using the putative DNA binding domain of *trx* and the 670 bp NdeI-PstI fragment containing the PRE are presented in Chapter IV.
The $trx$ proteins bind constitutively to their target genes

Binding of $trx$ to the $Ubx$ PRE in salivary glands where $Ubx$ is normally repressed suggests that $trx$ binding is constitutive. Some other signal along with $trx$ may be required to maintain the active complex. Binding of the $trx$ and the Pc-G proteins to the same DNA fragment of 670 bp suggests that the previously anticipated competition between these antagonistic transcriptional regulators is not at the level of DNA binding but could be at some higher level. It also indicates that the binding of the Pc-G proteins does not prevent binding of the $trx$ proteins over a small molecular distance of 670 bp.

The mode of action of the $trx$ and Pc-G proteins

The $Ubx$ PRE/TRE behaves differently than typical enhancer elements. In the $lacZ$ expression assay it does not show any recognizable $lacZ$ expression pattern. But when it is present along with the known $Ubx$ enhancers, it maintains the expression pattern of that enhancer element throughout development (Chan et al., 1994). This suggests that the PRE fragment is involved only in maintenance and not in establishing the expression patterns of homeotic genes by segmentation genes.
As mentioned before the PRE element is not involved in the early determination of homeotic gene expression patterns as suggested by the \textit{lacZ} expression pattern in PRE transformants. Binding of the Pc-G and the trx-G proteins to the \textit{Ubx} PRE/TRE indicates that these proteins maintain the expression pattern of \textit{Ubx} but are not involved in the initial determination of expression pattern. In contrast, the enhancer elements containing multiple binding sites for the segmentation gene products show very distinct early \textit{lacZ} expression patterns which decay prematurely. This suggests that the maintenance of the \textit{Ubx} expression by the Pc-G and the trx-G proteins through PRE/TRE is different than that done by the segmentation gene products.

These proteins may not act as typical transcription factors binding to the enhancers and controlling gene activity. Instead they may work by controlling the chromatin structure in the vicinity of a promoter by interacting with the PRE/TRE.

The detailed analysis of the UCR has revealed only one PRE (Chan \textit{et al.}, 1994). In contrast to this discrete PRE, in chromatin cross-linking experiments the \textit{Pc} protein has been shown to be distributed over the entire \textit{Ubx} gene in cells where it is repressed (Orlando and Paro, 1993). The PRE may constitute a strong Pc-G binding site where the Pc-G complexes are initially nucleated and from which additional complexes spread throughout out the gene. This complex could be equivalent to
heterochromatinization, shutting off the *UBx* gene in the anterior part of the embryo (Bienz, 1992; Chan *et al*., 1994).

We analyzed approximately 35 kb of the *UBx* UCR for the presence of a *trx* binding site with the exception of 4 kb DNA contained with the 6.5 kb HindIII fragment. This analysis identified a *trx* binding site(s) within a single fragment, the 670 bp PstI-NdeI fragment containing the PRE. This suggests that this may be the only *trx* binding site(s) in the UCR which is located approximately 25 kb away from the *UBx* promoter.

The presence of *trx* binding site(s) at a distance from the *UBx* promoter suggests that the *trx* proteins are able to act long distances in controlling the promoter activity. It is also possible that the *trx* proteins act locally as an "antirepressor" of the Pc-G proteins, preventing the spreading of the Pc-G complex, thus counteracting the repression. The antirepression activity has been demonstrated for one of the *trx*-G protein, *trithorax-like* which encodes the GAGA factor (Croston *et al*., 1991; Farkas *et al*., 1994).

**Determination of the expression state of a promoter**

One of the models proposed to explain the action of the *trx*-G and the Pc-G proteins presupposes that the late complexes formed at the PRE/TRE senses the state of activity of promoters or enhancers in the vicinity (Paro, 1990). If the enhancers are in
the active state, repression is not established and the enhancer-promoter complex may be locked in an active configuration. The active state may be viewed as the absence of nucleosomes in the vicinity or formation of the enhancer-promoter loop by the activation complex. The trx-G proteins would then maintain the active state by forming a chromosomal complex which allows the sequence-specific DNA-binding activator proteins to bind the enhancer efficiently. As described earlier, some of the other trx-G proteins characterized to date are likely to act directly on chromatin structure. If the region is not transcriptionally active, the repressive complex of the Pc-G proteins proceeds to assemble, locking the chromatin region in a configuration that prevents expression.

How is the active vs. repressed state of a promoter recognized and transduced by the trx-G and Pc-G proteins when our data shows both the activator and the repressor proteins are bound to the target DNA sequences at the same time in the same tissue? The active vs. repressed state is determined by the early acting segmentation gene products which bind the enhancer DNA in a sequence-specific manner (Qian et al., 1991; Zhang et al., 1991). The segmentation genes themselves are expressed in spatially restricted domains, in contrast to the genes of the Pc-G and the trx-G which are expressed ubiquitously in the embryo (DeCamillis et al., 1992; Knipple et al., 1985; Kuzin et al., 1994; Pignoni et al., 1990; Tautz, 1988; Zink and Paro, 1989). Thus the
expression patterns of the segmentation genes provide positional information within blastoderm cells which determines the transcriptional state of the homeotic genes (Irish et al., 1989; Reinitz and Levine, 1990; White and Lehmann, 1986). The transcriptional status of the homeotic genes is dependent on the concentration and combinatorial effects of different segmentation gene products (Ingham, 1988). How the transition from the early acting complexes to the maintenance complexes takes place is still unclear but the bound molecules of the segmentation gene products may help in transition (Bienz, 1992).

