

AN ANALYSIS OF POLYCOMB GROUP PROTEIN INTERACTIONS

by

MICHAEL STEPHEN KYBA

B.Sc., The University of Alberta, 1991

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Zoology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1998

© Michael Stephen Kyba, 1998

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Zoology
The University of British Columbia
Vancouver, Canada

Date June 25 1998

Abstract

The Polycomb Group (PcG) of proteins are global regulators of transcription. PcG mutants display posterior homeotic transformations, the result of ectopic expression of homeotic selector genes of the Bithorax and Antennapedia Complex, demonstrating that the PcG is required for the repression of target genes outside of their normal spatial boundaries of expression. Coimmunoprecipitation, cofractionation, and colocation on larval salivary gland chromosomes suggest that PcG proteins act through large multimeric complexes formed at their target sites. This thesis is a characterization of the protein interactions that underlie multimeric complex formation. Using the yeast two-hybrid system and an *in vitro* co-affinity precipitation assay, I demonstrate direct interactions between Polycomb (Pc) and Posterior Sex Combs (Psc), and between Psc and polyhomeotic (ph). I also show that Psc, ph, and Asx have self-interacting domains, and perform a detailed analysis of the self-interacting domain of ph. For the most part, these interacting domains are highly conserved between the *Drosophila* proteins and their mammalian counterparts. Because Asx shows no direct interactions with Pc, Psc, or ph, I screen Asx for interacting proteins within a two-hybrid library and within a two-hybrid panel of other chromatin proteins. Several interactors are identified, including the *Drosophila* homologue of cyclin G, and z40, a previously unknown protein which interacts strongly with Pc. In addition, an interaction is demonstrated between the respective carboxyl termini of Asx and trithorax (trx), a protein required for activation of homeotic selector genes. I show that Psc can repress transcription in *Saccharomyces cerevisiae*, and show that this repression does not require interactions with a variety of yeast proteins required for repression of various loci in the *S. cerevisiae* genome. These data enlarge our understanding of the structure of PcG complexes, and suggest that PcG proteins interact with one another promiscuously, enabling them, in theory, to form a large number of different complexes each tailored to a particular chromosomal neighbourhood.

Table of Contents

Abstract	ii
List of Figures	vii
Acknowledgments	ix
Introduction	1
The Polycomb Group	2
Cis Elements Recognized by the PcG	4
The Trithorax Group	5
The Mechanism of PcG Silencing or Repression	6
Other Silencing and Repression Systems	8
Chapter I: Interacting Domains of Asx, Pc, Psc, and ph.	12
Sequence Motifs Present in Asx, ph, Psc, and Pc.	13
Co-immunoprecipitation of Pc, ph, and Psc	14
Two-Hybrid Interactions	19
In-Vitro Interactions	27
Ternary vs. Binary Complexes	31
Isolated Domain Interactions can be Modulated by External Sequences	32
Interactions of the Vertebrate Homologues of Psc, Pc, and ph	35
The Role of Multiple Interacting Domains in PcG Complexes	36

Table of Contents, cont.

Chapter II: The SAM Domain.	38
The SAM Domain	38
The ph SAM Domain Mediates Self-Association In Vitro	39
Homologues of the ph SAM Domain	42
Homotypic and Heterotypic Self-Association of ph SAM	
Homologues	43
Mutations in Conserved Residues Have Different Effects on	
Binding to Different SAMs	49
SAM: A Self Association Motif	49
Key Structural and Functional Residues of the SAM Domain	52
The Potential for Promiscuous Oligomerization	53
 Chapter III: Asx Interactions	 55
A Conserved Domain in the C-terminus of Asx	55
The Asx Two Hybrid Library Screen	58
Interactions with Other PcG Proteins	60
Evaluation of AsxC Interacting Proteins	60
Possible Roles for Bowel and Cabeza	62
The Complete Sequences of z34 and z40	63
z34 Contains a Cyclin Box	63
An Asx-Cyclin G interaction?	73
The Asx-z40 Interaction	76
Asx Interacts with the trx SET Domain	76
The Transcriptional Consequence of an Asx-trx Interaction	82

Table of Contents, cont.

Chapter IV: PcG Functional Interactions in Yeast	85
Pc, Psc, and ph Can Repress Transcription Directly	85
Psc is a Transcriptional Repressor In Yeast	88
Telomeric Effects of the PcG and trxG in Yeast	94
Implications for the Mechanism of PcG Action	98
 Chapter V: Materials and Methods	99
Subcloning	99
Mutagenesis	102
Primers	103
Sequencing	105
GST-fusion protein expression and purification	106
In vitro coprecipitations	106
Binding to Ni-NTA agarose	107
Transformation and Culturing of Yeast Strains	108
Yeast Strains	108
Two-Hybrid Interaction Assays	109
β -galactosidase assays	110
Co-immunoprecipitation from Kc nuclear extracts	110
SL2 Transfections	111
CAT assay	111
Telomeric Variegation Assays	112

Table of Contents, cont.

Chapter VI: Conclusion	113
The Assembly of PcG Complexes	113
Dimerization	116
Silencing or Repression?	119
The Large Membership of the PcG	120
 Nomenclature	 124
 Bibliography	 127
 Appendix A: Interactor Preliminary Sequences	 142
 Appendix B: Interactor Sequence Comparisons	 151

List of Figures

Chapter I

Figure 1.1	Sequence motifs of the four PcG proteins studied.	15
Figure 1.2	Pc, ph, and Psc proteins coimmunoprecipitate.	17
Figure 1.3	Two-hybrid interaction assay results for ph, Psc, Pc, and carboxyl deletions.	21
Figure 1.4	Asx interactions.	23
Figure 1.5	Two-hybrid interaction assay results for conserved sequence constructs.	25
Figure 1.6	In vitro binding of reticulocyte lysate-generated ³⁵ S-labeled Psc constructs to bacterially produced GST-fusions.	29
Figure 1.7	Domains involved in interactions between Pc, ph, and Psc.	33

Chapter II

Figure 2.1	Self-binding activity of the carboxyl terminus of ph.	40
Figure 2.2	Alignment of relatives of the ph SAM domain.	44
Figure 2.3	Homotypic and heterotypic binding of various SAM domains.	46
Figure 2.4	Mutations in pSAM and their effects on binding.	50

Chapter III

Figure 3.1.	hAsx sequence.	56
Figure 3.2	Sequence of the z34 cDNA.	64
Figure 3.3	Sequence of the z40 cDNA.	67
Figure 3.4	z34 Sequence Alignments.	69
Figure 3.5	Asx-trx interactions.	79

Figures, cont.

Chapter IV

Figure 4.1	Transcriptional repression by PcG proteins in transiently transfected SL2 cells.	86
Figure 4.2	Transcriptional repression by Lex-Psc.	89
Figure 4.3	Transcriptional repression by Lex-Psc in various mutant backgrounds.	91
Figure 4.4	Telomeric silencing effects of trx sequences.	96

Conclusion

Figure 6.1	A model for regulated PcG complex assembly.	114
Figure 6.2	Self-association of PcG proteins may facilitate inter-homologue interactions	117
Figure 6.3	Different loci require different PcG proteins for silencing	121

Acknowledgments

Thanks are due to many people:

Dr. Hugh Brock, for allowing me great freedom throughout his supervision of this work.

Members of the Brock Lab, especially Dr. Jacob Hodgson, for much good advice, many stimulating discussions and encouragement.

Wilfred Lim for help with the two-hybrid library screen.

Michael O'Grady for his two-hybrid panel of chromatin proteins.

Erica Golemis, Fred Winston, Jasper Rine, Ira Herskowitz, Carrie Brachmann, Randall Mann, and David Stillman for providing yeast strains, Paul Adler, Rick Jones, Yasushi Matsui, Renato Paro, Ivan Sadowski, Rob Saint, Jeff Simon, Kazunori Shimada, and Don Sinclair for providing cDNAs, and Paul Adler and Jacob Hodgson for providing antibodies.

The National Science and Engineering Research Council of Canada for funding part of this work.

Mum and Dad for putting food on the table when my pockets were empty, and for their great moral and spiritual support.

Introduction

The central dogma of molecular biology postulates that information flows from DNA to RNA to protein. The relative rate at which different segments of information (genes) are transcribed through the first step in this progression is a critical element, sometimes the only necessary element, in defining a given biological state. Likewise, rate changes are linked to biological state changes: ontogeny, differentiation, response to external stimuli, and neoplasia, to name some of the more important ones. The understanding of transcriptional regulation, then, is of crucial importance to the understanding of biological phenomena.

If one considers the rate of transcription to be governed by a balance of positive and negative factors, then to date our understanding of the positive factors far outstrips our understanding of the negative factors. This is partly due to the relative diversity of positive factors, partly due to the existence of lower thresholds that can be crossed by the removal of positive factors by mutation, but a relative paucity of upper thresholds that can be crossed by the removal of negative factors, and partly due to the logical requirement for a functioning *in vitro* system of transcription to exist (the functioning of which presumes the contribution of certain positive factors) before one can study negative transcriptional factors *in vitro*. This biased apprehension has traditionally contributed to a view of transcription that overvalues the role of positive factors, and ignores or undervalues the role of negative factors. However, recent advances have begun to bring the role of negative factors into focus. This thesis addresses the less-charted negative side of transcription through a molecular study of one group of negative transcriptional regulators, the Polycomb group (PcG).

The Polycomb Group

Restricting the spatial and temporal expression of the transcription factors that control development is a central part of the mechanism by which pattern and structure come into being in a developing organism. In *Drosophila*, cellular identity along the anterior-posterior (AP) axis is specified by the homeotic selector genes of the Antennapedia and Bithorax Complexes (AntC, and BxC, respectively) [1; 2]. Particular patterns of homeotic gene expression are thought to define particular positions along the AP axis, and changes in these expression levels in a given axial segment cause corresponding changes of fate for that segment. The domains of expression of the homeotic selector genes are set up by gap and pair-rule genes which act early in development [3]. The proteins of the PcG maintain the fidelity of these domains of expression through later divisions when the early regulators are no longer present [4].

The PcG was originally identified in *Drosophila* as a group of genes whose mutation causes multiple homeotic transformations similar to gain-of-function mutations of the homeotic selector genes. This similarity is the result of derepression of the homeotic genes outside of their proper spatial boundaries [1; 5-8]. Because derepression in PcG mutants characteristically occurs between 5 and 6 hours of development, before which time selector gene expression is normal in both degree and pattern, the PcG is thought to be required for maintenance of the repressed state, but not for initiation of the repression [8-10].

There are 14 genetically characterized members of the PcG: *Polycomb* (*Pc*) [1], *extra sex combs* (*esc*) [11], *Polycomblike* (*Pcl*) [12], *Enhancer of Polycomb* (*E(Pc)*) [13], *super sex combs* (*sxc*) [14], *polyhomeotic* (*ph*) [15], *Sex combs on midleg* (*Scm*) [16] *Sex combs extra* (*Sce*) [17], *Posterior sex combs* (*Psc*) [18], *Additional sex combs* (*Asx*) [19], *Suppressor two of zeste* (*Su(z)2*) [20], *Enhancer of zeste* (*E(z)*) [10], *pleiohomeotic* (*pho*) [21], *multi sex combs* (*mxs*) [22], and *cramped* (*crm*) [23]. Each member of this group is

required for repression of one or more of the homeotic selector genes outside of their proper domains of expression (although in the case of *Su(z)2* and *E(Pc)* this requirement is minimal except in a genetically sensitized background, see below). The other hallmark of PcG mutants is that they display dominant enhancement [16; 20; 24; 25], and in some cases antipodal suppression [24; 25 ; 26; 27] of each other's phenotypes, implying that the function of the group as a whole is sensitive to the dosages of its members [16; 28]. However such interactions are not seen for every mutant combination, nor for every phenotype, and in many cases interactions are allele-specific [25]. With respect to this thesis, some details of these genetic interactions are worth pointing out. When all possible PcG (with the exception of *ph*) double heterozygotes were generated, *Asx*, *Pc*, *Pcl*, *Psc*, *Sce* and *Scm* enhanced each other's adult homeotic phenotypes in every pairwise combination (with the exception of *Psc/Asx* and *Psc/Sce*) [25]. Double heterozygous combinations of *ph*⁵⁰³, a null, with *Pc*⁴, *Pc*¹⁶, or *Su(z)2*¹ were lethal, with *Psc*¹ strongly semilethal, while with other PcG genes, non lethal [24]. Combinations hemizygous for *ph*⁴⁰⁹, a strong hypomorph, and heterozygous for *Pc*¹, *E(z)*¹, *Psc*¹, *Scm*, and *Sce*, were lethal, while other PcG genes did not interact lethally with *ph*⁴⁰⁹ [24]. Thus *ph*, *Pc*, and *Psc* form a strongly interacting leash within a larger set of mutually interacting elements, which includes *Pcl*, *Scm*, *Sce*, and *Asx*. Other PcG genes tend to interact more sporadically with members of this set and with each other.

A model that explains this genetic synergism postulates that the PcG acts as a multimeric protein complex with phenotypic enhancement being the result of increased perturbation of the complex with an increased number of mutant members [28]. All PcG proteins that have been tested immunohistochemically are present at specific sites along polytene chromosomes. *Pc*, *Pcl*, and *ph* bind the same approximately 100 sites [29; 30]. *E(z)* shares most of these sites but binds to unique sites of its own [31], as do *Psc* and *Su(z)2* [32; 33]. Colocation on polytene chromosomes, co-immunoprecipitation (co-IP) of *ph* and

Pc [29], as well as the presence of approximately 10 unidentified proteins in the Pc immunoprecipitate [29] support the idea that large, multimeric PcG protein complexes reside at various sites in the chromatin of *Drosophila*. However the fact that many loci stain for some members but not others, as well as the fact that different PcG mutants display different levels of selector gene derepression [8] suggests that if a complex exists it must be heterogeneous.

Mammalian Hox gene expression boundaries appear to be maintained through a mechanism similar to that seen in *Drosophila*, via mammalian homologues of PcG proteins. Targeted gene replacements of the *Psc* homologues *Bmi-1* [34] and *Mel-18* [35], the *ph* homologue *rae28* [36], and the *Pc* homologue *M33* [37] cause posterior transformations of the axial skeleton, due to the anterior shift of several Hox gene expression boundaries.

Overexpression of *Mel-18* in transgenic mice confers the opposite phenotype [38]. The surprising result that an *M33* transgene partially rescues a *Pc* mutation in *Drosophila* demonstrates that there has been remarkable conservation of the mechanism of PcG function between flies and mammals [39].

Cis Elements Recognized by the PcG

DNA sequences that restrict or reduce the expression of reporter genes in a PcG-dependent manner (PREs, for PcG Response Elements) have been found in the regulatory regions of the homeotics [40-44] as well as upstream of *ph* [45] and *engrailed (en)* [46]. Although DNA crosslinking experiments suggest that PcG proteins interact with DNA a few kb upstream and downstream in addition to the PRE [47], the vast distances between regulatory elements of the homeotics evoke an image of PcG complexes binding discretely to a few sites along the BxC and AntC. It was noticed that expression of the *white* gene, an eye colour marker used to score for P-element-mediated transformation, was suppressed when PRE-containing transgenes were made homozygous, and that this suppression was

dependent on the PcG [43; 45; 46] Such pairing-sensitive repression suggests that the PREs on homologous chromosomes interact with each other via PcG protein interactions.

By juxtaposing PREs with other *cis* elements, it has been shown that a PRE can block nearby binding of other DNA-binding proteins, and that this blockage is competitive, being overcome by higher expression levels of the protein being blocked [48]. Other experiments have shown that insulating sequences (*gypsy* and *scs*) interposed between the PRE and a reporter gene can block silencing [49]. Gypsy elements surrounding a PRE-reporter construct prevent silencing at some insertion sites, provided that the transgene is heterozygous, showing that silencing involves interactions with other sequences, perhaps other PREs in the vicinity, on the homologous chromosome, or even on non-homologous chromosomes [49]. The phenomenon of homing, whereby a transgene carrying a PRE inserts preferentially very near to an endogenous PRE [45], is also suggestive of promiscuous, interlocus PRE interactions.

The Trithorax Group

Mutations in a group of 12 other genes (collectively named the trithorax Group, *trxG*) suppress the homeotic transformations of the PcG [50]. The *trxG* is also required for proper expression patterns of the homeotics, and in many ways behaves similarly, although in the opposite direction, to the PcG. *trxG* mutations cause transformations similar to loss of function mutations of the homeotics [51], the result of reduced expression of the homeotics, with different homeotics showing different sensitivities to a given mutant [52]. Double heterozygotes for some pairs *trxG* mutants show enhanced homeotic phenotypes relative to single heterozygotes [53].

The *trxG* protein *brahma* (*brm*) contains six blocks of sequence similar to those found in the yeast general transcriptional activator, SWI2, a DNA-dependent ATPase [54]. SWI2 is

part of a multiprotein complex (the SWI/SNF complex) that enables transcription by relieving chromatin-mediated repression [55; 56]. Another trxG protein, *snr1*, is the homologue of SNF5 (also a member of the SWI/SNF complex) and is present with *brm* in a megadalton complex in *Drosophila* nuclear extracts [57]. *Trithoraxlike* (*Trl*) encodes GAGA factor [58], a protein required for transcriptional activation of many genes, including *hsp70* [59]. *In vitro*, GAGA factor can relieve repression by histone H1 [60], and can, in the presence of the ATP-dependent nucleosome remodeling factor (NURF), alter the chromatin structure of the *hsp70* promoter [61].

From these examples, it would appear that the function of the trxG is to enable transcription by altering the chromatin structure of a targeted locus. That is likely the case for the examples given above, however the mechanism of action of other trxG proteins is less clear. Interestingly, several trxG proteins share sequence motifs with PcG proteins. *E(z)*, *ash1* and *trx* contain SET domains [62]. *Pcl*, *ash1* and *ash2*, contain PHD fingers [62; 63]. This sequence homology may mean that certain PcG and trxG proteins compete for the same factors, a scenario that would explain the suppression of PcG haplo insufficient phenotypes by trxG mutations.

The Mechanism of PcG Silencing or Repression

Based on sequence similarity between Pc and HP1, a heterochromatin protein [64], Renato Paro has suggested that the PcG may organize the genes that they regulate into compact higher order heterochromatin-like structures that are inaccessible to RNA polymerase II (pol II) and transcriptional activators [65; 66]. Thanks to the trxG findings mentioned above, and a possible link to suppressors of position effect (mentioned below), heterochromatin has figured heavily into current thinking of PcG function. It is worth pointing out that the chromatin state of GAGA factor-, or SWI/SNF-regulated genes is not heterochromatic, even in the absence of transcription. Biochemical evidence of

heterochromatic changes induced by PcG proteins has yet to be presented, worse there is an absence of such change at the BxC in *Pc* mutants [67].

Vincenzo Pirrotta has suggested a model based on cooperative binding, whereby weak individual binding sites are distributed at strategic points along a transcriptional domain, and a PRE, which consists of multiple binding sites clustered together, serves as a nucleation center, looping in the single sites from adjacent DNA only when a complex of sufficient size has assembled at the PRE [68; 69]. The looping would prevent enhancers from interacting with the promoter of the gene that they regulate. Determination of DNA-binding activity of PcG proteins or interacting factors, and the fine scale mapping of the sites they recognize along the BxC will be necessary to judge the validity of this model.

Other models have been proposed. One PcG protein, *esc*, has been suggested to interact with and inactivate the basal transcription complex [70], although as yet there is no biochemical evidence for this suggestion. Subnuclear localization to a silent compartment has been suggested [67]. The evidence for this is shaky at present: *Pc*, *ph*, and *Psc* are present at multiple nuclear foci in interphase cells [71; 72]. When overexpressed in transfected cells, the mammalian *Psc* homologue Bmi-1 is visible at the nuclear periphery [72], however other studies of endogenous Bmi-1 and other mammalian PcG proteins report a speckled distribution [73]. If a silent PcG compartment exists, it is most likely distributed through the nucleus, thus rendering questionable the term "compartment".

Whether a single model can explain the biochemical function of the PcG has not really been questioned. However, early opinion held that at least two genes, *Pc* and *esc*, must have independent functions. This was based on the fact that the phenotype of individuals with a complete absence of *esc* (maternal⁻) was made more severe by reducing the dose of *Pc* [74], and by the fact that *esc* was required earlier than *Pc* [5]. Notwithstanding the genetic

interaction studies cited above, this result precludes a single model based on the formation of a complex with a threshold concentration, unless one assumes that there is redundancy built into the complex (i.e. that some members can substitute for one another).

Other Silencing and Repression Systems

Drosophila has two other classical systems that monitor transcriptional silencing or repression: Position Effect Variegation (PEV) and transvection. PEV is the stochastic, clonally inherited silencing of a gene that has been brought into proximity to heterochromatin by chromosomal rearrangement. This proximity is thought to convert the transcriptional domain into heterochromatin in some cells and their descendants. Mutations of chromatin modifiers or assembly factors are thought to enhance or suppress the frequency of this incorporation [75; 76]. *Trl* is an enhancer of variegation (E(var)) [58], while E(Pc) and Asx are suppressors of variegation (Su(var)s) [77]. Extra copies of the human homologue of *E(z)* are E(var)s [78]. Other PcG genes have not convincingly been shown to affect PEV. These data would lend strength to the idea that at least some PcG proteins are heterochromatin factors, except for the inability to determine whether the modifier effect is direct, or the result of derepression of a true modifier.

Transvection is the pairing-sensitive genetic interaction of homologous loci [79]. It is not a system of repression *per se*, however transvection-dependent repression of the *white* gene has been noted in a *zeste*¹ mutant background. Modifiers of this repression are thought to encode euchromatic chromatin proteins that influence inter homologue interactions.

Transvection has been shown to occur at the BxC [80], and interestingly, several PcG genes are modifiers of the *zeste-white* interaction: *Psc*, *Su(z)2*, *E(z)*, and *Scm*, [81-83]. It is tempting to speculate that the same PcG modifiers of transvection could be involved in crosstalk at the BxC, perhaps between PREs.

In the yeast *Saccharomyces cerevisiae*, two classes of loci, telomeres and the silent mating type loci, are kept transcriptionally inert through a mechanism involving a compact chromatin state. This type of permanent transcriptional inhibition is referred to as silencing, in contradistinction to regulatable transcriptional inhibition, or repression. Silencing at both telomeres and the silent mating type loci requires the action of Sir2, 3, and 4 (Sir, Silent Information Regulator) [84; 85], histones, and assorted other factors [86]. Sir3 and 4 are recruited by sequence-specific factors, and interact with the N-terminal tails of histones H3 and H4 [87], setting up a compact structural configuration that spreads some distance along the chromosome. The stochastic silencing of genes near telomeres is strongly reminiscent of PEV.

The conditional repression of many inducible genes in *S. cerevisiae* depends on the action of Ssn6 and Tup1. A pentamer containing four molecules of Tup1 and one molecule of Ssn6 [88] is recruited by a variety of sequence-specific DNA-binding factors including the $\alpha 2$ repressor, required for repression of a-specific genes [89], and Mig1, required for catabolite repression [90]. The mechanism by which Ssn6-Tup1 represses transcription is unclear, but may involve the positioning of nucleosomes. Deletion of *SSN6* or *TUP1* disrupts the positioned nucleosomes of the $\alpha 2$ operator, independently of whether transcription is occurring at the locus [91]. Tup1 interacts directly with the N-terminal tails of histones H3 and H4, and these interactions are strongest when the histones are underacetylated [92]. On the other hand, Tup1/Ssn6 repression can be seen in an in vitro system lacking nucleosomes, suggesting an interaction with the basal transcriptional machinery itself [93].

The study of a large number of factors identified through different repression assay systems in yeast has converged on the carboxyl terminal repeat domain of pol II (CTD) as being a target of inducible transcriptional repression. Many members of the protein

complex associated with the CTD (named the mediator [94; 95], for mediator of transcriptional activation) can be mutated to confer repression-insensitivity to pol II. The inducible genes in which these repression-defective mutants were originally identified include *SUC2* (*ssn* mutants), *HO* (*sin* mutants), Ty insertions (*spt* mutants), and *CYC7* (*rox* mutants) (reviewed in [96]). Repression in some of these systems also requires Ssn6/Tup1.

Understanding the relationship of the PcG to the well defined transcriptional repression and silencing systems of yeast will be helpful in piecing together the mechanism by which the PcG mediates transcriptional repression. Likewise, understanding the nature of the involvement of the PcG in PEV and transvection ought to shed light on its action at the BxC and AntC. Protein-protein interactions figure centrally in each of the other systems of transcriptional repression or silencing described above. Studying the protein interactions within the PcG then, is relevant not only to the structure of the PcG complex(es) but also to their function. This is especially true given the behaviour of the *cis* elements with which the PcG proteins interact, namely their PcG-dependent interactions with PREs on homologous chromosomes, and with other *cis* elements on the same or other chromosomes, and the influence that these interactions have on repression.

A clear understanding of how the PcG proteins carry out their function will ultimately require knowing not only the composition and structure of protein complexes that they form, but how these complexes are assembled, and the factors (DNA, chromatin, transcriptional activators, basal transcription factors, replication factors, or others) with which different members of the PcG interact.

My approach has been two-pronged: the first objective was to uncover and characterize interactions between known and cloned PcG proteins using the yeast two-hybrid system

and *in vitro* co-affinity precipitation. The second objective was to discover other interacting factors through a random screen of a two-hybrid library, a rational screen of a panel of chromatin proteins, and by assaying PcG proteins for function in a heterologous environment, and studying the requirement for other factors known play a role in silencing or repression in that heterologous environment.

Chapter I: Interacting Domains of Asx, Pc, Psc, and ph.

As a first step towards understanding the structure of PcG complexes, I tested four PcG proteins, Asx, ph, Psc, and Pc, for interactions with each other, and identified domains essential for these interactions. At the initiation of this work, there was a complete lack of knowledge about PcG intra-complex physical interactions. The current knowledge of these interactions in *Drosophila* is still very sparse. Temperature shift experiments with a temperature-sensitive allele of *E(z)* suggested that *in vivo* binding of ph, Psc, and Su(z)2 to most but not all of their sites on salivary gland chromosomes was dependent on *E(z)* protein [32]. This may mean that *E(z)* plays a role in targeting or fixing certain PcG complexes to their sites. An *E(z)*-esc interaction has been detected by Jeff Simon using the two-hybrid system (personal communication). Somewhat more work has been done with vertebrate PcG homologues. A possible ph-Psc interaction is suggested by recent two-hybrid experiments with the mouse homologues of these proteins, Mph1 and Bmi-1 respectively [73]. The C-terminal 292 amino acids of Mph1 interacted with Bmi-1, and a 220 amino acid putative helical domain of Bmi-1 interacted with Mph1. However these two domains were not tested against each other, so it is not known whether they interact with each other or with other parts of their respective proteins. The human homologues of ph, HPH1 and HPH2, were recently cloned using *Xenopus* Bmi-1 as bait in the two hybrid system [97]. This interaction was delimited to a 295 amino acid C-terminal fragment of HPH2. The conserved amino-terminal 188 amino acids of the *Xenopus* homologue of Psc, XBmi-1, has been shown to bind to the *Xenopus* homologue of Pc, XPc, and Xpc has been shown to bind to itself [98]. The mouse homologues of Psc and Su(z)2, Bmi-1 and Mel-18 have been shown to coimmunoprecipitate with the mouse homologue of ph, Mph, and the mouse homologue of Pc, M33 [73], (N. Hashimoto, H.W. Brock, M. Nomura, M. Kyba, J. Hodgson, Y. Fujita, Y. Takihara, K. Shimada, and T. Higashinakagawa, submitted).

Sequence Motifs Present in Asx, ph, Psc, and Pc

From the point of view of domain analysis, the four proteins studied have several interesting features. *ph* is a tandemly duplicated gene with the proximal and distal transcription units coding for two nearly identical proteins of 167 and 149 kDa. The proximal *ph* product has 193 amino-terminal amino acids that are absent from distal *ph*, and in addition makes use of internal initiation to give an alternate product shorter by 244 amino acids [99]. A notable feature of this unique proximal domain is the presence of a PxxPxxPxxP motif (aa 156-165) with proline spacing the same as that of the polyproline type-II helix recognized by the SH3 domain [100]. *ph* also has many glutamine repeats and a serine/threonine rich region. Near the carboxyl terminus are two blocks of sequence (aa 1297-1388 and aa 1511-1576) that are shared with the mammalian *ph* homologues [73; 97; 101]. The first sequence, named H1 consists of 28 highly conserved amino acids followed by an unusual C4 zinc finger with intercysteine spacing CX₂C...CX₃C. The second sequence has been variously referred to as H2 [101] or SEP [73] in the mouse homologue, SPM in the PcG protein Scm [102] as well as in the human *ph* homologues, HPH1 and HPH2 [97] and SAM in a variety of yeast signal transduction proteins [103]. I have shown that this domain can mediate homotypic and heterotypic self-association between *ph* and Scm proteins *in vitro* (Chapter 2). In view of this result, I refer to the domain in general as a Self Association Motif, and keep the acronym SAM, but refer to the specific subset of SAMs with greatest similarity to *ph* and Scm as SPM. The only internal region of sequence dissimilarity between proximal and distal *ph* are the 52 amino acids immediately preceding the SPM domain. This work (with the exception of chapter 2) has exclusively used the proximal isoform of *ph*.

Psc is a 170 kDa protein with several stretches of repeated amino acids. Strong similarity to amino acids 261-467 of Psc has been found in the *Drosophila* PcG protein Su(z)2 and the mammalian homologues Bmi-1 [104; 105], and Mel-18 [106]. This block of conserved

sequence includes a potential C2HC3 ring finger at the amino end and a putative helix-turn-helix (HTH) motif at the carboxyl. Interestingly, another *Drosophila* homologue has been cloned which consists of these two conserved sequences and nothing else [107].

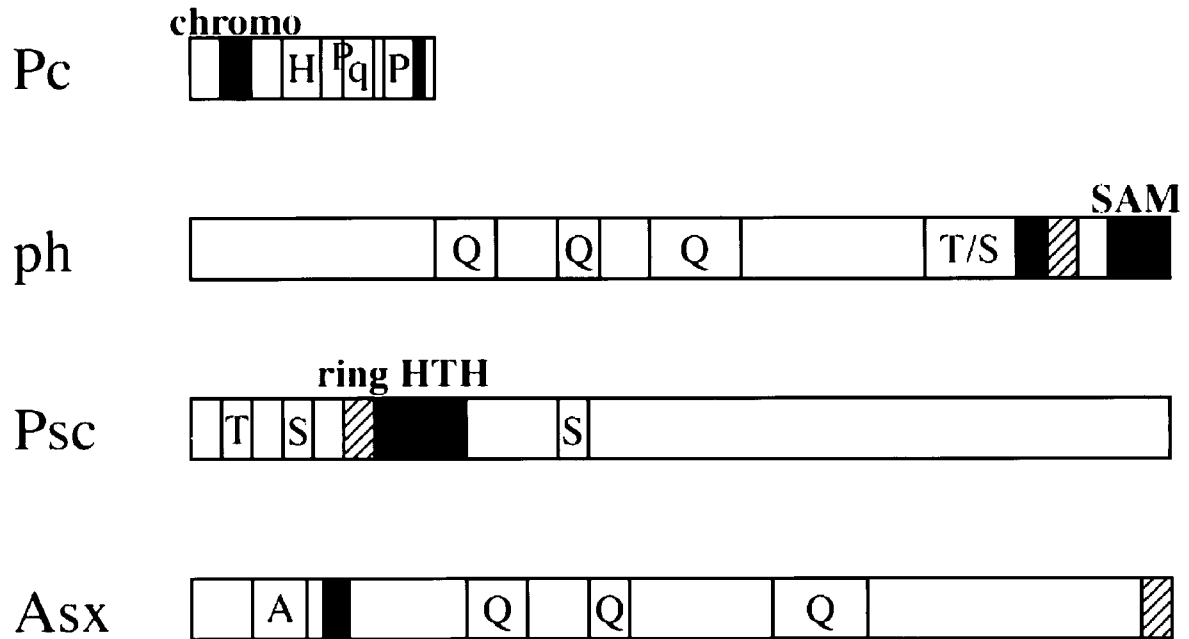
Pc is a 44kDa protein with two histidine repeats and two proline rich regions, the first of which partly overlaps with interspersed glutamine repeats. Amino acids 26-62 of Pc are conserved with HP1, a *Drosophila* heterochromatin protein, and the mammalian protein M33, and have been named the 'chromobox'. [64; 108] In addition, Pc and M33 share a short sequence near their respective carboxyl termini.

Asx is a 182 kDa protein. Like ph, it has extensive glutamine repeats. It also has a run of 20 alanines near the amino terminus. At the extreme carboxyl terminus is a cluster of cysteines with spacing C-x-C-x₇-C-x₂-C-x₃-C-H-x₂-C-x₆-C-x₂-C, which could contain a zinc finger. There are two domains that have a high degree of sequence conservation with mammalian ESTs (expressed sequence tags): the putative zinc finger just mentioned (and described in more detail in Chapter 3), and a stretch of sequence from aa 201-318. This latter sequence similarity was not known at the time that these experiments were initiated.

Co-immunoprecipitation of Pc, ph, and Psc

Pc and ph have previously been shown to colocalize on polytene chromosomes and to immunoprecipitate with each other as well as at least 10 unidentified proteins [29]. Given the high level of overlap between the polytene chromosome binding sites of these two proteins with Psc [32; 33] as well as the coimmunoprecipitation of the mammalian homologues of all three proteins [73] (N. Hashimoto, H.W. Brock, M. Nomura, M. Kyba, J. Hodgson, Y. Fujita, Y. Takihara, K. Shimada, and T. Higashinakagawa, submitted), it seemed likely that Psc would complex with ph and Pc *in vivo*. I therefore performed an immunoprecipitation of a nuclear extract with an antibody to Pc. As shown in Figure 1.2, ph and Psc proteins are both present in the immunoprecipitate of the Pc

Sequence Features



Scale: |-----| 1000 aa




-  Zinc or Ring Finger
-  Mammalian Homology
-  Amino Acid Repeats

Figure 1.1

Sequence motifs of the four PcG proteins studied. Regions of sequence conservation with mammalian homologues are marked black. Zinc-finger and ring-fingers are striped, and regions containing a predominance of a particular amino acid are shaded and labeled with the one letter designation for that amino acid. SAM: self-association motif, chromo: chromobox, ring: ring finger, HTH: helix-turn-helix

IP Antibody:

Pc pi

Pc pi

212—

156—

114—

97—

66—

55—

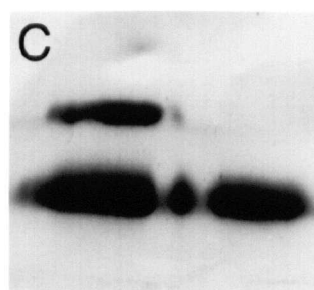
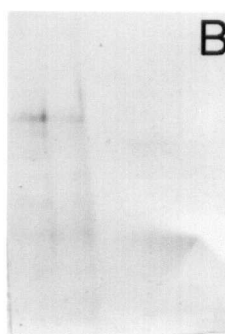
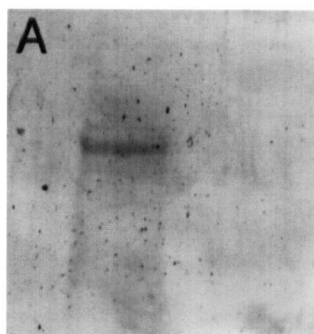


Figure 1.2

Pc, ph, and Psc proteins coimmunoprecipitate. The nuclear extract immunoprecipitate of a Pc antibody and its cognate pre immune serum were electrophoresed in two lanes each, and electrophoretically transferred to a nitrocellulose filter. The filter was then cut into three pieces, each probed with a different antibody. The reconstructed filter is shown where (A) is the part probed with ph antibody, (B) is the part probed with Psc antibody, and (C) is the part probed with Pc antibody. All three proteins are present in the Pc IP, but not in the preimmune IP. The large band at 55 kDa is the IgG heavy chain of the immunoprecipitating antibody, which reacts with the secondary antibodies.

antibody, but not present in the immunoprecipitate of the preimmune serum. Asx could not be tested due to the lack of an antibody. Very recently, the two reciprocal immunoprecipitations have been performed by Strutt and Paro [109], who show that the ph immunoprecipitate contains Psc, and the Psc immunoprecipitate contains ph, completing the circle of interactions.

Two-Hybrid Interactions

To identify potential direct contacts between the ph, Pc, and Psc proteins, I generated DNA-binding and activator fusions to all three proteins and carboxyl deletion derivatives and tested them for interaction in the yeast two hybrid system [110]. All possible pairwise combinations were tested. Shown in Figure 1.3 are the most informative pairs. All pairs not shown were negative. Three interacting combinations were detected: Psc-Pc (Figure 1.3a), ph-ph (Figure 1.3b), and ph-Psc (Figure 1.3d). There were no self-interactions seen for either Pc or Psc (Figure 1.3c). The deletion derivatives locate a ph-ph interacting region in the amino terminal 522 amino acids, although in Chapter 2, I demonstrate that ph also has a carboxyl terminal self-interacting domain. The ph-Psc interacting domains were mapped to between amino acids 523 and 1418 of ph, and amino acids 205 to 696 of Psc. This interaction occurred with deleted versions of each protein, but not with full length. The Psc-Pc interaction also mapped to amino acids 205 to 696 of Psc, and was similarly not observed with full length Psc.

Three Asx constructs were generated and tested for two-hybrid interactions. An amino construct, AsxA (aa 1-335), a construct from the glutamine-rich central portion of Asx, AsxQ (aa 611-1138), and a carboxyl construct, AsxC (aa 1139-1668). All three constructs were tested against the entire previous panel, in both DNA-binding and activator fusion combinations. No interactions were seen between any Asx construct and any ph, Psc, or

Pc construct. When the Asx constructs were tested against themselves, the only interaction detected was between with the AsxC construct, which interacted with itself (Figure 1.4).

To better define the ph-Psc-Pc interactions, I generated a set of smaller constructs. Because all of the interactions mapped to areas that contained sequence similarity to mammalian homologues, these sequences alone were tested against each other and against the previous panel of constructs (Figure 1.5).