Once a given state is established, then the Pc-G proteins bind only at the repressed promoter. At the active promoter the "open complexes" prevents the formation of the repressive complex. The open complex maintained by the trx-G proteins, allows the dedicated factors to bind their respective target promoters. Very little is known about the proposed dedicated activators.

The proteins of the Pc-G and the trx-G demonstrate the ability to form the appropriate complexes at the respective promoters after every cell division in the absence of segmentation gene products. The heritability of the maintenance complexes may be provided by at least two factors. One of the proteins from the trx-G and the Pc-G must act as a sequence-specific DNA-binding protein, thus tethering the complexes to
the DNA. Also the structure of these complexes may give the required stability for heritability. These multiprotein complexes may "linger on" during replication and reassemble on the DNA of both daughter strands after replication, due to the multiple cooperative interactions between the proteins constituting them (Bienz, 1992).

The analysis of the UCR of Ubx described here has identified a 670 bp fragment which contains a trx binding site(s). The trx proteins contain a putative DNA-binding motif homologous to the DNA-binding domain of the nuclear receptor family. This suggests that the trx proteins may bind DNA directly in a sequence-specific manner and help in anchoring the proposed trx-G protein complex onto the chromosomes.
CHAPTER IV: The \textit{trx} DNA-binding domain binds DNA with high affinity but uncertain specificity

INTRODUCTION:

Proteins of the \textit{trx-G} and the \textit{Pc-G} are required continuously for the maintenance of the spatially restricted expression patterns of homeotic genes. In the previous chapters we showed that the \textit{trx} proteins are associated with polytene chromosomes at 63 specific sites. These sites include those of its known targets, the genes of the BX-C, the ANT-C and \textit{engrailed}. Immunostaining of polytene chromosomes from transformant larvae carrying various fragments of the \textit{Ubx} regulatory DNA allowed us to map one \textit{trx} binding site to the 670 bp fragment containing the \textit{Ubx} PRE/TRE. The binding of \textit{trx} to specific sites on the polytene chromosomes and to the PRE/TRE indicates that \textit{trx} binding is sequence-specific. The \textit{trx} proteins may bind DNA directly or may do so indirectly by binding to another sequence-specific DNA-binding protein.

The putative \textit{trx} DNA-binding domain

Consistent with this possibility, the two \textit{trx} protein isoforms contain a novel variant of the DNA binding domain (DBD) motif found in members of the nuclear receptor superfamily (Figure
16). The \( trx \) proteins have no homology to the members of the nuclear receptor family outside of the DNA-binding domain. But there are certain features which are shared by the members of this family which are also shared by \( trx \). One of the common features is that the DNA-binding domain is encoded by two exons (Evans, 1988). In \( trx \) the putative DBD is encoded by separate exons, similar to the other members of the family. Nuclear receptors contain a highly acidic region N-terminal to the DBD which is involved in transcriptional activation (Evans, 1988). Similarly \( trx \) contains an acidic domain which is N-terminal to the putative DBD.

**Structure of the nuclear receptor DBDs**

The structures of the DBDs of glucocorticoid (GR), estrogen (ER) and the retinoid X (RXR) receptors have been determined by X-ray crystallography (Hard \textit{et al.}, 1990; Lee \textit{et al.}, 1993; Luisi \textit{et al.}, 1991; Schwabe \textit{et al.}, 1990). The structure of the DBD-DNA co-crystals have also been determined for GR and ER. Based on these structural data and site-directed mutagenesis, residues important for forming the compact globular structure and those which are responsible for specific base and phosphate backbone contacts have been identified.
Figure 16: Alignment of the residues of the putative trx DBD with the other members of the nuclear receptor family: Residues 762 to 860 of the trx proteins aligned with selected nuclear receptor superfamily members chosen to show conservation as well as variation within the superfamily. Shading indicates the residues highly conserved in all family members or trx residue conserved in at least one other. The location of the two α-helices after each finger in the glucocorticoid receptor structure is indicated by »» above. Residues contributing to glucocorticoid receptor DBD structural stability are indicated by 1) C above the 8 invariant Cys residues that form the Zn++ coordination sites 2) ± above highly conserved hydrophobic residues whose side chains interact in the glucocorticoid receptor to form an aromatic cluster at the crossing of the two α-helices 3) ↓ above residues (including the ninth invariant Cys) whose side chains participate in formation of a hydrophobic shell around the aromatic cluster in the glucocorticoid receptor, creating a very stable hydrophobic core which fixes and orients the two crossed α-helices. Highly conserved residues not conserved in trx are indicated by an * and include Asp/Glu6 (trx = Ala) and His18 (trx = Lys) in the first finger. List includes: odr-7, C. elegans odorant response protein; hGR and mGR, human and mouse glucocorticoid receptors; hMR, human mineralocorticoid receptor; hAR, human androgen receptor; hPR, human progesterone receptor; hER, human estrogen receptor; hERR1, hERR2, human estrogen receptor related; hTR2, human thyroid receptor; hVDR, human vitamin D receptor; hCOUP, human COUP homolog; hEAR1, human EAR homologs, NGF1, rat nerve growth factor induced protein; HNF4, rat hepatic nuclear factor 4; hRXR, mRXR, human and mouse retinoid X receptors; AD4BP, activator of P450 genes, EcR, Drosophila ecdysone receptor, E75A, Drosophila ecdysone induced protein; usp, Drosophila ultraspiracle protein; Ftzfl, Drosophila ftz activator protein; tll, Drosophila tailless protein. List also includes 3 Drosophila proteins with no receptor homology outside the DBD: egon, Drosophila embryonic gonad protein; kni, Drosophila knirps gap protein; knrl, Drosophila knirps-related protein.
All the DBDs contain nine invariant cysteine residues the first eight of which are coordinated by two Zn\(^{++}\) ions to form two Zn finger-like structural motifs. Substitution of these cysteines individually with serine or alanine, by site-directed mutagenesis, demonstrated that the mutated DBD was unable to coordinate Zn\(^{++}\) ions and did not bind DNA (Severne et al., 1988). The ninth cysteine is not involved in the Zn\(^{++}\) ion coordination but contributes to structural stability of the DBD fold (Severne et al., 1988). This presumptive trc DBD contains all the nine cysteine residues.

The crystallographic analysis of the GR DBD bound to its target DNA sequence revealed that it forms a compact globular fold comprised of two "modules" each composed of a Zn finger. Each module is nucleated by a Zn coordination center and is followed by an amphipathic α helix (Figure 17; Luisi et al., 1991).

Module I includes the first Zn finger and contains all the residues responsible for specific base as well as phosphate backbone contacts (Luisi et al., 1991; Umesono and Evans, 1989). This region of module I is called the Proximal or the P box (Umesono and Evans, 1989). These residues are highly conserved among nuclear receptor DBD. This conservation is reflected in the similarity of the recognition sequences of these DBDs.
Figure 17-1: Structure of the DNA binding domain of the Glucocorticoid receptor (Luksi et al. 1991): The two modules are indicated by the brackets. Eight out of nine conserved cysteines (C) are involved in Zn coordination. The α helical segments are boxed. Residues that make dimer interface contacts are indicated by the solid dots. Residues making phosphate backbone contacts at the specific and nonspecific sites are indicated by solid and open rectangles, respectively. Three amino acids that direct the discrimination of GRE and ERE are indicated with white lettering in solid boxes and those which discriminate between ERE and thyroid response elements are in discs.
Figure 17-2: The putative *trx* DNA binding domain showing the Cys-Zn coordination centers and α helical segments which are boxed: This structure is predicted from the structure of GR DBD. The extra residues in the *trx* DBD are indicated as "SAN.....C" represents the nine cysteine residues which are conserved in all the nuclear receptors are also conserved in *trx*. The highly conserved hydrophobic residues which are important for the formation of aromatic cluster are indicated as *. The putative P box of *trx* (squared letters) is different than the nuclear receptor P boxes suggesting that if *trx* DBD binds DNA it will recognize a different nucleotide sequence.
All nuclear receptors bind as homo or hetero-dimers to a palindromic or a tandemly repeated pair of "half sites" (Evans, 1988). The recognition sequences, called response elements (RE), of the GR, ER and thyroid (TR) receptors differ from each other in the spacing between the two half sites and few other bases (Berg, 1989). Other members of the superfamily such as *tailless* and *FTZ-F1* have variant P boxes reflecting the variant sequences of their binding sites (Liaw *et al*., 1993; Segraves, 1991). The residues comprising the *trx* P box differ from those most frequently found in the P boxes of the majority of the nuclear receptor DBDs. This suggests that if *trx* binds to DNA, the sequence of its binding site is likely to be different from the recognition sequences of other DBDs in this superfamily.

Module 2 or the second Zn finger contains sequences important for DBD dimerization and additional phosphate backbone contacts. The Distal or the D box within this module contains sequences which are involved in the DBD dimerization (Berg, 1989; Hard *et al*., 1990; Umesono and Evans, 1989). The second finger is more variable among the members of the family compared to the first finger (Luisi *et al*., 1991).

A linker region between the two modules typically contains 15-17 amino acids. The *trx* linker region differs in that it contains 14 extra amino acids. The *trx* DBD also has 6 additional residues in the loop of the first finger and 10 in the loop of the second finger. Some length variability is found in these regions.
of other nuclear receptor DBDs, but \( \text{trx} \) exhibits a much greater degree of variability than any others.

However, if the \( \text{trx} \) sequence is superimposed on the glucocorticoid receptor DBD structure, all these 'insertions' lie on exposed surfaces, away from DNA contact sites. In these positions the additional residues might be accommodated so as to preserve the basic form and integrity of this structural motif and its ability to interact with DNA. At least one other member of the family, \( odr-7 \), contains four extra residues in the second loop (Sengupta et al., 1994).

A conserved basic residue near the base of the second finger, which makes stabilizing phosphate contacts, is conserved in \( \text{trx} \). The \( \text{trx} \) DBD also contains three additional basic residues nearby, imported with the ten residue "insertion", which might be involved in additional stabilizing phosphate contacts.

Residues important for the structural stability of nuclear receptor DBDs

The similarities among the structures of the GR, ER and RXR suggest that the structures of other DBDs of this family will have fundamentally similar overall structures. The residues which are important for structural stability will be the same based on the high conservation of residues seen among the members.
In the GR, the two modules are held together at an approximately 90° angle to form a single compact globular domain. The main interactions stabilizing this globular fold are tetrahedral coordination of two Zn++ atoms by the eight invariant cysteines and interactions among the aromatic side chains of two invariant phe and two highly conserved tyr residues (See Figure 16; Luisi et al., 1991). The latter form part of an aromatic cluster that stabilizes the fold and fixes the orientation of the two modules.