The smallest fragment of ph to interact with Psc was the H1 domain, amino acids 1297-1418. The minimal Psc element required for the same interaction was the HTH fragment, amino acids 336-473. The minimal domains interacted with each other and are therefore sufficient. DNA-binding fusions to both phHD (amino acids 1297-1576) and the subfragment H1 activated transcription alone as assayed by their ability to promote growth on leucine deficient medium in the absence of any other plasmid. It was therefore impossible to test these domains reciprocally. However by using the ph Δ N construct (amino acids 1-1417), which contains the H1 domain and does not activate transcription alone, I could demonstrate reciprocity for the HTH domain of Psc. An interesting modulating effect was noted with the SPM domain of ph (amino acids 1511-1576): when the SPM domain was present in a construct, the interaction with Psc was weaker, or as in Figure 1.4d absent.

The domain of Pc required for the interaction with Psc was shown to reside in the 320 amino acids C-terminal to the chromobox (Figure 1.5c, referred to as Δ chrPc).

Surprisingly, the chromobox was not required for this interaction, nor did it or Δ chrPc show interactions with any Pc construct or with any ph construct from the panel (not shown). The Psc domain required for the Pc interaction was also located within the region of amino acid conservation. Minimally, the HTH domain showed interaction with Pc as

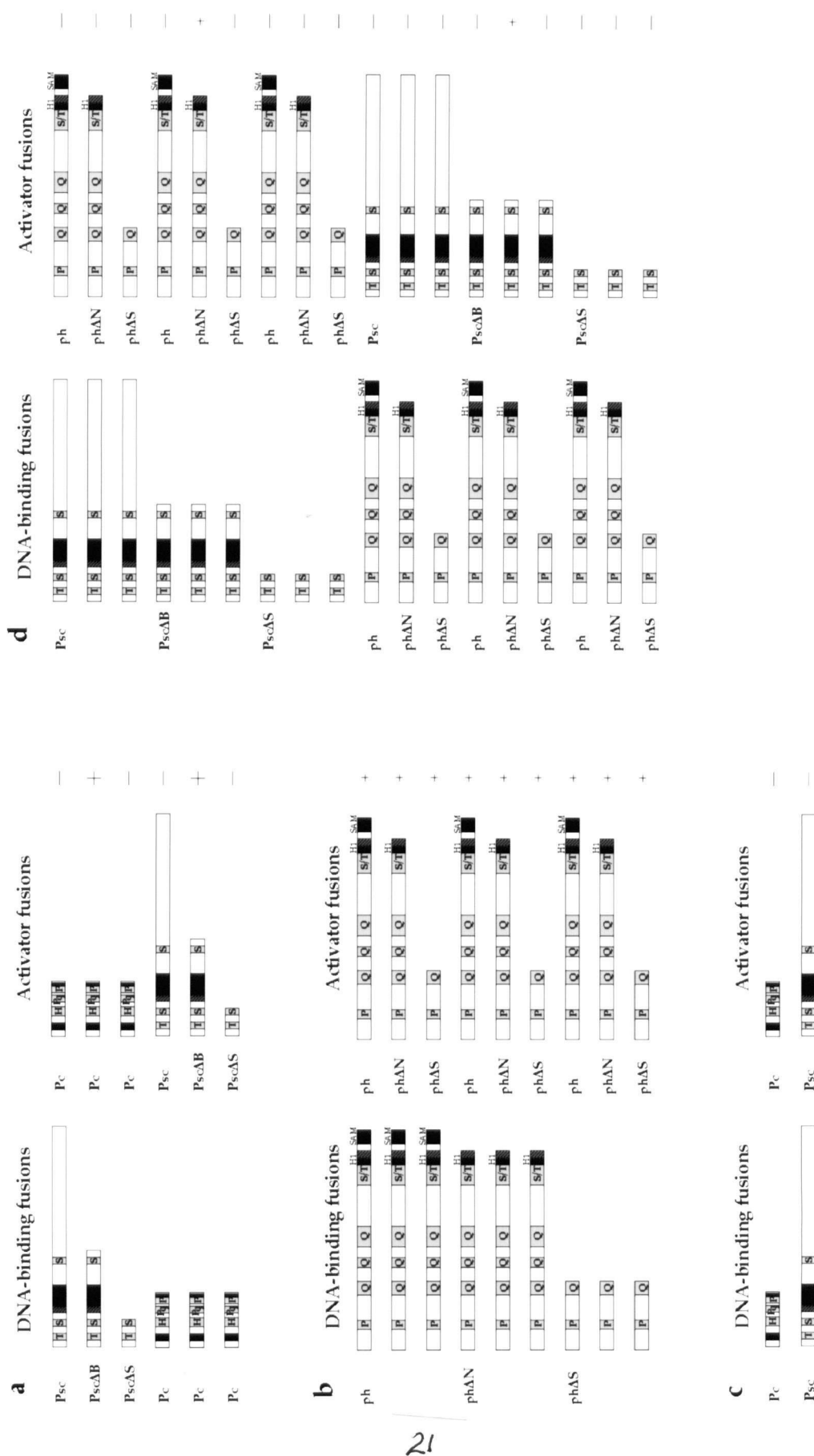



Figure 1.3

Two-hybrid interaction assay results for ph, Psc, Pc, and carboxyl deletions. DNA-binding fusions represent protein fusions to LexA, a bacterial DNA-binding protein, and activator fusions represent protein fusions to B42, a short acidic transactivation sequence. Constructs were expressed in the yeast strain EGY48, which has the LEU2 gene downstream from LexA binding sites. All pairwise combinations were tested. Combinations not shown were negative. Strong positives (1mm colonies after four days of growth on selective medium) are indicated by a large plus, weak positives (<1mm colonies after four days) by a small plus. (a) Psc interacts with Pc, however full length Psc must be deleted for this interaction to be seen. (b) ph interacts with itself through a domain or domains in the smallest amino-terminal construct. (c) Self-interactions were not seen with either Pc or Psc. (d) Psc interacts with ph, and this interaction requires carboxyl deletions of both proteins to be detected. Shadings are as described for Figure 1.1.

DNA-binding fusions

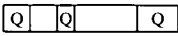
AsxA 

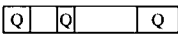


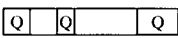


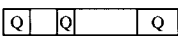


AsxQ

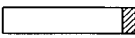


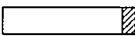


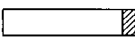


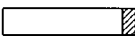


AsxC







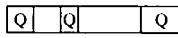


All other PcG

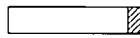
Activator fusions

AsxA 

AsxQ



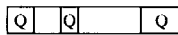
AsxC



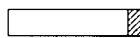
All other PcG

AsxA 

AsxQ



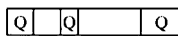
AsxC



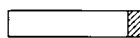
All other PcG

AsxA 

AsxQ



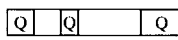
AsxC



All other PcG

AsxA 

AsxQ



AsxC




Figure 1.4

Asx interactions. No Asx construct interacted with any ph, Psc, or Pc construct. A self-interaction was seen, however, with AsxC. Shadings are as described for Figure 1.1

Figure 1.5

Two-hybrid interaction assay results for conserved sequence constructs. (a) Psc-ph interacting constructs. The interaction is delimited to the H1 domain of ph and the HTH-containing region of Psc. It is stronger in the absence of the SAM domain of ph. (b) Psc-Psc interacting constructs. This interaction was only seen with the isolated domains and was dependent on the ring finger. (c) Psc-Pc interacting constructs. The interaction appears dependent on sequences carboxyl to the chromobox of Pc and the HTH-containing region of Psc, although an interaction is seen with the ring finger in one pair. Shadings are as described for Figure 1.1

both a DNA-binding fusion, and as an activator fusion. The ring finger of Psc showed weak interaction with the activator-fusion of Δ chrPc. This may mean that although Pc makes contacts primarily with the HTH domain, it also makes weaker contacts with the ring finger domain.

When expressed in the absence of surrounding sequence, the ring finger of Psc dimerized (Figure 1.5b). This was surprising as dimerization of Psc had not been observed with any larger construct. A weak interaction between the ring finger construct and the HTH domain was seen in one orientation but not the other. This interaction may occur simply because these domains fit together naturally in the tertiary structure of the protein, or it may be part of a true Psc dimerization domain.

Caution should be used in relating the strength of interactions seen in the two-hybrid system with presumed affinities of individual proteins for one another. Two-hybrid analysis done with interactors of known affinities has shown that while interaction strength generally correlates with *in vitro* affinity, the response curve is not linear, and in many cases shows a threshold below which no response is seen [111].

***In-Vitro* Interactions**

The two-hybrid interaction assay takes place within the yeast nucleus. Because PcG proteins are transcriptional repressors, this environment is likely very close to their natural environment. However for the same reason it may also contain confounding influences. Any of these interactions could be mediated by an endogenous yeast nuclear protein with enough similarity to the *Drosophila* protein that actually functions as the mediator, hence the observed interaction may not be direct. Likewise, there may exist yeast proteins capable of interacting with the *Drosophila* fusion proteins which would occlude or prevent their interaction with each other. I therefore sought to test the identified interactions *in*

vitro. Interacting proteins and domains were subcloned into pGEX4T-1 for bacterial GST (glutathione-S-transferase)-fusion protein expression, and pET28a for T7 transcription and *in vitro* translation in a rabbit reticulocyte lysate. The T7 constructs were translated in the presence of ³⁵S-labeled methionine and incubated with GST-fusion protein immobilized on glutathione agarose. An interaction between the ³⁵S-labeled protein and the GST-fusion protein results in the co-precipitation of both on the affinity resin. Bound protein was then washed extensively, eluted with reduced glutathione, run on SDS-PAGE and autoradiographed.

The construct PscΔB (aa 1-696) originally implicated in the two-hybrid interaction was shown to interact specifically with both the minimal H1 domain of ph and with phHD (amino acids 1297-1576), the larger construct which contains the H1 domain and the SPM domain. It also bound the chromobox-deleted Pc fusion. However it did not bind a ph construct that does not contain H1, nor did it bind any Psc construct, nor GST alone (Figure 1.6a). These data corroborate the two-hybrid data. The construct PscHD (amino acids 250-473), which contains only the conserved sequences of Psc (the ring finger followed by the HTH-containing region) showed similar behaviour, although a new interaction with itself was detected (Figure 1.6b). When the homology region was broken into the ring finger and the HTH domain, the HTH domain interacted with H1 while weaker interactions were seen between HTH and Pc as well as HTH and ring-finger containing constructs (Figure 1.6c). The ring finger did not interact with H1, but did show an interaction with itself, and a weaker interaction with Pc and HTH (Figure 1.6e). Full length Psc interacted with both H1 and chromobox-deleted Pc (Figure 1.6f) recapitulating the behaviour of PscΔB.

In the translation of HTH-containing constructs of Psc, I observed smaller labeled fragments derived most likely from weak internal initiation or possibly from breakdown of

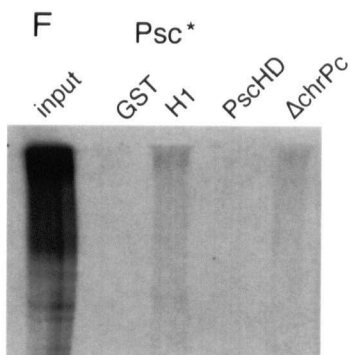
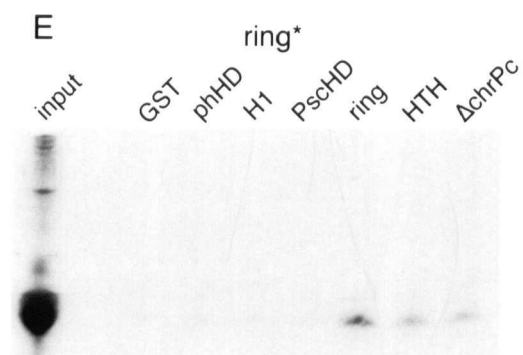
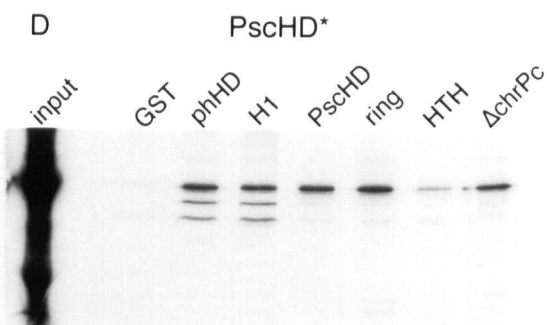
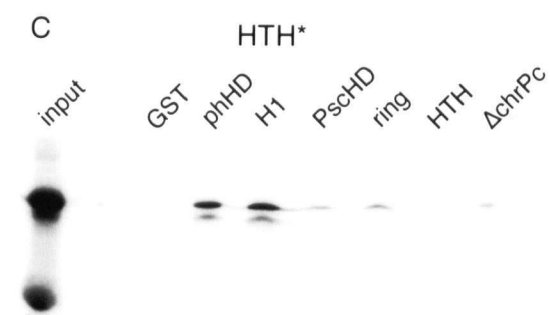
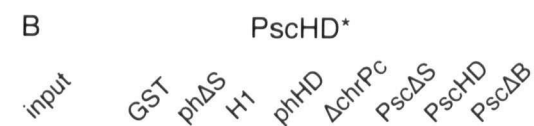
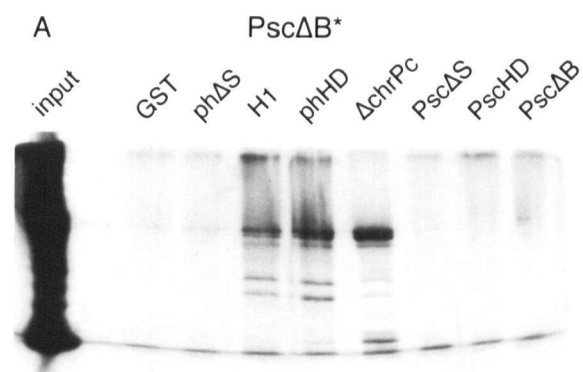


Figure 1.6

In vitro binding of reticulocyte lysate-generated ^{35}S -labeled Psc constructs to bacterially produced GST-fusions. GST-fusions were expressed in bacteria, bound to glutathione agarose beads, and blocked with BSA. ^{35}S -labeled Psc constructs were transcribed and translated in a reticulocyte lysate, and added to the blocked beads. Beads were washed, then eluted with reduced glutathione. The labeled construct used in each experiment is denoted by an asterisk. (a) Labeled Psc Δ B (aa 1-696) binds to GST fusions to regions of ph which contain H1 (aa 1297-1418) and to a GST-fusion of Pc deleted for the chromobox. It does not bind ph Δ S (aa 1-522 of ph) or other Psc constructs. (b) Labeled PscHD (aa 250-473) bind the same GST-fusions as well as GST-fusions containing aa 250-473 of Psc. (c) The Psc HTH-containing region (aa 336-473) binds H1-containing ph constructs strongly, and Pc and Psc constructs weakly. (d) PscHD (aa 250-473) binds as strongly to the ring finger alone (aa 250-335) as it does to the ring finger plus the HTH region, and only weakly to HTH. (e) The ring finger of Psc binds to itself, and also more weakly to Pc and HTH. (f) Labeled full length Psc binds GST fused to phH1 (aa 1297-1418) and to the GST-fusion of Pc deleted for the chromobox, but not to PscHD (aa 250-473).

the full length products (Figure 1.6a, b, c, d). These bound to H1-containing constructs but not to Pc. I interpret this as evidence that ph and Pc bind to different regions of PscHD. Furthermore, while H1-containing GST-fusions strongly bound both PscHD and the HTH domain, Pc strongly bound only the complete PscHD, and bound both HTH and the ring finger more weakly. This is further evidence that Pc makes use of a different interaction surface than that used by ph, and that this interaction surface is likely made up of elements from both the ring finger and HTH. The self-interaction of PscHD required the ring finger (Figure 1.6d), and was not seen with the larger construct Psc Δ B.

The amount of sample loaded in each experiment was such that a bound band of equal intensity to the input band represents approximately 10% of input labeled protein remaining bound through multiple wash steps of increasing stringency, and eluting with reduced glutathione. By comparing the relative intensity of bound band to input band between experiments, the most stable association under these conditions is seen between the labeled HTH fragment and ph constructs containing the H1 domain. This level of bound to input protein is similar to that seen in experiments with the SPM domain interactions of ph and Scm [112].

Ternary vs. Binary Complexes

Independently, the co-IP and domain analysis are consistent with both a ternary complex or multiple binary complexes, however a ternary complex seems more likely considering the data together. The co-IP demonstrates the existence of protein complexes containing Pc-ph and Pc-Psc, while the domain analysis only gives evidence for direct interactions between Pc-Psc and ph-Psc. A ternary complex with Psc as the bridge explains both sets of data. Alternatively a direct Pc-ph interaction may have eluded these assays, or may be mediated by another unidentified protein in the nuclear extract.

Isolated Domain Interactions can be Modulated by External Sequences

In the domain analysis, some interactions were affected by parts of the proteins not implicated in binding. In the case of the ph-Psc interaction, the presence of the ph SPM domain weakened the interaction in most two-hybrid combinations although not in the *in vitro* assay. Since the SPM domain has the potential for heterologous self-association, and since yeast proteins with this domain exist [103] the modulation might be an artifact of ph interacting with endogenous yeast proteins. In *Drosophila* there are at least two nuclear proteins that contain the SPM domain: Scm [102] and l(3)mbt [113]. Whether the Scm-ph interaction affects the ph-Psc interaction is an open question. The two-hybrid interactions were also attenuated by full length Psc. This may be due to the ability of full length Psc to repress transcription in yeast (Chapter 4): Consistent with this, the full length protein does interact with the expected domains of ph and Pc in the *in vitro* assay.

The greatest inconsistency between the two-hybrid results and the *in vitro* results was seen with the Psc-Psc interaction. In the two-hybrid system, self interactions were only seen with the isolated ring-finger domain. *In vitro*, self interactions were seen with the ring finger and with the complete conserved region which includes the ring finger, but not with larger constructs. The most likely reason for the discrepancy is the fact that these assays employ proteins produced from three different sources: yeast cells, bacterial cells, and a reticulocyte lysate. A protein expressed in a heterologous system will not necessarily have the same folding and covalent modifications as its native cognate. A given domain may be prevented by its expression context from attaining the fold or covalent modifications required for interaction. The fact that large parts of Psc from outside of the homology domain prevent the self-interaction may mean that the interaction is spurious, an artifact of the isolation of individual domains, or that dimerization is cryptic, and normally modulated by other parts of Psc, with dimerization only happening under certain conditions such as binding to DNA or binding other PcG proteins.