A "hydrophobic shell" surrounds this aromatic cluster. The residues which are involved in the formation of this hydrophobic shell and are conserved among all the members of the family are also conserved in trx with some exceptions. The important exceptions include the invariant his\textsubscript{12} (trx\textsubscript{18} = lys) and conserved asp/glu\textsubscript{6} (trx\textsubscript{6} = ala) in the loop of the first finger (Figure 16). Changing them to gly has only a slight effect on DNA-binding and activation of transcription (Hollenberg and Evans, 1988). In the glucocorticoid receptor DBD the invariant His\textsubscript{12} makes specific phosphate backbone contacts suggesting that the positively charged Lys at this position in trx may function similarly (Luisi et al., 1991).

The aromatic cluster and the hydrophobic shell are important for the structural stability of the DBD. The trx DBD has conserved virtually all the residues which are important for the form and stability of this structural element. It therefore would be
expected to form a fundamentally similar structure capable of interacting specifically with DNA.

The structure of the GR-DBD bound to its GRE demonstrated that the GR-DBD interacts with the major groove of the DNA molecule (Hard et al., 1990; Luisi et al., 1991). The DBDs of nuclear receptors undergo dimerization after making contact with the DNA molecule. The dimerization promotes cooperative binding which is necessary for sequence-specific DNA binding (Luisi et al., 1991). Some nuclear receptor DBDs contain a third α helix which lies immediately after the second module (Lee et al., 1993). This third helix in RXR mediates both protein-protein and protein-DNA interactions required for cooperative, dimeric binding of the RXR DBD to DNA. Residues N-terminal to the first conserved cysteine are not important for DNA binding (Luisi et al., 1991).

Does the putative trx DBD bind DNA?

The sequence similarities among the DBDs of the other members of the nuclear receptor family and the trx DBD discussed above suggest that this putative trx DBD may bind DNA in a similar sequence-specific manner. To test this, we expressed the trx-DBD as a glutathione-S-transferase (GST) fusion protein. The DBD of most receptors when expressed in E. coli are sufficient to bind DNA (e.g. Freedman et al., 1988;
Schwabe et al., 1990). We analyzed DNA-binding by the GST-DBD by electrophoretic mobility shift assay (EMSA) (Carey, 1991). We used the 670 bp PstI-NdeI fragment from Ubx UCR which contains the Ubx PRE and trx binding site(s) (Chapter III).

The EMSA using GST-DBD fusion protein and radiolabeled 670 bp PstI-NdeI fragment demonstrated that the fusion protein binds DNA at 10 nM concentration. The DNA probe was used at 1 nM. Binding of the fusion protein to the DNA fragment at such low concentration suggests that the fusion protein binds DNA with high affinity. The GST protein alone does not bind DNA indicating that the trx residues present in the fusion protein are responsible for DNA binding. Competition assays where unlabeled, non-specific, linear DNA molecules were used indicated that the GST-DBD exhibits low sequence-specificity. But when supercoiled DNA constructs containing specific and non-specific inserts were used as competitor DNAs, the fusion protein exhibited some sequence specificity. These contrasting results suggest that the GST-DBD fusion protein has some sequence-specificity but also exhibits non-specific binding, perhaps to the ends of the linear fragments used.
Materials and Methods:

Isolation of a cDNA encoding the *trx* DBD: An embryonic cDNA library from 3-12 hr. embryos was screened with a 0.9 kb HindIII genomic fragment containing the DBD. After a tertiary screen the positive phage recognized by this probe was grown in a liquid culture and the DNA was isolated by the following method.

Lambda DNA preparation: The isolated plaque was resuspended in 500 µl of lambda dilution buffer (10 mM Tris-Cl, pH 7.5; 10 mM MgSO4). After adding a drop of chloroform it was stored at 40°C for at least one hour before proceeding further. The competent cells grown in LB and maltose (100 µl) were mixed with 100 µl of the phage suspension and 100 µl of 10 mM CaCl2 and 10 mM MgCl2. This mixture was incubated at 37°C for 15 min. and was added to 4 ml of the prewarmed LB medium. The incubation was carried out at 37°C for 5-6 hr. Once the cells were lysed, two drops of chloroform were added and the incubation was continued for another half an hour. The culture was spun down at 3K for 15 min. and the supernatant was stored at 40°C (phage stock).

For large DNA prep, (50 ml culture), 100 µl of the competent cells were mixed with 100 µl of 10 mM CaCl2 and 10 mM MgCl2 and two different dilutions of the phage stock prepared earlier.
One hundred μl of 1:10 and 1:100 dilutions were used here. The cultures were incubated at 37°C for 15 min, and were mixed with 50 ml of prewarmed LB. The cultures were incubated at 37°C for 8 hr. Once the lysis occurred, 100 μl of chloroform was added and incubated for another half an hour. The cultures were spun at 8K for 15 min to pellet cell debris.

Ten μl of 5 mg/ml DNase and 25 μl of 10 mg/ml RNase was added to the supernatant and incubated at 37°C for 1 hr. The supernatant was spun at 20K for 2.5 hr in a JA 20 rotor to pellet the phage particles. The supernatant was discarded. The pellet was resuspended in 200 μl of lambda dilution buffer after air drying. It was extracted with an equal volume of phenol by vortexing for half hour at 40°C. It was then extracted twice with equal volumes of phenol:chloroform and once with chloroform. The phage DNA was ethanol-precipitated, dried and resuspended in TE.