Interacting domains

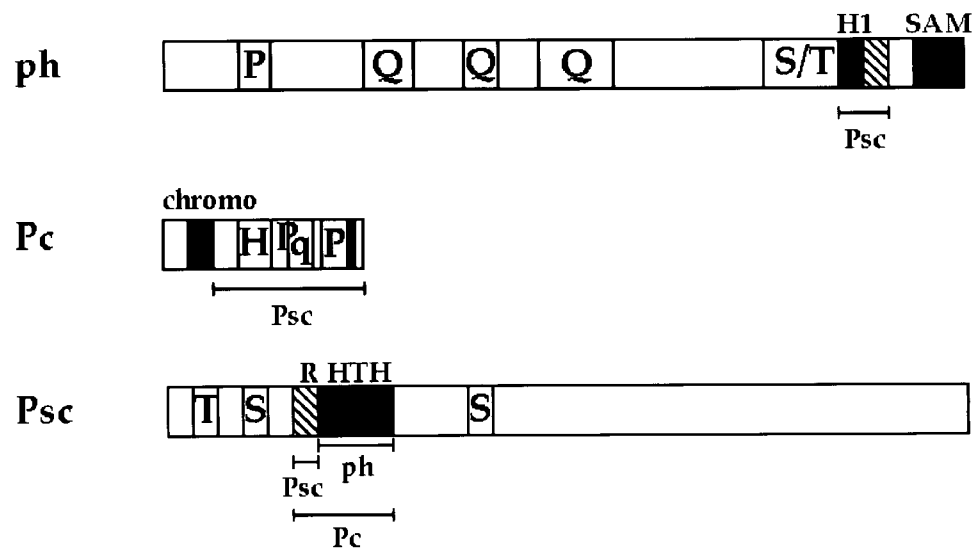


Figure 1.7

Domains involved in interactions between Pc, ph, and Psc. The Psc-interacting domain of ph spans aa 1297-1418. The Psc-interacting domain of Pc is within aa 70-390. The homology domain of Psc (aa 250-473) binds to Pc, while the HTH subregion (aa 336-473) binds to ph. Shadings are as described for Figure 1.1.

Interactions of the Vertebrate Homologues of Psc, Pc, and ph

My results are similar in general, but differ in detail from those reported for the various mammalian homologues of ph and Psc. Although the isolated Mph H1 domain and Bmi-1 HTH domain were not tested with each other, the presence of both H1 and the SPM domain of Mph was required for the interaction with Bmi-1 [73], leading the authors to speculate that Mph dimerization was a prerequisite for Bmi-1 binding. I do not see such a requirement for ph binding to Psc. The issue is complicated by the fact that besides Psc, there are two other ring-HTH containing proteins in the fly, Su(z)2 [104; 105], and L(3)Ah [107], and at least one other in the mouse, Mel-18 [106]. The mammalian complex members may truly behave differently from their fly cognates, or perhaps Mel-18, and not Bmi-1 is the functional homologue of Psc.

In this work, the Psc-Pc interaction was seen with both the ring-finger and the HTH domain of Psc. Alkema *et al.* [73] do not see a two-hybrid interaction between the mouse homologues, Bmi-1 and M33. However Hashimoto *et al.* (N. Hashimoto, H.W. Brock, M. Nomura, M. Kyba, J. Hodgson, Y. Fujita, Y. Takihara, K. Shimada, and T. Higashinakagawa, submitted) have reported such an interaction with an *in vitro* binding assay similar to that used in this work, and in one orientation in the two-hybrid system, and show that the HTH domain-containing region is required. The *Xenopus* homologues, XPsc and XPc, have been shown to interact with each other, however this interaction was shown not to require the HTH domain of XPsc [98], requiring instead the 188 upstream amino acids which contain the ring finger. While these differences may reflect true differences between fly, frog and mouse, given the sequence conservation of these domains, it is more likely that the differences arise from differences in the assays, specifically in the sizes and imprecise overlap of the constructs used. Since I have seen interactions with both the ring-finger and the HTH-containing region in both two-hybrid and *in vitro* assays, I speculate that Pc primarily contacts the HTH-containing region but

also contacts the ring-finger domain weakly. Alternatively, Pc may contact the region in between the ring-finger and the HTH domain proper, and some level of binding to each half is seen even when this region is divided. Pc and XPc differ also in their observed self affinity: Reijnen *et al.* [98] reported that full length XPc was able to interact with both its amino terminus and its carboxyl terminus, whereas I see no Pc-Pc self-interaction.

It has been shown that full length Mel-18 has the ability to bind DNA whereas a deleted version of Mel-18 lacking the ring finger does not [106]. It is possible that Psc also has this ability, and it would be interesting to know whether the binding of ph and Pc, so close to and perhaps directly on the putative DNA-binding domain would influence the putative DNA-binding properties of Psc.

The Role of Multiple Interacting Domains in PcG Complexes

Using a formaldehyde crosslinking assay, Strutt and Paro [109] have recently shown that the composition of PcG complexes is not the same at all target loci. The partially but not completely overlapping patterns of PcG protein binding to polytene chromosomes also suggest PcG complexes that are heterogeneous in composition, being different at different target sites. The interaction domains that I have described may facilitate this heterogeneity. Psc has a domain with the ability to bind either ph or Pc, or perhaps both, while ph has two very distinct domains with the ability to bind Psc on the one hand, and ph or Scm (Chapter 2) on the other. These interaction domains make possible multiple protein contacts, not all of which necessarily occur at every site. By allowing different complexes to form at different sites, more complex regulation of target genes is permitted.

All of the conserved sequences of ph and Psc have now been shown to function as protein-binding domains. This raises the question of what purpose the nonconserved sequence, which forms the vast majority of these proteins, serves. A putative complex involving only

a single copy of each of ph, Scm, Psc, and Pc would be on the order of 0.5 MDa, although the interacting amino acid sequences would account for less than 80 kDa. One possibility is that the nonconserved sequence has a direct transcriptional repression function that is conserved in the absence of sequence conservation. An alternative is that transcriptional repression is an indirect result of the bulk of the protein complex, which either excludes transcriptional activators from the vicinity of their binding sites, or prevents their interaction with the basal transcription machinery. If this were the case, the PcG proteins could be described as very large molecules with small domains that can interact with each other promiscuously, allowing bulky heterogeneous complexes to form at their various sites of action.

Chapter II: The SAM domain

To explore interaction-space outwards from the initially defined ph-Psc-Pc and Asx interactions, I tested other PcG proteins against the two-hybrid panel. LexA-fusions of esc and Pcl provided by Jeffery Simon and Rob Saint, respectively, did not interact with any members of the panel. An activator fusion of E(z) showed a weak interaction with Lex-esc (an interaction also noted by Jeffery Simon, personal communication), but not with any other member of the panel. Finally, in a collaboration with Jeffery Simon, an interaction between Scm and ph was discovered. The minimal domains required for this interaction were the carboxyl terminal SPM domains of ph and Scm (aa 1511-1576 of ph, and 797-877 of Scm).

The SAM Domain

The SPM domain is unusual among conserved domains of PcG proteins in that there are distinct paralogous sequences not only in other chromatin proteins but also in cytoplasmic proteins. This domain was in fact originally identified in comparisons of cytoplasmic proteins, and named the SAM domain [103]. The name SAM is an acronym for Sterile Alpha Motif, reflecting the putative alpha helical structure that is strongly predicted for this sequence [103]. Because the proteins Byr2p [114; 115] and C33B4.3 [116], have SAM domains at their extreme N (amino acids 1-66) and C (amino acids 1045-1110) termini respectively, the boundaries of this domain are clearly defined. Database searches have identified SAM domains in over 60 other proteins [117] that share no obvious common function.

A possible function of the SAM domain is to associate with other SAM domains, either homotypically whereby two identical SAM domain-containing proteins associate, or heterotypically whereby two different SAM domain-containing proteins associate through

the interaction of their SAM domains. Two SAM domain-containing proteins required for mating in *S. pombe*, Ste4p and Byr2p, have been shown to interact with each other [118]. The interaction occurs through regions of both proteins that contain SAM domains: amino acids 1-160 of Ste4p and amino acids 1-392 of Byr2p. The SAM domain of Byr2p is essential for this interaction, as a single base substitution in the Byr2p SAM domain abolishes Ste4p-binding activity [118]. A subset of the ETS family of transcription factors including ETS-1, ERG-2, and TEL from vertebrates, and Pointed-P2 and Yan from *Drosophila* have a SAM domain near their amino termini, referred to previously as the B domain [119] or the *pointed* domain [120]. The SAM domain of TEL when fused to either the PDGF-b receptor t(5;12) or the AML1 gene t(12;21) induces oncogenic transformation [121; 122] and oligomerization through the SAM domain is essential for the constitutive activation of TEL-PDGF-b's protein kinase activity and mitogenic properties [123]. The presence of this domain in two PcG proteins, as well as in a wide variety of other proteins, prompted a more detailed functional analysis. I was interested to know what sequence features of this domain governed its association behaviour, and if this behaviour had implications for the complexes formed by the PcG.

The ph SAM Domain Mediates Self-Association *In Vitro*

I generated a variety of GST-fusions to carboxyl sequences of ph (Figure 2.1a) and tested each for the ability to bind the *in vitro* translated ³⁵S-labeled polypeptides phHD and pSAM from the carboxyl terminus of ph. The GST-fusions were immobilized on glutathione agarose, mixed with labeled polypeptide, washed extensively in 500mM NaCl and eluted with reduced glutathione. Labeled bound polypeptide was detected by gel autoradiography. Both labeled polypeptides tested, phHD (amino acids 1297-1576) which contains the H1 domain and the SAM domain, and the smaller pSAM (amino acids 1511-1576) containing only the SAM domain, showed similar behaviour (Figure 2.1b and c). They bound only to GST-fusions with an intact SAM domain. phHD showed aberrantly

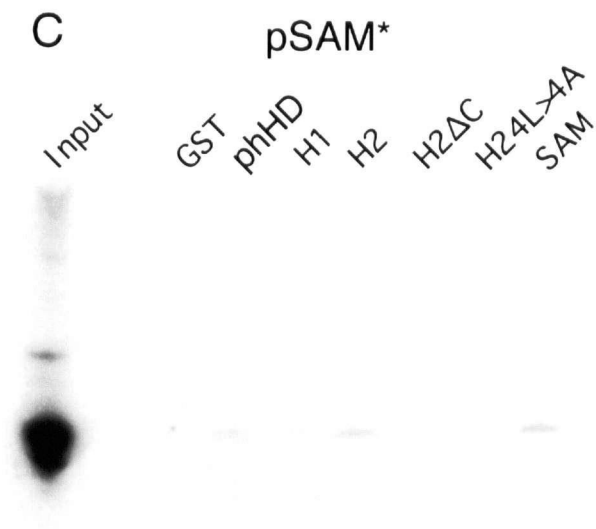
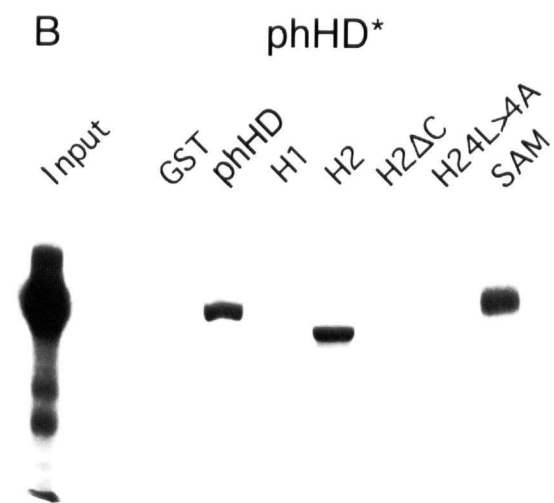
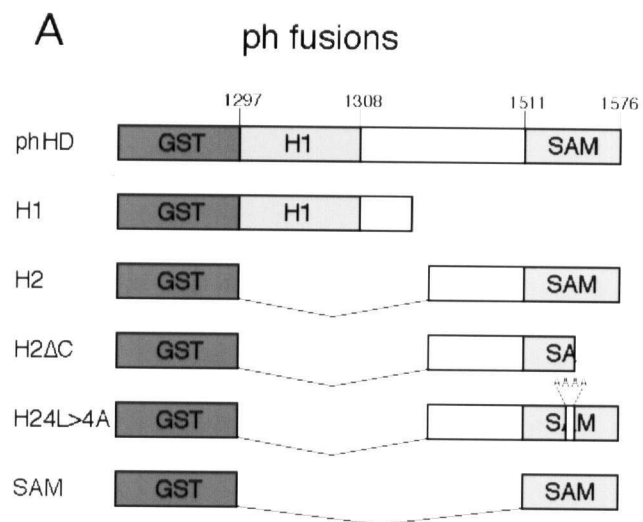


Figure 2.1

Self-binding activity of the carboxyl terminus of ph. (a) GST-fusion constructs. (b) Binding of ^{35}S -labeled pHHD to the GST fusions. (c) Binding of ^{35}S -labeled pSAM to the GST fusions. ^{35}S -labeled constructs were transcribed and translated in a rabbit reticulocyte lysate, and binding and elution was done as described above (Figure 1.6). The labeled construct used in each experiment is denoted by an asterisk.

slow migration due to the high content of proline in the sequence just upstream of the SAM domain. On binding to and elution from the H2 construct, pHHD showed a slight and reproducible shift in migration which I attribute to a conformational change in the proline rich region (possibly a peptide bond isomerization) that is resistant to denaturation in the SDS loading buffer.

The first lane in all gels contains a fraction of the total labeled polypeptide representing one tenth of the amount added to each individual binding reaction. A bound band equal in intensity to the input band therefore represents 10% of the total labeled polypeptide binding, remaining bound through the wash steps, and eluting with reduced glutathione. By comparing the input lanes with the bound lanes (Figure 2.1b and c), the notable difference between the larger pHHD construct and the minimal pSAM construct is that binding of the larger construct under these conditions is much stronger than binding of the minimal SAM domain. Nevertheless the fact that the minimal SAM domain binds to itself demonstrates that this domain alone is sufficient for the self-association.

Homologues of the ph SAM Domain

The SAM domain of ph has several close relatives from within *Drosophila* and from other species. Figure 2.2 shows an alignment of SAM domains, where amino acid identities are boxed and conservative substitutions relative to ph residues, as determined by the Kyte Doolittle matrix and visual inspection, are shaded. Characteristic features of the domain sequence are at the amino-terminal end, where it initiates with a conserved tryptophan followed 5 residues later by a valine, in the middle where a perfectly conserved glycine is followed closely by a hydrophobic block with a strong preference for the sequence ALLLL, and towards the carboxyl-terminal end where there is an almost perfectly conserved glycine. Features in particular that separate this group from other proteins

falling under the broad consensus suggested by Schultz *et al.* [117] are the conservation of both glycines previously mentioned (neither of which are part of the consensus of Schultz *et al.*), and a strong preference for the conserved tryptophan to be followed by the sequence [ST]-X-[DE]-[DE].

Homotypic and Heterotypic Self-Association of ph SAM Homologues

I selected several members of this group of SAM domains to determine whether the self-association function was conserved with the sequence. Proximal and distal ph are semi-redundant genes [124]. Therefore I was interested to determine whether the binding properties of their SAM domains differed, particularly when the domain was present in the context of its upstream sequence, the only region of interstitial sequence dissimilarity between proximal and distal ph. The SAM domain of distal ph is nearly identical to that of proximal, with only three conservative substitutions (Figure 2.2). RAE-28 was chosen as it is the recognized mammalian orthologue of ph [73; 101] and has 40/64 amino acid identities. Scm represented a paralogous SAM domain from within the same species [102], with 25/64 identities. BEB1 represented the most related SAM domain from *S. cerevisiae* [125; 126] and shares 16/64 amino acids with ph over the SAM domain. The proximal and distal ph SAM domains were tested as isolated domains as well as in the context of their unique upstream sequences. I use the nomenclature bSAM, rSAM, sSAM, pSAM, and dSAM to designate the minimal SAM domains of BEB1 (amino acids 263-329), RAE28 (amino acids 945-1012), Scm (amino acids 797-877), proximal ph (amino acids 1511-1576), and distal ph (amino acids 1338-1403) respectively. Scm2, pH2, and dH2 represent sequences from Scm (amino acids 767-877), proximal ph, (amino acids 1427-1576) and distal ph (amino acids 1249-1404) respectively, which include the SAM domain and upstream sequence.

ph_prox(d)	1	P	P	I	S	S	W	S	Y	D	D	V	S	N	F	I	R	E	L	P	G	C	Q	D	Y	Y	D	D	-	F	I	Q	Q	E	I	D	G	35
ph_dist(d)	1	P	P	I	S	S	W	S	Y	E	E	V	S	N	F	I	R	E	L	P	G	C	Q	D	Y	Y	D	D	-	F	I	Q	Q	E	I	D	G	35
Scm(d)	1	S	Q	P	I	D	W	T	L	E	E	Y	I	Q	Y	I	E	S	N	D	N	S	L	A	Y	H	G	D	L	F	R	K	H	E	I	D	G	36
L(3)mbt(d)	1	T	N	P	L	H	W	T	S	W	D	Y	C	E	Y	I	E	R	A	L	D	S	T	D	I	A	K	Y	I	F	-	E	Q	D	I	D	G	35
389108(m)	1	T	D	P	V	E	W	T	Y	M	D	Y	Y	E	Y	F	T	-	-	E	A	G	F	P	E	Q	A	T	A	F	Q	E	Q	E	I	D	G	34
RAE28(m)	1	S	N	P	S	Q	W	S	Y	E	E	V	Y	E	F	I	A	S	L	Q	G	C	Q	E	I	A	E	E	-	F	R	S	Q	E	I	D	G	35
HIBBU50(h)	1	-	-	-	T	K	W	N	Y	E	D	Y	Y	E	F	I	R	S	L	P	G	C	Q	E	I	A	E	E	-	F	R	A	Q	E	I	D	G	32
TEL(h)	1	L	Q	P	I	Y	W	S	R	D	D	Y	A	Q	W	L	K	W	A	E	N	E	F	S	L	R	P	I	D	S	N	T	F	E	M	N	G	36
TEL(m)	1	L	Q	P	I	Y	W	S	R	D	D	Y	A	Q	W	L	K	W	A	E	N	E	F	S	L	R	P	I	E	S	N	K	F	E	M	N	G	36
ETS(d)	1	S	D	P	R	L	W	S	R	E	D	Y	L	V	F	L	R	F	C	Y	R	E	F	D	L	P	K	L	D	F	D	L	F	Q	M	N	G	36
YAN(d)	1	S	D	P	R	L	W	S	R	E	D	Y	L	V	F	L	R	F	C	Y	R	E	F	D	L	P	K	L	D	F	D	L	F	Q	M	N	G	36
C-ETS-2(h)	1	K	N	P	W	L	W	S	E	Q	Q	Y	C	Q	W	L	L	W	A	T	N	E	F	S	L	Y	N	Y	N	L	Q	R	F	G	M	N	G	36
BOI1(y)	1	K	M	A	K	S	W	S	P	E	E	Y	T	D	Y	F	S	L	Y	G	F	D	Q	S	T	C	N	K	-	F	K	E	H	Q	Y	S	G	35
BEB1(y)	1	Y	E	A	E	F	W	S	P	E	E	I	T	A	Y	F	I	M	E	G	Y	D	Y	Q	S	A	S	R	-	F	Q	K	H	K	I	S	G	35
DAGkinase(h)	1	A	P	Y	H	L	W	G	T	E	E	Y	A	A	W	L	E	H	L	S	L	C	-	E	Y	K	D	I	-	F	T	R	H	D	I	R	G	34
mMG11(m)	1	T	D	L	R	T	W	E	P	E	D	Y	C	S	F	L	E	N	R	G	F	R	E	K	K	Y	L	D	I	F	R	D	N	K	I	A	G	36
ph_prox(d)	36	Q	A	L	L	L	L	K	E	K	H	L	Y	N	A	M	G	M	K	L	G	P	A	L	K	I	V	A	K	V	E	S	I	K	E	Y	70	
ph_dist(d)	36	Q	A	L	L	L	L	K	E	N	H	L	Y	N	A	M	G	M	K	L	G	P	A	L	K	I	V	A	K	V	E	S	I	K	E	Y	70	
Scm(d)	37	K	A	L	L	L	L	N	S	E	M	M	M	K	Y	M	Q	L	K	L	G	P	A	L	K	I	C	N	L	Y	N	K	V	N	G	R	71	
L(3)mbt(d)	36	R	A	L	L	M	L	G	R	K	E	L	D	T	Y	L	K	L	K	Y	G	P	A	V	K	L	Y	S	L	I	L	N	L	R	I	A	70	
389108(m)	35	K	S	L	L	L	M	Q	R	T	D	Y	L	T	G	L	S	I	R	L	G	P	A	L	K	I	Y	E	H	H	I	K	Y			66		
RAE28(m)	36	Q	A	L	L	L	L	K	E	E	H	L	M	S	A	M	N	I	K	L	G	P	A	L	K	I	C	A	K	I	N	Y	L	K	E	T	70	
HIBBU50(h)	33	Q	A	L	L	L	L	K	E	D	H	L	M	S	A	M	N	I	K	L	G	P	A	L	K	I	Y	X	R	I	S	M	L	K		65		
TEL(h)	37	K	A	L	L	L	L	T	K	E	D	F	R	Y	R	-	S	P	H	S	G	D	Y	L	Y	E	L	L	Q	-	H	I	L	K	Q	R	69	
TEL(m)	37	K	A	L	L	L	L	T	K	E	D	F	R	Y	R	-	S	P	H	S	G	D	Y	L	Y	E	L	L	Q	-	H	I	L	K	Q	R	69	
ETS(d)	37	K	A	L	C	L	L	T	R	A	D	F	G	H	R	-	C	P	G	A	G	D	Y	L	H	N	Y	L	Q	-	M	L	I	I	E	S	69	
YAN(d)	37	K	A	L	C	L	L	T	R	A	D	F	G	H	R	-	C	P	G	A	G	D	Y	L	H	N	Y	L	Q	-	M	L	I	I	E	S	69	
C-ETS-2(h)	37	Q	M	L	C	N	L	G	K	E	R	F	L	E	L	A	P	D	F	Y	G	D	I	L	W	E	H	L	E	-	Q	M	I	K	E	N	70	
BOI1(y)	36	K	I	L	L	E	L	E	L	E	H	L	K	E	L	D	I	N	S	F	G	I	R	F	Q	I	F	K	E	I	R	N	I	K	S	A	70	
BEB1(y)	36	K	I	L	L	E	L	E	L	Y	H	L	K	E	L	D	I	N	S	F	G	T	R	F	E	I	F	K	E	I	E	K	I	K	E	A	70	
DAGkinase(h)	35	S	E	L	L	H	L	E	R	R	D	L	K	D	L	G	V	T	K	Y	G	H	M	K	R	I	L	C	G	I	K	E	L	S	R	S	69	
mMG11(m)	37	S	F	L	P	F	L	D	E	D	R	L	E	D	L	G	V	S	S	L	E	E	R	K	K	M	I	E	C	I	Q	Q	L	S	Q	S	71	

Figure 2.2

Alignment of relatives of the ph SAM domain. Amino acid identities are boxed.

Conservative changes from the SAM domain of ph are shown shaded. The organism from which the protein sequence is taken is denoted by a letter in parenthesis: d - *Drosophila*, m - Mouse, h - Human, y - yeast (*S. cerevisiae*). The sequence shown begins 5 amino acids upstream of the SAM domain proper, of which the conserved tryptophan is residue number 1.

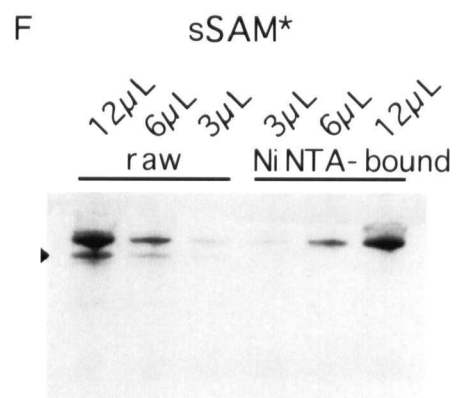
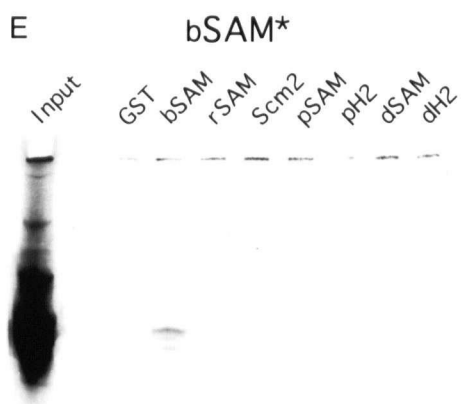
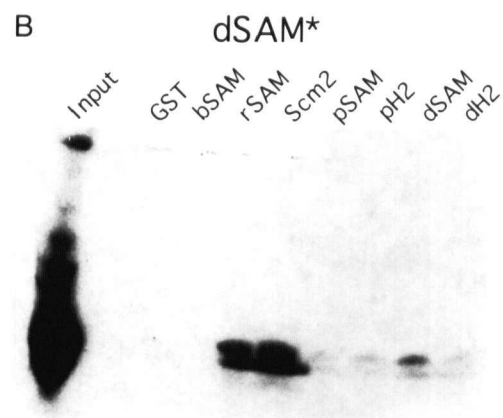
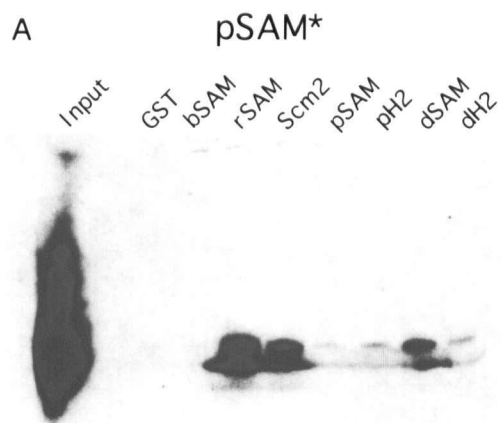


Figure 2.3

Homotypic and heterotypic binding of various SAM domains. (a) ^{35}S -labeled polyhomeotic proximal SAM domain tested for binding to GST and 7 GST-SAM domain fusions. The lane marked 'input' contains a fraction of the ^{35}S -labeled polypeptide before binding. (b) ^{35}S -labeled polyhomeotic distal SAM domain. (c) ^{35}S -labeled RAE-28 SAM domain. (d) ^{35}S -labeled Scm SAM domain. (e) ^{35}S -labeled BEB1 SAM domain. (f) ^{35}S -labeled Scm SAM domain tested for binding to Ni-NTA agarose. The arrowhead indicates the shorter product lacking the 6xHis tag.

The labeled polypeptides pSAM and dSAM showed similar behaviour. They both showed weaker binding to ph SAM domains than to those of sSAM or rSAM. Interestingly, both pSAM and dSAM showed stronger binding to dSAM than to pSAM. Neither bound bSAM or GST alone (Figure 2.3a, b). rSAM and sSAM behaved similarly to each other, but differently from pSAM and dSAM. Both bound pSAM and dSAM more strongly than they bound themselves and each other, and their respective homotypic interactions were stronger than those of pSAM or dSAM (Figure 2.3c, d). Neither bound bSAM or GST alone. bSAM was distinctive in that it bound only to itself, but not to any of the other SAMs, nor to GST alone (Figure 2.3e). These data confirm the hypothesis that the SAM domain is a self-association motif. Each SAM domain tested has the ability to bind to itself *in vitro*, albeit weakly in some cases. Furthermore with the exception of the yeast SAM domain which is most divergent in sequence, each is also capable of binding other SAM domains. However these heterotypic interactions occur with different affinities, and do not occur in any combination with bSAM. Because these domains all share the amino acids of the consensus sequence yet behave differently, the specificity of association must be derived from the nonconserved amino acids.

A protein doublet was seen in the *in vitro* translation products of some constructs, particularly with sSAM. The *in vitro* translated products were transcribed from the vector pET28a, which in addition to providing an initiation methionine, also adds a 6xHis tag followed by a thrombin cleavage site to all proteins translated. Only the upper band of the doublet (the full length product) bound to Ni-NTA agarose (Figure 2.3f). The lower band therefore lacks the 6xHis tag, and is consistent in size with the proteolytic product resulting from cleavage at the thrombin site, most likely by minor contaminating proteases in the reticulocyte lysate.

Mutations in Conserved Residues Have Different Effects on Binding to Different SAMs

The evidence presented above shows that the SAM domain self-associates and that specificity arises from the non-conserved amino acids. I wished to investigate the role of the conserved amino acids in the SAM domain. Therefore I created 5 mutations in residues that were conserved within the group of proteins tested and within the broader group of SAM domains (Figure 2.4a). The mutations were made in the pSAM domain and expressed as fusions to GST. When tested for binding to labeled pSAM, all 5 mutations abolished binding (Figure 2.4b). However when the mutants were tested with the larger phHD construct, weak binding was seen with the mutant I62D (Figure 2.4c). The strongest binders to wild type pSAM were rSAM and sSAM (Figure 2.3). They showed some binding to I62D, and in addition showed binding to L33A and L41A (Figure 2.4d,e). No construct bound the mutants W1A or G50A. By comparing the intensities of mutant-bound bands to wild-type-bound bands, it is apparent that no mutant SAM bound as strongly as wild type. An allelic series can be constructed where (W1A, G50A) > (L33A, L41A) > I4D with severity being measured by the number of constructs whose binding is abolished, or by the levels of binding compared to wild type pSAM.

SAM: A Self Association Motif

The carboxyl terminal region of ph can self-associate *in vitro*, and this self-association is a function of the SAM domain. I have demonstrated *in vitro* self-association of four other SAM domains from different proteins derived from fly, mouse, and yeast. In accordance with these observations, I propose to keep the name suggested for this sequence by Ponting, namely SAM [103], but to redesignate the acronym: Self Association Motif.

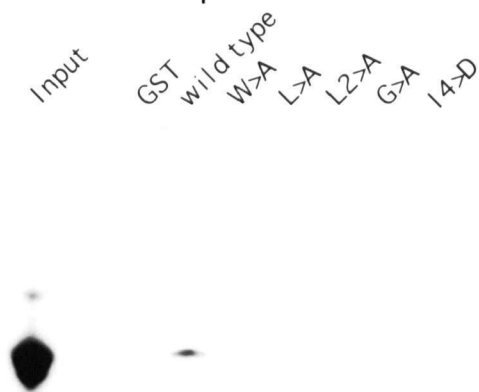
Although homotypic self-association was seen for all SAM domains tested, their relative affinities for self varied, with the SAM domains of RAE28 and Scm showing high affinity

A

W1A				L33A	L41A	G50A	I62D
WSVDDVSNFI	RELPGCQDYV	DDFIQQEIDG	QALLLLKEKH	LVNAMGMKLG	PALKIVAKVE	SIKE	DK*
A			A	A	A		

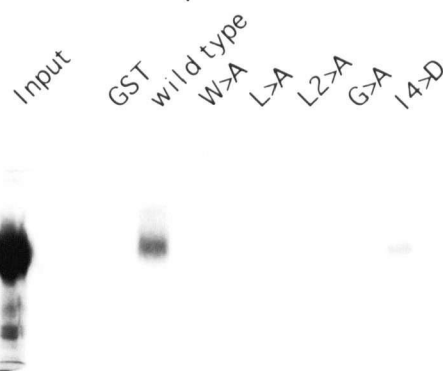
B

pSAM*



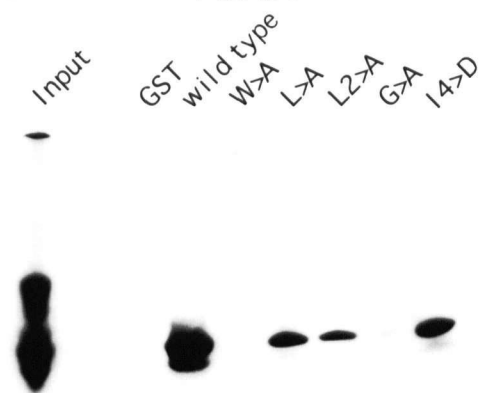
C

phHD*



D

rSAM*



E

sSAM*

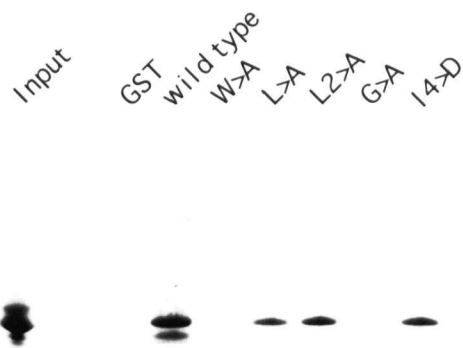


Figure 2.4.

Mutations in pSAM and their effects on binding. (a) The amino acid sequence of the wild type pSAM domain and substitution mutations. The mutation names are given above their respective positions in the wild type ph SAM domain and the substituted amino acids are given below. (b) ^{35}S -labeled polyhomeotic proximal SAM domain tested for binding to GST, GST-pSAM, and mutants of pSAM. (c) ^{35}S -labeled phHD binding to the mutant panel. (d) ^{35}S -labeled RAE-28 SAM domain binding to the mutant panel. (e) ^{35}S -labeled Scm SAM domain binding to the mutant panel.

and proximal and distal ph, and BEB1 SAM domains showing low affinity. In addition, various heterotypic interactions were noted between all SAM domains with the exception of that of BEB1. These interactions varied in strength, with affinities between RAE28-ph SAMs and Scm-ph SAMs being even stronger than the strongest homotypic interactions. These observations demonstrate that while association in general may be a function of the conserved residues of the domain, the specificity of the interaction is determined by non-conserved residues. Thus, particular sets of nonconserved residues generate particularly good interaction surfaces, such as sSAM-pSAM, while others such as bSAM-pSAM do not. The lack of any heterotypic interaction with the SAM domain of BEB1 is perhaps not surprising, given that it is the most divergent of the group. It is notable that bSAM is the only member tested in which the four leucine run is broken by a charged amino acid. Incompatibility between bSAM and the other SAMs might be expected if this region of the domain formed part of the binding surface.

The GST fusions, pH2 and dH2, which contain the SAM domains of proximal and distal ph preceded by their respective upstream unique sequences behaved similarly. Since the upstream unique sequences contain the most significant interstitial sequence divergence between proximal and distal ph, it has been suggested that the functional differences between the two isoforms arise from this region [99]. The similar binding behaviour of pH2 and dH2 suggests that the functional differences between proximal and distal ph do not arise as a consequence of SAM-mediated interactions.

Key Structural and Functional Residues of the SAM Domain

Without knowing the structure of the domain it is difficult to speculate which particular amino acids comprise the interaction interface, however the mutational analysis provides some hints. Two mutations abolished interaction between pSAM and any other construct: W1A and G50A. It is most likely that the conserved tryptophan and glycine are key

elements of the domain fold, and altering them prevents the domain from folding properly. However three mutations showed weaker binding to some constructs, and the absence of binding to others. The fact that binding was seen at all demands that the domain fold itself must be intact. Assuming this, these three mutations, L33A, L41A, and I62D, must alter key residues of the binding interface, either directly by side chain substitution, or indirectly by inducing a change in neighbouring residues that are themselves part of the binding interface.

The Potential for Promiscuous Oligomerization

The SAM domain has been found in proteins of very diverse function. In this study, pSAM, dSAM, sSAM, and rSAM come from chromatin proteins involved in transcriptional repression, while bSAM is from a cytoskeletal protein that interacts with BEM1 and is required for proper cell polarization [125; 126]. The SAM domain-containing proteins from *S. pombe*, Ste4p and Byr2p are involved in mating pheromone signal transduction. One feature that unifies this group of proteins is that they are all involved in multimeric protein complexes. ph is in a complex with at least 10 other proteins [29]. Beb1p interacts with Bem1p, and has several protein interaction domains besides the SAM domain: an SH3 recognition motif, an SH3 domain of its own, and a PH domain [125; 126]. Ste4p and Byr2p from *S. pombe* and their homologues, Ste50p and Ste11p from *S. cerevisiae* form a complex web of interactions with other components of the MAP kinase module, including Ste5p, Ste11p, Ste7p, and Fus3p [118; 127-130].

From these examples, it would appear that the SAM domain is a protein interaction domain that is particularly well suited to joining members of multiprotein complexes. In the case of the PcG, the SAM domain could facilitate heterodimerization of ph and Scm, or homodimerization of either. In addition it is possible that the SAM domain could mediate an interaction between PcG complexes containing ph or Scm with SAM domain-containing

transcription factors such as dETS, Yan or Pointed. As yet no interaction between a PcG protein and a sequence-specific DNA-binding transcription factor has been demonstrated, and it will be interesting to know whether these proteins associate with ph or Scm.

The different affinities noted for particular SAM domains may have biological relevance. In the case of proximal and distal ph and Scm, which are present in the same organism, heterotypic interactions appear to be more stable than homotypic. This would suggest that PcG complexes might prefer ph-Scm heterodimers to Scm-Scm or ph-ph homodimers. The promiscuous nature of SAM domain binding suggests that the complexes containing ph and Scm could be highly heterogeneous. Whether PcG protein complexes take advantage this potential heterogeneity for regulatory purposes whereby the exchange of one SAM domain-containing protein for another would enable or disable the silencing activity of the complex, or functional complexity whereby the complex would employ certain SAM domain-containing proteins for certain tasks such as recognizing a specific target locus, remains to be determined.

Chapter III: Asx Interactions

In the two-hybrid matrix, none of the three Asx constructs interacted with any other PcG construct. This raised the possibility that Asx might function independently from ph, Pc, and Psc at the molecular level, perhaps being a member of a different protein complex, or that Asx might function at a step in the regulatory hierarchy above or below the ph-Pc-Psc complex. In an attempt to clarify the issue, I undertook a broader search for Asx-interacting proteins by screening Asx against a two-hybrid library of embryonic cDNAs, and by screening Asx against a panel of chromatin proteins generated by Michael O'Grady.