**Preparation of the GST-DBD fusion protein:** By using primers with the EcoRI restriction sites on the end, a 1 kb PCR product was generated using the cDNA template isolated above. The primers were designed such that the PCR product was in frame with the GST-fusion protein vector (Pharmacia, a gift from Dr. Mark Johnson). The GST-fusion vector used here is modified to contain a site for heart muscle kinase and a "flag" epitope. The anti-flag antibodies (Kodak) were used to detect the fusion
protein by the Western analysis. Western analysis was carried out as described in chapter II.

The 1 kb PCR product was cloned into the GST-fusion vector and was used to transform XL-1 blue bacterial cells. The fusion protein is under the control of a 'taq' promoter which has a lac operator. The fusion protein is induced by adding IPTG to the medium. The induction and purification was essentially carried out as described in Current Protocols in Molecular Biology (Ausubel et al., 1991).

Cells containing the GST-DBD construct were grown overnight at 37°C in LB + 50 μM ZnSO₄. Zn is added to the medium to ensure proper formation of the Zn finger structure of the DNA-binding domain (Del Rio, 1991). The following day 1 ml of the overnight culture was diluted with 100 ml of fresh LB + 50 μM ZnSO₄. The culture was allowed to grow to an O.D. of 0.7 at 600 nm. The cells were induced with 1.5 mM IPTG and were grown for three hours.

At the end of the incubation cells were harvested by centrifugation at 5K for 10 min. at 4°C. The cells were resuspended in 2 ml of Buffer I (PBS + 50 μM ZnSO₄ + 5 mM DTT) and were lysed by sonication. TritonX-100 at 0.1% was added and mixed by swirling. The suspension was incubated on ice for 5 min. The cell debris was pelleted by centrifuging at 10K, 4°C for 5 min. The supernatant contains soluble GST-fusion protein.
The GST-DBD fusion protein was purified from the supernatant by using glutathione-agarose beads (Pharmacia). Glutathione is a substrate for Glutathione-S-Transferase, thus enabling the fusion protein to be associated with the beads. The beads were first equilibrated with PBS at room temperature for 1 hour and were resuspended in PBS to make a 50% slurry.

One hundred μl of the 50% slurry was added to 2 ml of the above supernatant containing the fusion protein and allowed to incubate at room temperature for 5 min. The beads were separated by centrifugation at 2 k for 30 sec and washed twice with buffer A + 0.1% tritonX-100 followed by three washes with buffer A. The fusion protein is eluted from the beads with 50 mM reduced glutathione in 10 mM Tris-Cl, pH 8.0 and stored at 4°C.

**Electrophoretic Mobility Shift Assay (EMSA):** EMSA was carried out by using guidelines from the protocol described in the Current protocols in Molecular biology (Ausubel et al., 1991). The 670 bp PstI-NdeI fragment containing the PRE (Chapter III) was used in these experiments. The DNA fragment was end-labeled using P\(^{32}\). A typical binding reaction consists of 1 nM of labeled DNA, 1X DNA binding buffer (4 mm Tris-Cl pH 7.5, 80 mM NaCl, 0.5 mM dithiothreitol, 1 mM EDTA, 0.5 mM ZnSO₄; (Gogos et al., 1992)), 2μg BSA, 10% glycerol, 10-20 nM of
purified GST-DBD protein, 7 mM MgCl₂, and water to make the reaction volume to 20 µl.

The reactions were incubated at room temperature for 1 hour and were electrophoresed through a 4% low-ionic strength polyacrylamide gel with 1X running buffer (4% acrylamide gel mix: 0.8 ml of 50x running buffer, 4 ml of 40% acrylamide solution (Amresco), 1 ml of 2% bisacrylamide, 2 ml of 50% glycerol and water up to 40 ml. The mixture was degased for 15 min. and 300 µl of ammonium persulphate and 30 µl of TEMED were added. 50X Low-ionic-strength buffer: 0.335 M Tris-Cl, pH 7.9, 0.165 M sodium acetate, pH 7.9, 0.05 M EDTA, pH 8.0). Electrophoresis was carried out at a constant current of 30 mA. The gel was pre-run for an hour at 100 volts before loading the samples. The samples were electrophoresed until the blue dye front was 2/3rd into the gel. After electrophoresis the gel was dried and autoradiographed.
RESULTS:

The *trx* DNA-binding domain (*trx*-DBD) is encoded by two exons

The *trx* proteins contain a region of homology with the DNA binding domain of the nuclear receptor family (Stassen *et al.*, 1995: Figure 16). Like genes for the other nuclear receptor DBDs whose genome structure has been determined, the putative *trx*-DBD is encoded by two exons, separated by a 63 bp intron. An embryonic cDNA library was screened using a 0.9 kb HindIII genomic fragment containing the *trx* DBD sequences, to isolate a cDNA containing the *trx*-DBD. After purifying the phage containing the *trx*-DBD, the insert was isolated by performing a polymerase chain reaction (PCR) on the phage DNA. The 1 kb PCR fragment containing *trx* nucleotides 3113- 4114 was then cloned into a GST fusion vector which has a site for the heart muscle kinase and the "flag" epitope.

The expression and purification of the GST-DBD fusion protein

The *trx*-DBD was expressed as a GST fusion protein (GST-DBD) using the clone containing the 1 kb PCR product from the cDNA template. The expected molecular weight of the fusion protein is
66 kD. But the GST-DBD fusion protein construct after induction gave a 42 kD induced protein instead of the expected 66 kD. Western analysis using anti-Flag antibodies confirmed that the 42 kD protein seen on the coomassie gel is the GST-DBD fusion protein (data not shown).

DNA sequence analysis of this construct showed that the DNA binding domain is intact but contains a one base pair deletion after the putative DBD. This deletion caused a frame shift which introduced a stop codon at nucleotide 3534. The deletion is presumably a PCR artifact. Since it contained the entire DBD, we decided to use it in electrophoretic mobility shift assays (EMSA). This fusion protein, referred here as GST-DBD I, contains six amino acids before the first conserved cysteine and 38 residues after the end of the homology.

The GST-DBD I fusion protein was purified from the total cell extract using affinity chromatography on glutathione sepharose beads (Pharmacia). It is highly soluble and thus was easier to purify by affinity chromatography. The purified GST-DBD I was used in the EMSA to determine if it binds the DNA fragment containing the Ubx PRE.
The GST-DBD I fusion protein binds DNA with high affinity but low specificity.

EMSA using the end labeled 670 bp PstI-NdeI fragment and increasing concentrations of the GST-DBD I protein showed that the minimum concentration of the fusion protein required to shift the DNA is 10 nM (Figure 18). The GST protein alone at 100 nM does not shift the labeled 670 bp PstI-NdeI probe (Figure 19). This indicates that the DNA binding activity of the fusion protein is due to the trx residues present in the fusion protein.

To determine the specificity of this DNA-protein interaction, we performed competition assays using three different linear unlabeled non-specific DNA fragments and a constant protein concentration of 10 nM. The unlabeled 670 bp PstI-NdeI fragment was used as a specific competitor DNA. The difference between the concentration of the unlabeled specific and non-specific fragments required to abolish the DNA binding observed as a "shifted" band on the gel is an indicator of the specificity of the DNA binding. This difference between the concentrations of the specific and non-specific DNA fragments required to abolish the "shift" was negligible (Figure 20).
Figure 18: The GST DBD fusion protein binds DNA with high affinity: EMSA using increasing amounts of fusion protein and constant amount (1 nM) of the radiolabeled 670 bp PstI-NdeI fragment demonstrated that the minimum amount of protein required to shift some of the DNA molecules is 10 nM. Lane 1: DNA fragment alone without the fusion protein. This is a control for the mobility of the unbound DNA. Lanes 2-7: Increasing amounts of fusion protein with 10 nM, 20 nM, 30 nM, 40 nM, 50 nM, 100 nM and 500 nM respectively.
Figure 19: The \textit{trx DBD binds DNA}: Lane 1: The radiolabeled 670 bp PstI-NdeI fragment without the fusion protein. Lanes 2 and 3: DNA fragment with 50 nM and 100 nM GST DBD fusion protein respectively. Lane 4: DNA fragment with 100 nM GST protein. Electrophoretic mobility of the DNA in the presence of GST protein was not changed suggesting that it does not bind DNA. The shift in mobility observed in presence of the fusion protein (lanes 2 and 3) suggest that the \textit{trx} residues of fusion protein are responsible for DNA binding.
Figure 20: **Determination of specificity of the GST-DBD I protein by using linear fragments:** The unlabeled 670 bp PstI-NdeI DNA fragment was used as a specific competitor and three different DNA fragments of comparable length were used as non-specific competitor DNAs. 1 nM labeled 670 bp PstI-NdeI fragment and 10 nM GST-DBD I fusion protein was used in each reaction. Increasing amounts of the unlabeled DNA fragments were added. Lane 1: Labeled DNA without the fusion protein. Lane 2: 10 nM of GST-DBD and 1 nM of radiolabeled 670 bp PstI-NdeI fragment. Competitor DNA was not added and serves as a positive control for the DNA binding reaction. Lanes 3-7: 1, 10, 25 and 50 nM of specific unlabeled DNA was added to the reaction respectively. Lanes 8-12: 1, 10, 25 and 50 nM of a non-specific unlabeled DNA was added to the reaction respectively. The specific and non-specific DNA competed the specific DNA binding at approximately same concentration. This suggests that the GST-DBD I fusion protein binds DNA with low sequence specificity. Similar results were obtained with the other two non-specific DNAs (Data not shown).
The other non-specific DNA samples such as poly dI-dC and salmon sperm abolished the observed "shift" at a very low concentration of 0.5 μg/ml (Figure 21). This indicates that these non-specific DNAs compete well with the specific DNA binding of the fusion protein. These results suggests that the GST-DBD I fusion protein binds with high affinity (Kd ≤ 10 nM) but exhibits low specificity, under the conditions used.

The GST-DBD I contained the DBD with only few amino acids after the homology. We tested whether extra amino acids after the homology will give the required specificity of the DNA-protein interaction. We replaced the 0.5 kb DraIII-HindIII fragment of the GST-DBD I construct with an equivalent fragment of the genomic DNA to eliminate the stop codon introduced by the PCR error described earlier. This construct when induced gave the expected 66 kD fusion protein, GST-DBD II.