A Conserved Domain in the C-terminus of Asx

From the three LexA-fusions to different domains of Asx, I chose Lex-AsxC as the bait for the library screen after discovering that it contained a short domain of high similarity to a previously unknown mammalian gene: I performed an exhaustive BLAST search of dbEST using the entire Asx protein sequence, and identified two human ESTs whose conceptually translated sequence matched that of the putative Asx zinc finger (Figure 3.1b). No other region of non-redundant Asx sequence gave any significant matches to any other ESTs at the time, or to any other known proteins. Both cDNAs were from the IMAGE consortium: cDNA 42515 (Genbank Accession T16795) and cDNA 840471 (Genbank Accession AA485878). I obtained cDNA 42515 from the IMAGE consortium, and sequenced it. The 1.5 kb insert encoded an open reading frame that was open through the 5' end and ended in two stop codons very near the 3' end (Figure 3.1a). Apart from the putative zinc finger, there was no similarity to Asx. Given that the interactions discovered so far between Psc, ph and Pc had been mediated by conserved domains, the C-terminus of Asx, which contained the sequence conservation, seemed the best candidate to use as bait in a two-hybrid screen.

a hAsx sequence:

GCTTGGAAACGANGCAAGGNCAGATG GNATNGTTGGTCCTCAGAGATGGGT GTCTCGAGTATGTGCGGTCCGCCAA
A W N X A R X D G X V G P Q R W V S R V C A V R Q>

100
AAGATCCCAGATTCCCTACTGCTGG TCAGTACTGAGTACCAGCCAAGAGC CGTGTGCCTGTCCATGCCTGGGTCC
K I P D S L L L V S T E Y Q P R A V C L S M P G S>

200
TCAGTGGAGGCCACTAACCCTACTTG TGATGCAGTTGCTGCAGGGTAGCTT GCCCCTAGAGAAGGTTCTTCCACCA
S V E A T N P L V M Q L L Q G S L P L E K V L P P>

300
ACCCACGATGACAGCATGTCAGAAT CCCCAAGTACCCTGACCAAAGA CCAGAGCCATGGCTCGCTACGCATG
T H D D S M S E S P Q V P L T K D Q S H G S L R M>

GGATCTTTACATGGTCTTGGAAAAA ACAGTGGCATGGTTGATGGAAGCAG CCCAGTTCTATAAGGGCTTTGAAG
G S L H G L G K N S G M V D G S S P S S I R A L K>

400
GAGCCTCTTCTGCCAGATAGCTGTG AAACAGGCACTGGTCTTGCCAGGAT TGAGGCCACCCAGGCTCCTGGAGCA
E P L L P D S C E T G T G L A R I E A T Q A P G A>

500
CCCCAAAAGAATTGCAAGGCAGTCC CAAGTTTGTACTCCCTCCATCCAGT GACAAATCCCATTACATCCTCTAGG
P Q K N C K A V P S F D S L H P V T N P I T S S R>

600
AAACTGGAAGAAATGGATTCCAAAG AGCAGTTCTCTTCTTTAGTTGTGA AGATCAGAAGGAAGTCCGTGCTATG
K L E E M D S K E Q F S S F S C E D Q K E V R A M>

TCACAGGACAGTAATTCAAATGCTG CTCCAGGAAAAGAGCCAGGAGATCT TACTACCTCGAGAACACCTCGTTTC
S Q D S N S N A A P G K S P G D L T T S R T P R F>

700
TCATCTCCAAATGTGATCTCCTTTG GTCCAGAGCAGACAGGTCGGGCCCT GGGTGATCAGAGTAATGTTACAGGC
S S P N V I S F G P E Q T G R A L G D Q S N V T G>

800
CAAGGGAAGAAGCTTTTTGGCTCTG GGAATGTGGCTGCATCCCTTCAGCG CTCCAGACCTGCGGACCCGATGCCT
Q G K K L F G S G N V A A S L Q R S R P A D P M P>

900
CTCCCTGGTGAGATCCCTCCAGTTT TTCCAGTGGGAAGTTGGGACCAAG CACAACTCCATGTCTGGTGGGGTA
L P G E I P P V F P S G K L G P S T N S M S G G V>

CAGACTCCAAGGGAAGACTGGGCTC CAAAGCCACATGCCTTTGTTGGCAG CGTCAAGAATGAGAAGACTTTTGTG
Q T P R E D W A P K P H A F V G S V K N E K T F V>

hAsx sequence, continued:

```

                                1000
GGGGGTCCTCTTAAGGCAAATACCG AGAACAGGAAAGCTACTGGGCATAG TCCCCTGGAAGTGGTGGGTCAGCTTG
G G P L K A N T E N R K A T G H S P L E L V G H L>

                                1100
GAAGGGATGCCCTTTGTTCATGGACT TGCCCTTCTGGAAATTACCCCGAGA GCCAGGGAAGGGGCTCAGTGAGCCT
E G M P F V M D L P F W K L P R E P G K G L S E P>

                                1200
CTGGAGCCTTCTTCTCTCCCTCCC AACTCAGCATCAAGCAGGCATTTTA TGGGAAGCTTTCTAAACTCCAACCTG
L E P S S L P S Q L S I K Q A F Y G K L S K L Q L>

AGTTCCACCAGCTTTAATTATTCCT CTAGCTCTCCACCTTTCCCAAAGG CCTTGCTGGAAGTGTGGTGCAGCTG
S S T S F N Y S S S S P T F P K G L A G S V V Q L>

                                1300
AGCCACAAAGCAAACCTTTGGTGCGA GCCACAGTGCATCACTTTCCTTGCA AATGTTCACTGACAGCAGCACGGTG
S H K A N F G A S H S A S L S L Q M F T D S S T V>

                                1400
GAAAGCATCTCGCTCCAGTGTGCGT GCAGCCTGAAAGCCATGATCATGTG CCAAGGCTGCGGTGCGTTCTGTAC
E S I S L Q C A C S L K A M I M C Q G C G A F C H>

                                1500
GATGACTGTATTGGACCCTCAAAGC TCTGTGTATTGTGCCTTGTGGTGAG ATAATAAATTATGGCCATGGGAAAC
D D C I G P S K L C V L C L V V R * *
ATTGT

```

b Asx-hAsx homology:

```

Asx:      1634 CACSLNAMVICQCGAFCHDDCIGAAKLCVAC-VIR* 1669
           CACSL AM++C+ CGAFCHDDCIG +KLCV+C V+R*
hAsx:     1522 CACSLKAMIMCQGCGAFCHDDCIGPSKLCVLCLVVR* 1558

```

Figure 3.1.

hAsx sequence. (a) hAsx sequence derived from cDNA 42515 of the IMAGE consortium, Genbank Accession: T16795. The DNA sequence is given above, and the conceptual translation below. The region of sequence conservation with Asx is underlined. (b) Sequence similarity between hAsx and Asx proteins at the carboxyl terminus of both proteins.

The Asx Two Hybrid Library Screen

I screened the activator fusion library RFLY1 [131] which was derived from poly(A⁺) RNA of 0-12 hour *Drosophila* embryos. The library contained 4.2×10^6 independent transformants (Russell Finley, personal communication). From approximately 10^5 cDNAs, 62 colonies that grew on selective medium in the presence of galactose as a sugar source were picked. These were named z1-z62. Of these, 20 grew when tested on dextrose and were discarded (transcription of the activator fusion is shut off by dextrose, therefore true interactions should only result in growth on galactose). Plasmids were rescued from the remaining 42, and retransformed into the bait strain, as well as a strain containing LexA alone as bait. Of these, 20 were specific to AsxC. Insert size and preliminary sequence from the 5' end classified these clones into 11 groups (Table 3.1). Clones were named after the number of the first member picked from the group.

Of the 11 interacting cDNAs identified, z28, z38, and z46 corresponded to previously identified *Drosophila* genes, *bowel*, *klett*, and *Cabeza*, respectively, and z7, z60, and z3 were obvious *Drosophila* homologues of genes cloned from other organisms, ribosomal protein L34, thioredoxin, and spermidine synthase, respectively. Other interactors showed smaller scale sequence similarities to known proteins: z34 contained glutamine repeats, shared by other members of the PcG and by many other transcription factors, and a run of seven alanines, similar to that seen in Asx itself, z2 showed similarity to thrombospondin through a cysteine-rich region, and z11 showed similarity to phosphatidylinositol (4,5)bisphosphate 5-phosphatase. Structural features of note are the presence of zinc fingers in at least two interactors, Cabeza and Bowel, and a cluster of cysteine residues in z2. z2 also contained a leucine-rich region strongly predicted to be α -helical, although not conforming to the leucine zipper motif. Preliminary sequence of z1 and z40 showed no significant similarities to any known proteins. The preliminary

Clone	Insert Size	Sequence Features	PcG interactions	AsxC	AsxC1
Strong Interactors					
<u>1</u> ,6,13,25,29,57	3.3kb		-	++	++
<u>11</u> ,27,41,50	0.9kb		-	++	+
<u>56</u> , <u>60</u>	0.6kb	thioredoxin	-	++	-
<u>7</u> ,21	0.6kb	Aides ribosomal prot. L34	-	++	na
<u>34</u>	2.2kb	polyQ, polyA, cyclin box	PscΔB*	++	++
<u>34ΔB</u>	1.3kb	polyQ, polyA	na	++	++
Weak Interactors					
<u>2</u>	1.1kb	L-helix, cysteine cluster	-	+	+
<u>3</u>	0.75kb	spermidine synthase	-	+	*
<u>28</u>	2.0kb	Bowel (multiple Zn-fingers)	-	+	-
<u>38</u>	2.5kb	Klett	-	+	na
<u>40</u>	1.1kb		Pc, ΔchrPc	+	-
<u>46</u>	0.6kb	Cabeza (single Zn-finger)	na	+	*

Table 3.1

Grouping of the Asx interactors. The underlined clone number is used to designate the entire group. The fourth column designates interactions with any members of the PcG LexA fusion panel. The rightmost column designates interaction with the construct AsxC1, an amino derivative of AsxC, representing amino acids 1139-1420. ++ strong interaction, + weak interaction, - no interaction, na not assayed, * some very weak growth. z34ΔB is a deletion construct retaining the amino 306 amino acids of z34.

sequence for each interactor not shown in this chapter is given in Appendix A, and sequence similarities between interactors and known proteins are shown in Appendix B.

To determine which interactions, if any, might require the conserved sequence at the extreme carboxyl terminus, AsxC was divided roughly in half to create the constructs AsxC1 (aa 1139-1420) and AsxC2 (aa 1412-1669). Unfortunately Lex-AsxC2 activated transcription alone as measured by its ability to promote growth on leucine deficient medium in the absence of any other plasmid, and was therefore not used in this assay. However AsxC1, which lacks the conserved sequence, was tested against each of the interactors. z1, 2, 11, and 34 retained an interaction with AsxC1, while the interaction was abolished for z28, 40, and 60 (Table 3.1, column 5). A much reduced interaction occurred for z3 and z46.

Interactions with Other PcG Proteins

If Asx were a member of the same complex(es) containing ph, Pc, and Psc, then Asx-interacting proteins might also interact with any of these three. I tested z1, 2, 3, 11, 28, 34, 40, and 60 for interaction with each member of my panel of full length PcG constructs and carboxyl deletions. One strong interaction appeared between z40 and Pc as well as a weak interaction between z34 and Psc (Table 3.1, column 4).

Evaluation of AsxC Interacting Proteins

Both the preliminary sequence and the interaction results from the PcG panel were helpful in deciding which interactors to pursue further. Ribosomal proteins (z7) are well-known false positives in two-hybrid screens (Erica Golemis, personal communication) and thioredoxin (z60) and spermidine synthase (z3) were excluded because the link between these enzymatic activities and PcG function is tenuous. From the set of previously cloned genes, *klett* had been isolated by Gunter Reuter and Michael O'Grady (personal

communication). The clone that I isolated is in fact the very clone that they had isolated as a Lex-Su(var)3-9 interactor. There are as yet, no *klett* mutants or information on *klett* expression. While its separate isolation in a *Su(var)3-9* screen could be indicative of a physical link between *Asx* and *Su(var)3-9*, the fact that it interacted with many constructs from the O'Grady panel (Michael O'Grady, personal communication) suggested that it might just as easily be a sticky protein, showing spurious interactions with many different Lex-A fusions.

On the other hand, *bowel* and *Cabeza* were more promising. *Bowel* is a multiple zinc-finger transcription factor, a member of a paralogous gene family including *odd-skipped* and *sob* [132]. *bowel* homozygotes show cephalopharyngeal and hindgut defects [133] consistent with the restricted anterior and posterior expression pattern of *bowel* mRNA in developing embryos. *Cabeza* is a glycine rich RNA-binding protein made up of five domains: An amino-terminal glycine-rich domain followed by an RRM (RNA-Recognition Motif) domain, a central glycine-rich domain, a C₂C₂ zinc finger, and a carboxyl-terminal glycine rich domain [134]. The z46 clone, which encodes amino acids 246 through to the stop codon of *Cabeza*, lacks the RRM and amino-terminal glycine-rich domain. *Cabeza* is expressed ubiquitously early, is later enriched in the embryonic brain and CNS, but is absent from larval brain. It is also present in eye-antennal, wing, and leg imaginal discs, and is enriched in adult head vs. body [134]. It is located in the nucleus of all cell types examined (larval fat body and imaginal disc).

Asx protein is ubiquitous in the early embryo, and concentrated in the neurectoderm and CNS in late embryos [135]. *Asx* mRNA levels drop markedly in larvae and increase again in pupae and adults. *Asx* mutants die as larvae with mild posterior transformations of the cuticle of abdominal and thoracic body segments. They show striking cephalopharyngeal defects including the complete failure of head involution [19]. In addition, *Asx* mutations

cause derepression of a *Ubx-lacZ* reporter gene in the CNS of parasegments 2-5 [8; 135] indicative of a requirement for Asx in transcriptional repression in the CNS of these parasegments. The expression pattern of Asx as well as the domain of effect of Asx mutations coincide well with those of both *bowel* and *Cabeza* making these two proteins attractive prospects for *in vivo* binding partners of Asx. At the time of this writing, studies by others are underway to ascertain the role of Asx in Bowel and Cabeza function.

Possible Roles for Bowel and Cabeza

Since Bowel is likely to be a sequence-specific DNA-binding protein, it may play a role in recruiting Asx to target genes in Bowel-expressing tissues. It remains to be seen what genes Bowel regulates, and whether it is a positive or negative transcriptional regulator.

If Asx interacts with Cabeza *in vivo*, one would assume that RNA is a component of Asx complexes, or that Asx complexes have the ability to bind to RNA. Renato Paro has claimed (unpublished) that Polycomb fractionates differently in the presence vs. the absence of RNase, suggesting that RNA is a component of Polycomb-containing complexes. It may be that Cabeza is the RNA-binding protein responsible. One could also imagine a regulatory function for RNA. One of the perennial questions about PcG protein function is why some targets are silenced in one set of cells but not in others, when the PcG proteins responsible for maintenance of the silenced state are present in both cell types. The corollary is also puzzling: how is it that some loci are active in salivary glands, while others are silenced, when PcG proteins are found at both? One solution to this problem is to postulate that the PcG proteins are present at all target sites, but are either active or inactive, depending on the transcriptional state of the locus at the time of complex assembly. Negative regulatory proteins such as Hunchback, bound to DNA at the time of PcG programming, could provide input favouring the silencing state, while positive factors could provide antipodal input. mRNA itself, transcribed from the locus in question, could

be such a positive factor, signaling into the complex through an RNA-binding protein such as Cabeza.

The Complete Sequences of z34 and z40

The interactions noted for z34 and z40 with the PcG panel, as well as the presence of glutamine and alanine repeats in z34 prompted the complete sequencing of these two interacting clones. I subcloned the inserts into pBluescript, and using a combination of forward/reverse primers and 5' and 3' restriction site deletions, sequenced them in their entirety (Figures 3.2 and 3.3). Both contained open reading frames that extended through the 5' end of the sequence, and eventually terminated before reaching the 3' end. Although neither gene has previously been described in *Drosophila* or any other organism, both genes are represented by multiple *Drosophila* ESTs. By analyzing 5' EST sequences, the 5' ends of both open reading frames could be determined. EST sequence 5' to the z34 clone contained stop codons in all three reading frames. The first methionine codon of the z34 cDNA is in my interacting clone, therefore the z34 clone contains a complete open reading frame (Figure 3.2). z40 on the other hand, was found to be missing 89 nucleotides of open reading frame at the 5' end. Figure 3.3 shows the melded sequence of my z40 clone and the *Drosophila* EST 1032424 (Genbank accession: AA391083).

z34 Contains a Cyclin Box

I rescreened the protein sequence databases using the full length sequences, and was still unable to find any significant match for z40. z34 however, now gave high scoring matches to cyclins, with the highest being to a human G-type cyclin. The sequence similarities were strongest at the cyclin box, which encodes the CDK (cyclin dependent kinase)-binding domain of the cyclin. Figures 3.4a and 3.4b show the cyclin box sequence of z34 aligned with the same for all of the *Drosophila* cyclins (a) and human and mouse cyclin G (b). Cyclin G has not yet been described in *Drosophila*. As is clear from inspection of the

z34 cDNA sequence

```

      *                               *
GAATTCGGACGAGGCGATTTTTTG GAAATAAGAAGCAAGAAAAGCAGAT ACTGATCAAAACGCAGAGGCATCCG
      *                               *
      100
      *                               *
GGTACTAGGCCAGCCCTCACAATGT CTGTCCCTGTACGCTACTCCTCTGC TGCCGCCGAATACGCCGCCGAAGTT
      M   S   V   P   V   R   Y   S   S   A   A   A   E   Y   A   A   E   V>
      *                               *
      200
      *                               *
GATTGTGAGTTGGAGAGCACTCTGC AACAGCAGCAACAGTTGCACTTGCA ACAGCAATACGAGCAATACCAGCAC
      D   C   E   L   E   S   T   L   Q   Q   Q   Q   Q   L   H   L   Q   Q   Q   Y   E   Q   Y   Q   H>
      *                               *
      300
      *                               *
TACCACTACCAACGGGAGCAGGATA TCGCCTACTATTGCCAGTTGCAGGC GGCGCGGCAGCAGGAGCAGTTGATG
      Y   Q   Y   Q   R   E   Q   D   I   A   Y   Y   C   Q   L   Q   A   A   R   Q   Q   E   Q   L   M>
      *                               *
      *                               *
CAGCAGAGGACATCGATGTCGTCGT CGGTTATGCCAGGCCTAGCCTTGCC CCAGGATCACCAGGACCACCCAGCC
      Q   Q   R   T   S   M   S   S   S   V   M   P   G   L   A   L   P   Q   D   H   Q   D   H   P   A>
      *                               *
      400
      *                               *
GCCCTTTTGAACGGACCCCAACAACA ACAACATCGGACTCGCCATGGACGC CCACAGCATCAACGCCATTCTGGTC
      A   L   L   N   G   P   H   N   N   N   I   G   L   A   M   D   A   H   S   I   N   A   I   L   V>
      *                               *
      500
      *                               *
GACGACGAGCAGGCCTCGACTTCGG CCCAGGCTGCCGCCGCTGCTGCCGC ATCCGCGGGTGGATCTGCTGGTGGC
      D   D   E   Q   A   S   T   S   A   Q   A   A   A   A   A   A   A   S   A   G   G   S   A   G   A>
      *                               *
      600
      *                               *
GGATCGGGATCGGGATTGGGTGGTG CTATCCGTGGGGGCAAGCTGGGCAA CGCGATTAAACGCAATGCAGAGATG
      G   S   G   S   G   L   G   G   A   I   R   G   G   K   L   G   N   A   I   N   R   N   A   E   M>
      *                               *
      *                               *
CCAAGTATTGGATGAGGATTGCGG ACGAGGGCCGGTATGGGACACCGGG TGCTGCTGGCTTGGAAATATCAGAAG
      P   T   D   W   M   R   I   A   D   E   G   R   Y   G   T   P   G   A   A   G   L   E   Y   Q   K>
      *                               *
      700
      *                               *
TACGAACAACAACAACAACCTGGAGG ATCTGGCGGAGTCCGAGGCAGGAGC GTACGGTGGAGCCAGCAACAACAAC
      Y   E   Q   Q   Q   Q   L   E   D   L   A   E   S   E   A   G   A   Y   G   G   A   S   N   N   N>
      *                               *
      800
      *                               *
GGCGAATCGTCGTCGTCCTTGA AAA AGCTAGAGGATCAGCTGCACGCCCT CACCTCGGACGAGTTGTACGAAACC
      G   E   S   S   S   S   L   K   K   L   E   D   Q   L   H   A   L   T   S   D   E   L   Y   E   T>
      *                               *
      900
      *                               *
CTCAAGGAGTACGACGTCCTGCAGG ACAAGTTCCACACGGTGTCTGTT GCCCAAGGAATCAAGGCGTGAGGTT
      L   K   E   Y   D   V   L   Q   D   K   F   H   T   V   L   L   L   P   K   E   S   R   R   E   V>
      *                               *
      *                               *
      Bam
      *                               *
ACTGCCGGAGGACGAGATGGATCCG CCTACGTGCTGCGCTGCCTGAAGAT GTGGTACGAGCTGCCCTCCGACGTC
      T   A   G   G   R   D   G   S   A   Y   V   L   R   C   L   K   M   W   Y   E   L   P   S   D   V>