The minimum concentration of the GST-DBD II fusion protein required to shift the 670 bp PstI-NdeI probe is 10 nM, the same as that seen with GST-DBD I (Figure 22). The competition assays performed with the unlabeled linear DNA fragments gave similar results as those obtained with the GST-DBD I (data not shown). But the competition experiment with the unlabeled supercoiled constructs showed 100-fold difference between the specific and the non-specific supercoiled DNA (Figure 23).
Figure 21: **Determination of specificity of the GST-DBD I protein by using Poly dI-dC and salmon sperm DNA**: Poly dI-dC and salmon sperm DNA were used as non-specific DNAs to determine the specificity of DNA binding observed for the GST-DBD fusion protein. 10 nM of the fusion protein and 1 nM of radiolabeled DNA fragment were used in each reaction. Lane 1: Unbound 670 bp PstI-NdeI fragment; Lane 2: 670 bp PstI-NdeI fragment + GST-DBD I, no competitor DNA. Lanes 3-6: Salmon sperm DNA at increasing concentrations of 0.05, 0.5, 2.5, 5 and 12.5 μg/ml respectively was used as a competitor DNA. Lanes 7-11: Poly dI-dC at concentrations of 0.05, 0.5, 2.5, 5 and 12.5 μg/ml respectively was added to the reaction. Both Poly dI-dC and salmon sperm DNA compete away the specific binding at 0.5 μg/ml. Lanes 12-15: Total fly DNA at increasing concentrations of 0.05, 0.5, 2.5 and 5 μg/ml respectively was used as a competitor DNA.
Figure 22: The GST-DBD II fusion protein binds DNA with high affinity: Lane 1: Unbound probe; Lanes 2-11: Increasing amounts of fusion protein was added to 1 nM of DNA. Concentrations tested are 1, 5, 10, 20, 30, 40, 50, 75, 100, 250 nM respectively. Similar to the GST-DBD I fusion protein, GST-DBD II binds DNA at 10 nM suggesting that it has high affinity for DNA binding.
Figure 23: Determination of specificity of the GST-DBD II protein by using unlabeled supercoiled DNAs: Supercoiled DNA constructs containing 670 bp PstI-NdeI DNA fragment insert was used as a specific competitor. Supercoiled DNA constructs containing heterologous insert fragments were used as non-specific competitor DNAs. Lane 1: Unbound probe. Lane 2: GST-DBD II + 670 bp PstI-NdeI, no competitor. Lanes 3-6: Specific supercoiled competitor DNA at 0.01, 0.1, 0.5 and 1 nM concentrations respectively. Lanes 7-12: Non-specific supercoiled competitor DNA at 0.01, 0.1, 0.5, 1, 5 and 10 nM concentrations respectively. Lanes 13-17: Another non-specific supercoiled competitor DNA at 0.01, 0.1, 0.5, 1, and 5 nM concentrations respectively. The specific construct competes away the detectable binding at 0.01 nM, whereas the non-specific construct competes away at 1 nM. The difference between the specific and non-specific DNA binding here is 100 fold.
In all these experiments the labeled DNA was the linear 670 bp PstI-NdeI fragment.

The experiments using salmon sperm DNA and the poly dI-dC as non-specific DNA competitors demonstrated that they compete with the "specific" binding at the concentration of 5 \( \mu \text{g/ml} \). This is ten fold higher specificity than seen with the GST-DBD I fusion protein but is still lower than the values seen for the other sequence-specific DNA binding proteins.

The higher specificity seen when the supercoiled DNA was used as a competitor probably suggests that the fusion protein may recognize the structure along with the sequence of the DNA fragment. It also suggests that the fusion protein exhibits "end binding" which is observed when linear fragments are used as competitors. The fusion protein may have sequence-specificity which is masked by this "end-binding" activity.
DISCUSSION:

The *trx*-DBD binds DNA

The electrophoretic mobility shift assay using the 670 bp PstI-NdeI fragment containing the PRE/TRE showed that the GST-DBD fusion protein binds DNA at a very low concentration. The low concentration required for binding DNA suggests that the GST-DBD fusion protein has high affinity for DNA. The GST protein does not bind the 670 bp PstI-NdeI fragment implying that the GST-DBD binds DNA due to the *trx* DBD part of the fusion protein.

The ability of the GST-DBD fusion protein to bind DNA suggests that the *trx* DBD may adopt a structure similar to the nuclear receptor DBD structure which is capable of recognizing DNA. The similarities between the *trx* DBD and the steroid hormone receptor DNA-binding domains and its ability to bind DNA, suggests that this could mediate binding of *trx*. It also suggests that the full length, endogenous *trx* protein associate with specific sites on the polytene chromosomes by binding directly to DNA.
The GST-DBD fusion protein exhibits sequence-specific DNA-binding along with some non-specific binding.

The results of the competition assays using unlabeled specific and non-specific linear DNA fragments suggest that the GST-DBD fusion protein has little or no sequence-specific DNA-binding activity. In contrast the experiments carried out using supercoiled DNA as competitors showed 100-fold difference between specific and non-specific binding. A 100-fold difference between concentrations of supercoiled DNA constructs containing the 670 bp PstI-NdeI insert and the non-specific heterologous DNA insert suggests that the fusion protein exhibits some sequence-specific DNA-binding. The supercoiled DNA does not have free ends and has different structure than the linear fragments used in the other competition experiments.

The non-specific DNA-binding observed when linear fragments were used suggests that, the fusion protein also binds non-specifically, probably to the ends available on linear fragments. It is also possible that the fusion protein recognizes the specific supercoiled structure or some other property of the construct DNA. Thus taken together, these results suggests that the *trx* DBD has two distinct activities, one is sequence-specific DNA-binding and the other is non-specific binding probably to the free ends of DNA fragments. But the experiments described
here do not allow us to conclusively say whether the \( \text{trx} \) DBD binds DNA in a sequence-specific manner \textit{in vitro}.

Further experiments are required to confirm the presence of end-binding and also to conclusively demonstrate sequence-specific binding of the \( \text{trx} \) DBD. The DNA fragment used here as a probe is very large and the fusion protein will bind to some sequence on this fragment non-specifically in addition to the proposed end-binding. EMSA using small overlapping fragments from the 670 bp PstI-NdeI fragment in the presence of supercoiled, competitor DNA can be carried out to identify the smaller fragment containing the specific DNA sequence recognized by the fusion protein. DNase I footprint analysis to determine the DNA sequence recognized by the \( \text{trx} \) DBD fusion protein will confirm if \( \text{trx} \) DBD has sequence-specific DNA-binding activity.

\textbf{The \( \text{trx} \) proteins exhibit sequence specificity \textit{in vivo}}

Using anti-\( \text{trx} \) antibodies we have shown that the \( \text{trx} \) proteins are associated with 63 discrete chromosomal sites. This suggests that their binding is likely to be sequence-specific. If the binding of \( \text{trx} \) proteins was not sequence-specific, then one might expect to see the HRP signal all over the chromosomes including the small fourth chromosome, which does not contain any \( \text{trx} \) binding sites.
The binding of the \textit{trx} proteins to the specific \textit{Ubx} regulatory sequences outside of their normal sequence contexts also suggests that this binding is sequence-specific (Chapter III). The absence of binding to the 22UZ transposon and the 500 bp Styl-PstI transposon along with binding to the 670 bp PstI-NdeI transposon containing the PRE strongly indicates that the \textit{trx} binding to DNA seen \textit{in vivo} is sequence-specific.

Thus any non-specific or end-binding activity of the \textit{trx} DBD fusion protein may not be relevant \textit{in vivo}. It is possible that the \textit{trx} proteins are complexing with other proteins in which a more promiscuous high-affinity binding by \textit{trx} alone contributes to net affinity of a complex, while the other proteins contribute the sequence-specificity. One such likely candidate with which \textit{trx} may form a complex is the \textit{ash1} protein, member of the \textit{trx-G} proteins. Co-localization of \textit{trx} and \textit{ash1} at most of their chromosome binding sites raises the possibility that they may bind DNA as a complex (Chapter II).
Our results demonstrate that the *trx* proteins bind to specific sites on the polytene chromosomes including its known targets, the BX-C and ANT-C. One of the *trx* binding site is located in a 670 bp fragment from the 5' regulatory region of the *Ubx* gene. These results indicate that the *trx* proteins exert their effect by binding directly or indirectly to the specific DNA sequences in the target genes.

Do *trx* proteins exhibit sequence-specific DNA binding in *vitro*?

Gel shift assays using the GST-DBD fusion protein and the 670 bp fragment of *Ubx* indicate that the putative *trx* DBD binds DNA with high affinity. The use of different competitor DNA showed that the *trx* DBD may have some sequence-specificity but it also has non-specific DNA-binding. To resolve this issue and to confirm whether *trx* DBD binds DNA in a sequence-specific manner, further experiments are required. As suggested in the Chapter IV, DNaseI footprint analysis using smaller DNA fragment and the fusion protein in presence of supercoiled competitor DNA could be performed. If this gives a nice protected sequence it will confirm the specificity of DNA-binding of the fusion protein.
Are \textit{trx} and \textit{ashl} proteins part of a multiprotein complex?

Co-localization studies indicate that the \textit{trx} and \textit{ashl} proteins are co-localized at many sites on the polytene chromosomes. To test if their binding is dependent on each other, immunostaining of polytene chromosomes from \textit{trx}^l and \textit{ashl} mutant larvae could be carried out. If the binding is mutually affected then reduction in number of binding sites for both proteins may be expected. This will suggest that they are part of a multiprotein complex.

Analysis of the \textit{Ubx} regulatory region for the presence of binding sites of \textit{ashl} similar to that done for \textit{trx} will demonstrate if they are truly co-localized. Immunoprecipitation using either anti-\textit{trx} or anti-\textit{ashl} antibodies will be a direct test for the presence of these two proteins in a complex.

\textbf{Identifying regions of \textit{trx} proteins required for its chromosomal association.}

The \textit{trx} proteins contain novel motifs which are present in proteins from different organisms. We have mutants which truncate or create a point mutation in these regions (David Bailey, personal communication). Immunostaining of the
polytene chromosomes from these mutants will determine if a
given region is necessary for chromosomal binding.

Do the *trx* and *Pc* proteins bind to the same DNA sequence?

Simultaneous localization of *trx* and *Pc* proteins on the
polytene chromosomes showed that the two proteins are co-
localized at 32 sites. The 670 bp fragment of the *Ubx* regulatory
region contains binding sites for the *trx* and *Pc-G* proteins. By
immunostaining the polytene chromosomes from transformants
containing smaller DNA fragments of this 670 bp DNA the *trx*
and *Pc* binding site sequence can be determined. This will also
show if the binding sequence of the *trx* and *Pc-G* proteins can be
separated or not. Experiments such as these will support or rule
out the possibility that the *trx* and *Pc-G* proteins are part of a
complex.
LITERATURE CITED


Evans, R. M. (1988). The steroid and thyroid hormone receptor superfamily. Science 889-895:


potentiates transcriptional activation by the glucocorticoid receptor. EMBO J. 12:4279-4290.