```

z34 cDNA sequence, cont.

```

1000
*
CTGTTCTCGGCCATGAGCCTGGTGG ACCGCTTCCTGGATCGCATGGCCGT CAAGCCGAAGCACATGGCCTGCATG
L F S A M S L V D R F L D R M A V K P K H M A C M>

1100
*
AGCGTGGCCTCGTTCCACTTGGCCA TCAAGCAGCTGGACTTGAAACCCAT TCCCGCCGAGGATCTGGTTACAATA
S V A S F H L A I K Q L D L K P I P A E D L V T I>

1200
*
TCTCAGTGTGGTTGTACCGCTGGTG ATCTGGAACGCATGGCCGGCGTGAT TGCCAACAAGCTGGGCGTCCAGATG
S Q C G C T A G D L E R M A G V I A N K L G V Q M>

*
GGACATGCACCGATCACTTCTGTGA GCTACCTGCGCATCTACTACGCCCT CTTCCGCAACTTGGCGAAGGAGATC
G H A P I T S V S Y L R I Y Y A L F R N L A K E I>

1300
*
GGCGGCGACTTTTTCAGTTCTACC AGCAGCTCATCAAGCTGGAGGAACT GGAGAACC GCCTGGAGATCCTGATG
G G D F F K F Y Q Q L I K L E E L E N R L E I L M>

1400
*
TGCGACGTGAAGACCACGGTGATCA CGCCCTCGACGCTGGCGCTGGTGCT CATCTGCCTGCACCTGGACTTCCAC
C D V K T T V I T P S T L A L V L I C L H L D F H>

1500
*
ATCAAGGAGTCGTACACCGCGGCA GTCCGGAGCTGAACACTCTTCAATT ACATTCTCTCCTGCAGCAGTACATG
I K E S Y T R G S P E L N T L Q L H S L L Q Q Y M>

*
AGGATTCTGATCGCGTTTTCACCT GCGGCTTCAGCATCGTTTCGGGTAT TCTGTCCCATTACAACGGGCAGAAC
R I P D R V F T C G F S I V S G I L S H Y N G Q N>

1600
*
AAGGCGCCCTACAAGCAGCGGCTTG TCTGGAAGCTGTCCAGTCGCACGCT GCGCGTCTTGCGCCCGATCAACCGC
K A P Y K Q R L V W K L S S R T L R V L R P I N R>

1700
*
TTCTCCTCCGACCTGCCACCATTG AGGAAGGCATCCCCAACGCCCTCGA CGATGGCCTGCGTTCTAGAACCAG
F S S D L P T I E E G I P N A L D D G L R S R T E>
PEST -2.1

1800
*
AGTATTAGCTCCGAGGAGGAGGAGG ACTGGCCCACTTCACCCATAATTCC AATTTTCGAACAATGTAGTATCCA
S I S S E E E E D W P T S P I I P I F E Q C *

*
GCGACAGCCAGCAGCATTAGAGCAG CAGGGTCAACAGCTGGGCAAACTG GAGCAGCGGACAAAAGGACCATCAA

1900
*
CCAGACGTAATCGCGGGACACACCT GACACCTCACACCTCCTGCTTTCCT GGCCTAGGCATATCCCATAGCAGC

```

z34 cDNA sequence, cont.

```

                                2000
                                *
AGCAGGAGGGTGGAAATCTCCGCAG AGGATGTGTGTCCGTGAGAGCGTTC CGCGTGCGAGTGTAGTAGTGCTGGT
                                *
                                2100
                                *
TGAGGTAGAGTGTGTATATTTTGCT AAAGTGCATCATAATTCTTTTGGCA TACACACAAATGTGTATTTGGTAGC
                                *
                                *
GCGCTGGTTCTAACATTTAAGAAAC TATTGAGATACGGAAATGTGGACGC CAAGGTGACGCAGCCACGCCACCCC
                                *
                                2200
                                *
CTTTTGAAAAGTTACCCGAGAGTAG CGCCCCACATTTTTTCGATTATT TGCGGCAGACAAAAAATACGTGAA
                                *
AAAAAAACAAAAGCAAAAAAAAAA AAAAAAAAAA

```

Figure 3.2

Sequence of the z34 cDNA. The DNA sequence is given above, and the conceptual translation below. The internal BamHI site used to generate the ΔB construct is shown as a horizontal bar. The PEST sequence is underlined and the PESTfind algorithm score given below.

z40 cDNA sequence

```

*                               *
AAAAAGAACACAAAAGTTATAATTC TGTCATTTAATGTGCTGCAAAATAG GACTTTCCAGCATCATTCATGTAG
100
*                               *
TTAAATCCAGATTTCTGAAAAACCT TATTGGCAGCGATATGTTAGCCA AAAAATAGGAATTTGCGGATAAATT
200
*                               *
ACCCCCCGGAGTGACAAAGATCTGT ATTGTCATACAGACGCCTGCAGAAA AAAGGGTTTACAAGAAGAAAATGC
300
*                               *
GCTATTTTGCAAAGGAATGATTTCT AGAATTCAACGTCAAATATAATGGA TAACATAGTCTACGACTTTGCCAAG
M D N I V Y D F A K>
*                               *
ATCACGTTTCAAGCAAAAGACAACA GGTCACCCCTACTAACTCGAATCT GTCGTGGCAACTAAATCAGATGGCT
I T F Q A K D N R S P T T N S N L S W Q L N O M A>
400
*                               *
TTGTCCGACATGGAGGAGATGCAGG ACACATCCGAGCCCATAGCTCCACC CGAATCCGATGACAATGTCAGCAGT
L S D M E E M O D T S E P I A P P E S D D N V S S>
PEST +13.4
500
*                               *
GAATCGCAAGACTCCGACGATGTGG ACTCGCAATTGAGTCGCTGCGAGGA CAACGATGACGACAGCGATTGCATC
E S O D S D D V D S O L S R C E D N D D D S D C I>
600
*                               *
AGTGGATCCTCCAGACGAGTTCCA CTTCTGGAGCTCGAGCGGGCGTGGC TCGTCGCAGAATGCCCCGAGGGTG
S G S S R R S S T S G A R A G V A R R R M P A R V>
700
*                               *
TCCAAGGACAACCTTTAACC GGATCT GGAGCGCCATCATGAAACCCATCAA AAAGAAGCAACGCAAAGAGCTGAAC
S K D N F N R I W S A I M K P I K K K Q R K E L N>
800
*                               *
ACAAATGCCCAAACCTTAAAGCA TCGAAAGGATCCACACCAGCAGGCG CATGAAAAAGTTACGCCCCACCAAT
T N A Q T L K S I E R I H T S R R M K K F T P T N>
900
*                               *
CTGGAGACAATCTTCGAGGAACCCA GCGATGAGAATGCCGCCGATGCAGA GGACGACAGCGAGGAGTGCTCCATC
L E T I F E E P S D E N A A D A E D D S E E C S I>
PEST +11.7
*                               *
AGCAGCCAAGTGAAAGTAGTTAAGG TGTGGGTCGCAAACTCCGCCGGGC AATATCCTTCAGCGATGGCCTGAAC
S S O V K V V K V W G R K L R R A I S F S D G L N>

```

z40 cDNA sequence, cont.

```

          *                               *                               *
AAGAACAAAATCCTGTCTGAAGAGAC GCCGCCAGAAGGTGAAGAAGACCTT TGGCAAGCGTTTCGCACTCAAGAAA
K N K I L S K R R R Q K V K K T F G K R F A L K K>

          1000
          *                               *                               *
ATCTCCATGACCGAGTTCCACGATC GTCTGAATAAGAGCTTCGACAGTGC CATGCTGGAGGGGGATGATGCCAGA
I S M T E F H D R L N K S F D S A M L E G D D A R>

          1100
          *                               *                               *
GGGCGGGAGGATCGGCGGAGGCCGT CAACATTCCCCAAGACATCCATGAC CATGGAGGACATACAGCTGCCGACA
G R E D R R R P S T F P K T S M T M E D I Q L P T>

          1200
          *                               *                               *
ATGAGCAGCCAGCACCAGTTCTTCA TGCAACCGCGGGCTTTGAGTAGAG AGACTGAATGATCCATCAAATACGC
M S S Q H Q F F M Q P A G F E *

          *                               *                               *
CCCACATTGATTTGCATTGCATTAA AACTAGGTAAATAGTGCCCAAAAAT AAATGTACTGATTACCAATTATGAA

          1300
          *                               *                               *
GTTAGGATTAACGTTTCGTTTGGTTA ACTTCTCACCTTAGTCTTAAGCCCC ATAAAAGTTATAAATGAGTGTAAT

          1400
          *                               *                               *
AGCATGTAGAAGAAAAGAAAATAAG AGCTATACCTAGAGCTAAACTTATC CAGCCATAGAATACGATTCTGTGCT

TAGCCATTAAGATAATAAATAAAA

```

Figure 3.3

Sequence of the z40 cDNA. Note: nt 1-360 are from the *Drosophila* EST 1032424, Genbank accession: AA391083 nt 360-1449 are from the two hybrid interacting clone, z40. The DNA sequence is given above, and the conceptual translation below. The caret denotes the first in frame amino acid of the z40 construct from the interacting clone. PEST sequences are underlined with the PESTfind algorithm score given below.

a

z34	1	V	T	A	G	G	R	D	G	S	A	Y	V	L	R	C	L	K	M	W	Y	E	L	P	S	D	V	L	F	S	A	M	S	L	V	D	R	F	L	D	R
A	1	I	S	H	N	M	R	S	I	L	I	D	W	L	V	E	V	S	E	E	Y	K	L	D	T	E	T	L	Y	L	S	V	F	Y	L	D	R	F	L	S	Q
B	1	V	S	H	K	M	R	A	V	L	I	D	W	I	N	E	V	H	L	Q	F	H	L	A	A	E	T	F	Q	L	A	V	A	I	I	D	R	Y	L	Q	V
C	1	E	Y	Q	K	V	F	I	F	F	A	N	V	I	Q	V	L	G	E	Q	L	K	L	R	Q	Q	V	I	A	T	A	T	V	Y	F	K	R	F	Y	A	R
D	1	I	T	P	P	M	R	K	I	V	A	E	W	M	M	E	V	C	A	E	N	C	Q	E	E	V	V	L	L	A	L	N	Y	M	D	R	F	L	S	S	
E	1	L	Q	P	R	M	R	A	I	L	L	D	W	L	I	E	V	C	E	V	Y	K	L	H	R	E	T	F	Y	L	A	V	D	Y	L	D	R	Y	L	H	V

z34	41	M	A	-	V	K	P	K	H	M	A	C	M	S	V	A	S	F	H	L	A	I	K	Q	L	D	-	-	L	K	P	I	P	A	E	D	L	V	T	I	S
A	41	M	A	-	V	V	R	S	K	L	Q	L	V	G	T	A	A	M	Y	I	A	A	K	Y	E	E	-	-	I	Y	P	P	E	V	G	E	F	V	F	L	T
B	41	V	K	D	T	K	R	T	Y	L	Q	L	V	G	T	A	L	F	I	A	T	K	Y	E	E	-	-	L	F	P	P	A	I	G	D	F	V	F	I	T	
C	41	N	S	-	L	K	N	I	D	P	L	L	A	P	T	C	I	L	L	A	S	K	V	E	E	-	-	F	G	V	I	S	N	S	R	L	I	S	I	C	
D	41	K	S	-	V	R	K	T	Q	L	Q	L	A	A	C	L	L	L	A	S	K	L	R	E	-	-	P	S	C	R	A	L	S	V	D	L	L	V			
E	41	A	H	K	V	Q	K	T	H	L	Q	L	I	G	I	T	C	L	F	V	A	A	K	V	E	E	-	-	I	Y	P	P	K	I	G	E	F	A	Y	V	

z34	78	Q	-	-	C	G	C	T	A	G	D	L	E	R	M	A	G	V	I	A	N	K	L	G	V	Q	M	G	H	A	P	I	T	S	V	S	Y	L	R	I	Y	Y
A	78	D	-	-	D	S	Y	T	K	A	Q	V	L	R	M	E	Q	V	I	L	K	I	-	-	L	S	F	D	L	C	T	P	T	A	Y	V	F	I	N	T	Y	A
B	79	D	-	-	D	T	Y	T	A	R	Q	I	R	Q	M	E	L	Q	I	F	K	A	-	-	I	N	C	N	L	S	R	P	L	P	I	H	F	L	R	R	Y	S
C	78	Q	S	A	I	K	T	K	F	S	Y	A	Y	A	Q	E	F	P	Y	R	T	N	-	-	H	I	L	E	C	E	F	Y	L	L	E	N	L	D	C	C	L	
D	78	Y	T	D	N	S	I	Y	K	D	D	L	I	K	W	E	L	Y	V	L	S	R	-	-	L	G	W	D	L	S	S	V	T	P	L	D	F	L	E	L		
E	79	D	-	-	G	A	C	T	E	R	D	I	L	N	H	E	K	I	L	L	Q	A	-	-	L	D	W	D	I	S	P	I	T	I	T	G	W	L	G	V	Y	

b

z34	1	V	T	A	G	G	R	D	G	S	A	Y	V	L	R	C	L	K	M	W	Y	E	L	P	S	D	V	L	F	S	A	M	S	L	V	D	R	F	L	D	R
mG2	1	L	C	S	R	L	R	N	A	K	V	E	D	L	R	S	L	T	S	F	F	G	S	G	T	E	T	F	V	L	A	V	N	I	L	D	R	F	L	A	
hG1G2	1	M	T	A	R	L	R	D	F	E	V	K	D	L	L	S	L	T	Q	F	F	G	F	D	T	E	T	F	S	L	A	V	N	L	L	D	R	F	L	S	

z34	41	M	A	V	K	P	K	H	M	A	C	M	S	V	A	S	F	H	L	A	I	K	Q	L	D	-	-	L	K	P	I	P	A	E	D	L	V	T	I	S
mG2	41	M	K	V	K	P	K	H	L	S	C	I	G	M	C	C	F	L	A	A	R	L	A	E	E	E	G	D	V	P	P	T	H	D	V	I	R	I	S	
hG1G2	41	M	K	V	Q	P	K	H	L	G	C	V	G	L	S	C	F	Y	L	A	V	K	S	I	E	E	E	R	N	V	P	L	A	T	D	L	I	R	I	

z34	79	C	G	C	T	A	G	D	L	E	R	M	A	G	V	I	A	N	K	L	G	V	Q	M	G	H	A	P	I	T	S	V	S	Y	L	R	I	Y	A
mG2	81	C	K	C	T	A	S	D	I	K	R	M	E	K	I	S	E	K	-	-	L	H	Y	E	L	E	A	T	T	A	L	N	F	L	H	L	Y	H	
hG1G2	81	Y	R	F	T	V	S	D	L	M	R	M	E	K	I	V	L	E	K	-	-	V	C	W	K	V	K	A	T	T	A	F	Q	F	L	Q	L	Y	

z34	119	F	R	N	L	A	K	E	I	G	G	D	F	F	K	F	Y	Q	Q	L	I	K	L	E	E	L	E	N	R	L	E	I	L	M	C	D	V	K	T
mG2	119	-	-	-	-	-	-	-	-	V	F	C	H	T	S	E	R	K	E	I	L	S	L	D	K	L	E	A	Q	L	K	A	C	N	C	R	V		
hG1G2	119	-	-	-	-	-	-	-	-	L	Q	E	N	L	P	L	E	R	N	S	I	N	F	E	R	L	E	A	Q	L	K	A	C	H	C	R	I		

z34	159	I	T	P	S	T	L	A	L	V	L	I	C	L	H	L	D	F	H	I	K	E	S	Y	T	R	G	S	P	E	L	N	T	L	Q	L	H	S
mG2	151	A	R	P	S	V	L	A	L	C	L	L	N	L	E	I	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
hG1G2	152	A	K	P	S	V	L	A	L	S	T	I	A	L	E	I	Q	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				

z34	199	Q	Y	M	R	I	P	D	R	V	F	T	C	G	F	S	I	V	S	G	I	L	S	H	Y	N	G	Q	N	K	A	-	P	Y	K	Q	R	L
mG2	182	K	H	L	K	L	S	D	T	E	F	F	Y	W	R	E	L	V	S	K	C	L	A	E	Y	S	S	P	R	C	C	K	P	D	L	K		
hG1G2	183	K	H	S	K	I	N	G	R	D	L	T	F	W	Q	E	L	V	S	K	C	L	T	E	Y	S	S	N	K	C	S	K	P	N	V			

z34	238	L	S	S	R	T	L	R	V	L	R	P	I	N	R	F	S	S	D	L	P	T	I	E	E	G	I	P	N	A	L	D	D	G	L	R
mG2	222	V	S	R	R	T	A	Q	N	L	H	S	S	Y	Y	S	V	P	E	L	P	T	I	E	G	G	C	F	D	G	S	E	S	E	D	
hG1G2	223	V	S	G	R	T	A	R	Q	L	K	H	S	Y	Y	R	I	T	H	L	P	T	I	E	M	V	P	-	-	-	-	-	-	-		

z34	278	I	S	S	E	E	-	E	E	D	W	P	T	S	P	I	I	P	I	F	R	T	M	L	V	S	S	D	S	Q	Q	H	
mG2	262	S	C	G	E	E	S	L	S	S	S	P	P	S	D	Q	E	C	T	F	F	F	D	F	Q	V	A	Q	T	L	C	F	P

c

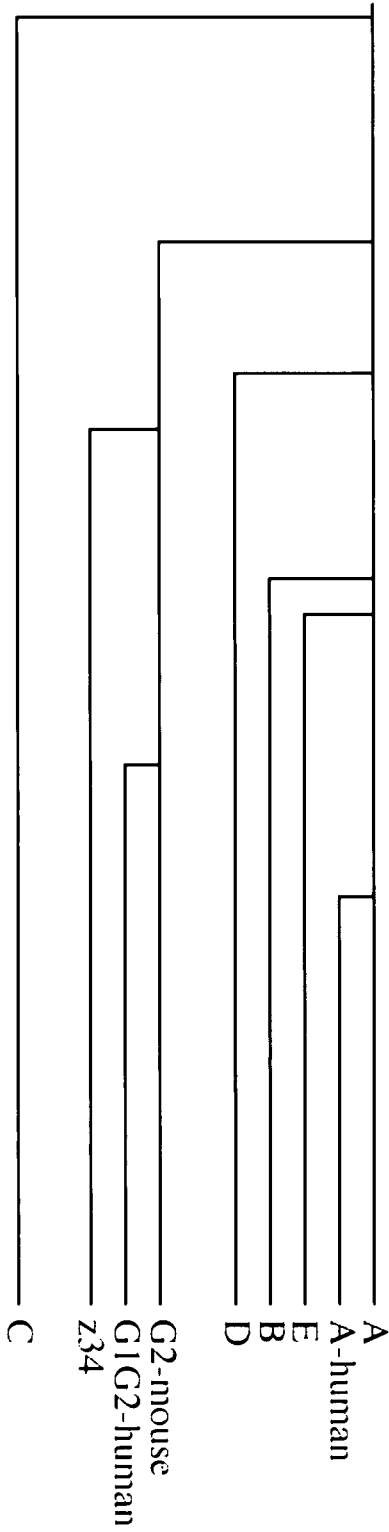


Figure 3.4

z34 Sequence Alignments. (a) Comparison of z34 cyclin box sequence to *Drosophila* cyclins A-E. Boxes surround sequence identities within the group. Shading represents sequence similarity to z34. (b) Comparison of z34 cyclin box and following sequence to G cyclins from mammals; h, human; m, mouse. Numbering begins at the first residue of the cyclin box for each sequence given. Arrowheads mark the conserved alanines at putative interhelical crossing points. Asterisks mark two charged residues critical for cyclin A-CDK contact; z34 has an alanine substitution for the second of these. Circles identify the positions of residues that make the up hydrophobic pocket of cyclin A responsible for binding the central portion of the PSTAIRE helix. The vertical bar separates the cyclin-box domain (on the amino side) from the C-terminus of the protein. (c) Phylogenetic tree of cyclin box sequences generated by the cluster algorithm. z34 was compared to *Drosophila* cyclins A, B, C, D, and E, as well as mammalian cyclins A and G.

alignments (Figure 3.4a and 3.4b) or the phylogenetic tree generated using cyclin-box sequences (Figure 3.4c) z34 is more closely related to mammalian cyclin G than to any known *Drosophila* cyclin. z34 can be divided into three sequence domains. The amino domain contains the poly-glutamine and poly-alanine stretches, the central domain is the cyclin box, and the carboxyl domain contains sequence that is conserved with cyclin G but to a lesser extent than the cyclin box. Mouse cyclin G2, the mammalian G cyclin most similar to z34, is 35% identical and 73% similar in sequence over the cyclin box, and 24% identical and 55% similar over the sequence following the cyclin box. A BamHI site lies at the junction between the amino domain and the cyclin box, and was used to generate a deletion construct, z34 Δ B, that contained only the amino domain. This construct interacted strongly with AsxC and AsxC1 (Table 3.1). Asx therefore interacts with sequences amino to the cyclin box.

The cyclin box folds into a 5 helix bundle with short inter-helical distances [136]. In particular, the tight packing between the second and third helices requires alanine residues at their crossing points. These are conserved in z34 (arrowheads, Figure 3.4-b). The cyclin A-CDK2 interface consists of many interactions, with the focal point being the PSTAIRE helix of CDK2 [136]. Key among these are four hydrogen bonds involving two conserved cyclin residues, a lysine and a glutamate (asterisks, Figure 3.4b). The lysine donates two hydrogen bonds to the PSTAIRE helix, the glutamate accepts one hydrogen bond, and there is an additional hydrogen bond between the lysine and glutamate themselves. Whereas these two residues are conserved in human and mouse cyclin G1 and G2 [137], z34 has an alanine substitution for the glutamate. In addition, the alanine is followed by glycine, suggesting that the pattern of interactions for a z34-CDK binding interface would be significantly different in this region from that of cyclin A-CDK2. This could mean that the z34-CDK interaction is weak, perhaps requiring a cofactor, or that the

CDK bound by z34 or z34 itself has compensatory substitutions that make up for the lack of hydrogen bonds involving this glutamate. It may also mean that z34 does not bind a CDK, however given the striking conservation of other residues throughout the cyclin box of z34, this is unlikely. Identified with circles in Figure 3.4b are the positions of residues that make the up hydrophobic pocket of cyclin A which binds the central portion of the PSTAIRE helix. The hydrophobic character is conserved in all but one of the five residues. While this nonconserved residue is a glutamate in z34, in the mouse and human homologues it is a leucine and a serine, respectively, indicating that even among the human cyclin Gs, this residue does not necessarily need to be hydrophobic.

An Asx-Cyclin G interaction?

Cyclin G was originally cloned accidentally in a low stringency screen for *src* family kinases from rat. Its sequence was found to be most similar to the A cyclins, however it was shown to lack a PEST sequence, destruction box, or any other sequence implicated in protein turnover, and to have relatively constant mRNA levels through the cell cycle in a rat cell line [138]. Mouse cyclin G was discovered in a screen for genes activated by p53 in a leukemic cell line [139]. Cyclin G mRNA levels increased in response to an increase in p53 expression levels or γ -irradiation (which induces p53) prompting the suggestion that cyclin G may regulate apoptosis. Two types of cyclin G were subsequently shown to exist in mammals, G1 and G2 [137]. Both are tissue specific: G1 is expressed at high levels in skeletal muscle, ovary and kidney, while G2 is expressed in cerebellum, thymus, spleen, kidney and prostate. In contrast to cyclin G1, cyclin G2 mRNA levels oscillate with the cell cycle, showing maximum expression in late S phase. The C-terminus of cyclin G2 contains a PEST sequence. z34 has a weak PEST sequence at its C-terminus, as determined by the PESTfind algorithm [140]. Cyclin G immunoprecipitates with the kinases CDK5 and GAK [141]. It also immunoprecipitates with the regulatory subunit of protein phosphatase 2A (PP2A-B' α), whereas other cyclins do not [142]. B' α is

predominantly nuclear and is thought to play a role in the translocation/localization of PP2A in the nucleus [143] making it likely that cyclin G is also nuclear (no data on the subcellular location of cyclin G is yet published.) Whether cyclin G is concurrently or alternately associated with a kinase and a phosphatase, and whether these associations are generally applicable or specific to particular cell types, is unknown.

Two questions arise: "Is z34 *Drosophila* cyclin G?" and "What is the significance of an Asx-cyclin (cyclin G) interaction?" The answer to the former is probably yes. Until we better understand the function and properties of mammalian cyclin G, and are able to determine the same for z34, one cannot be sure. While z34 is more similar to the G type cyclins than to any other known proteins, there may exist a protein in *Drosophila* that is more similar to mammalian cyclin G. As Figure 3.4c shows, z34 and human cyclin G2 are significantly more divergent than are *Drosophila* and human cyclin A. In addition, the alanine substitution at a critical PSTAIRE-interacting residue raises the possibility that z34 will not show the same CDK-binding activity as mammalian cyclin G. On the other hand, I have done BLAST searches for mouse cyclin G2-related sequences on the *Drosophila* EST database, and z34 ESTs (three as of February 10, 1998) are the most similar sequences in the database. Given that the cyclin boxes of cyclins A-E are all represented multiple times over in the EST database, it seems unlikely that there exists a more similar cyclin box sequence expressed in *Drosophila*.

The significance of an Asx-cyclin G interaction will remain somewhat unclear until the properties of cyclin G are better worked out. Assuming that Asx interacts with some type of cyclin, it would then recruit a CDK (or perhaps PP2A) to its chromosomal sites of action. At least three potential targets of such a kinase (or phosphatase) can be imagined. The first potential target is histone H1. Phosphorylation of histone H1 is a property of MPF, being accomplished by CDC2, and is in fact the classic test for CDK activity [144].

In the non-mitotic phases of the cell cycle, H1 phosphorylation is thought to be associated with opening up of chromatin for transcriptional competence [145]. An H1 kinase activity could then antagonize the predicted function of Asx, which is transcriptional repression. The second potential target is the RNA polymerase holoenzyme or initiation complex. There are in fact two CDKs already known to be associated with RNA polymerase: CDK7 and cyclin H are part of TFIIH [146; 147] and SRB10 and SRB11 are a kinase-cyclin pair in the RNA polymerase II holoenzyme itself, required to phosphorylate the carboxyl-terminal domain (CTD) of pol II [148]. A cyclin G-CDK pair brought to a site of transcription could enhance or potentiate transcription by helping the pol II-resident CDKs, while a cyclin G-PP2A pair could suppress or prevent transcription by antagonizing the same. Finally, the PcG proteins themselves are likely targets for a CDK brought in by Asx. Many PcG proteins, including Asx itself, have serine, threonine, or serine/threonine repeats. Although it has not been shown, regulation of PcG protein activity through phosphorylation of these repeats is an attractive possibility. The PcG proteins would be the most likely targets of a CDK for another reason, however. PcG proteins have a definite requirement to be sensitive to the phase of the cell cycle, and thus could take advantage of a cell-cycle phase dependent kinase activity. Whereas the transcriptional repression of a PcG target is maintained through many cell divisions, rendering the polymerase or chromatin at that target no different from those at any other locus in their requirement for input from the cell cycle, the PcG protein complexes must duplicate themselves with each cell division in order to ensure this maintenance. Phosphorylation events associated with S phase would be very useful in triggering a change in a given complex that would ensure its replication following cell division. Indeed, the behaviour of PcG proteins is reported to be cell cycle-dependent. Some stay at their chromosomal sites through nuclear division, while others are shunted out of the nucleus, only to return after cell division is complete [72]. Phosphorylation and dephosphorylation could be initiating these translocations. Those that stay behind must in some way allow polymerase to pass through their sites, and then

reform after replication, spreading to both daughter strands. Such spreading, either along a DNA molecule, or from one to another, would be disastrous if not restricted to the appropriate phase of the cell cycle.

The Asx-z40 Interaction

Although the sequence of z40 does not give any clues to its function, it is nevertheless the most salient of the interactors by virtue of the very strong interaction that it shows with Pc. This interaction addresses the question posed at the beginning of this chapter namely, "Can evidence be found supporting a physical connection between the complex containing ph, Psc, and Pc, and Asx?" The answer is yes, with z40 as the putative bridge. Evidence has since come out that Asx and Pc do colocalize. On polytene chromosomes, 64 of the 90 Asx binding sites reliably detected correspond to previously determined Pc binding sites, including the AntC and BxC loci [135]. Whether Asx or z40 are intimately associated with ph and Pc, or are peripheral remains to be determined. Given that the z40-Asx two-hybrid interaction is much weaker than the z40-Pc interaction, it may be that z40 is an integral component of complexes containing Pc, while Asx is more peripheral. Evidence has recently been obtained for z40 being required for maintenance of homeotic gene expression boundaries. Homozygotes for a deficiency that uncovers the z40 locus at 65A show extensive ectopic expression of the homeotic gene Scr (Tom Milne, personal communication). If this phenotype can be narrowed down to the z40 locus itself, z40 would become the newest *bona fide* member of the PcG, being required for homeotic gene regulation and showing an interaction with Pc.

Asx Interacts with the trx SET Domain

Several PcG genes have been shown to be modifiers of position effect variegation. Asx is an E(var), *E(Pc)* is a Su(var) [77], and a transgene expressing the human homologue of

E(z) is an *E(var)* [78]. This pointed to *Su(var)s* and *E(Var)s* as potential candidate binding partners of *Asx*. I therefore collaborated with Michael O'Grady to test his unpublished two-hybrid panel of *Su(var)s* and *E(var)s* for interaction with the three *Asx* constructs as both DNA-binding fusions and as activator fusions. The O'Grady panel consisted of LexA fusions to *Su(var)3-9* [149], and klett (discovered in a two-hybrid screen using *Su(var)3-9*) and activator fusions to klett, PP1 (the product of the *Su(var)3-6* locus) [150], and *trxC* (the C-terminal 553 amino acids of *trx* which has a domain of high similarity to *Su(var)3-9*, but behaves as an *E(var)*, Sarb Ner, personal communication). Two interactions were seen, both with Lex-*AsxC*. This fragment interacted weakly with the activator fusions to klett (which had independently been isolated as z38 in my library screen) and with the activator fusion to *trxC* (Figure 3.5a).

The *trx* interaction is very significant. In *PcG/trxG* double heterozygotes, homeotic transformations are suppressed, producing a wild type fly [51; 151]. The basis of this antagonistic behaviour is not understood at the molecular level. The *Asx-trx* interaction is the first evidence of a direct protein-protein interaction between the *PcG* and the *trxG*.

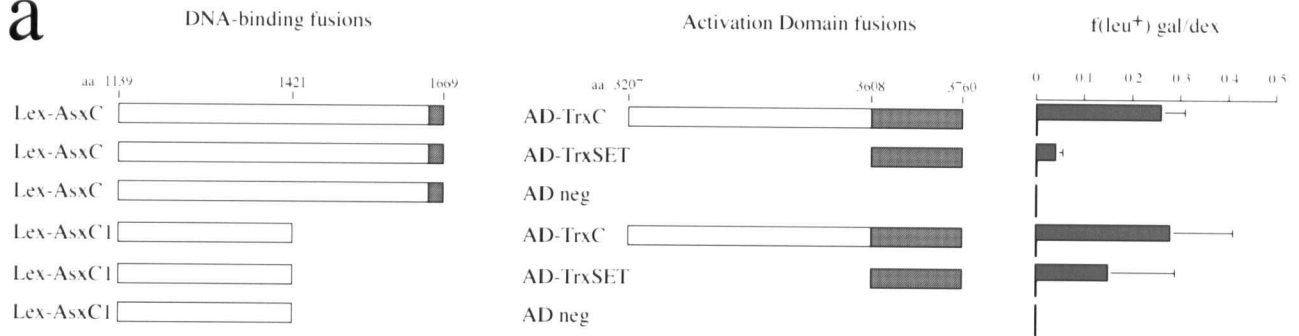
The *trxC* construct contains a sequence motif known as the SET domain. This domain is shared in common with another *trxG* gene, *ash-1* [62], the *PcG* gene, *E(z)* [31], and *Su(var)3-9* [149] from *Drosophila*. It is also present at the C terminus of ALL-1, the human homologue of *trx* [152-154]. A clue to the function of the SET domain comes from the yeast gene, SET1 [155]. Strains lacking this SET domain-containing gene are defective in telomeric silencing. This defect is corrected by expressing a mini-gene that consists of the SET domain alone. It is also corrected by expressing h*E(z)*, the human homologue of *E(z)* [78]. Another clue comes from ALL-1. Chromosomal translocations involving ALL-1 are seen in 80% of cases of infantile acute lymphoblastic leukemia [156] and are involved in many other leukemias [153]. These rearrangements replace the C terminus of the

protein, including the SET domain, with in-frame sequence from a variety of other genes (12 to date) [157]. Many of these other genes have been shown to be transcriptional activators [158; 159]. In fact a recombinant gene composed of the amino terminus of ALL-1 and a minimal transcriptional activation domain is sufficient for cellular transformation *in vitro* [160]. The final clue comes from the *trx*C construct itself. As a LexA fusion, this construct strongly activates transcription from LexA reporter genes (Figure 3.5e). These observations point to the SET domain being involved in regulating both chromatin and transcription, although its presence in both enhancers and suppressors of position effect, and activators and repressors of transcription makes it difficult to speculate what its mode of action is likely to be.

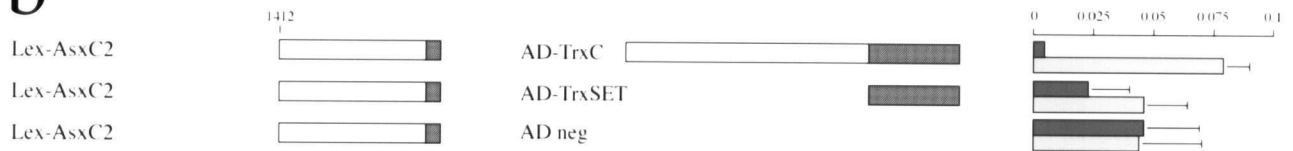
To further refine the interaction between *Asx* and *trx*, I generated three smaller constructs: *AsxC1* (aa 1139-1420), *AsxC2* (aa 1412-1669), and *trxSET* (the SET domain alone, aa 3608-3759). Figure 3.5a shows the *Asx*-*trx* interaction with the constructs Lex-*AsxC* and Lex-*AsxC1* assayed by measuring the frequency of leucine prototrophs (see below). Both interact strongly with *trx*C, and with *trxSET*, although in the case of Lex-*AsxC*, the latter interaction is significantly weaker than the former. In both cases, transcriptional activation is dependent on galactose (compare the upper bar with the lower bar in the graph for each experiment) meaning that they require expression of the activator fusion. These data point to an interaction between the SET domain and *AsxC1*.

Unfortunately, Lex-*AsxC2* and Lex-*trxSET* both activated transcription alone, although not as strongly as *trx*C, and so could not easily be used to assay for interaction. In an attempt to circumvent the problem of transcriptional activation, I developed a limiting dilution assay to determine the frequency of cells that were converted to leucine auxotrophy. By comparing the number of colony forming units in a given drop spotted on leucine deficient galactose and dextrose plates with the number spotted on leucine supplemented plates, I

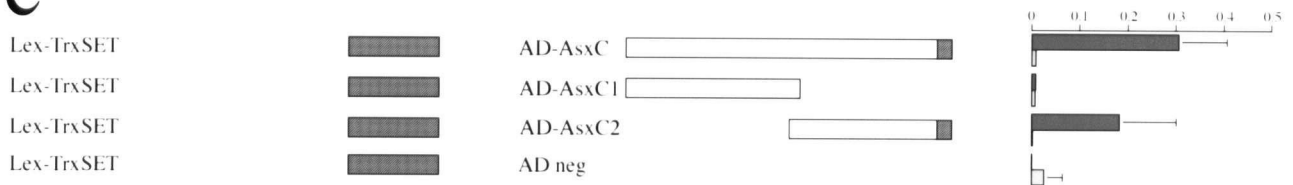
a



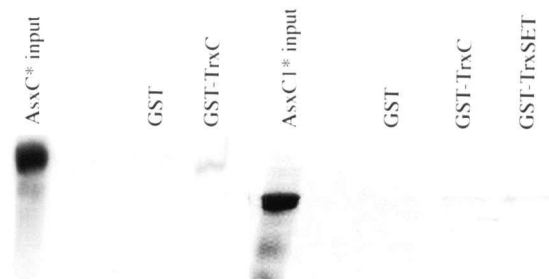
b



c



d



e

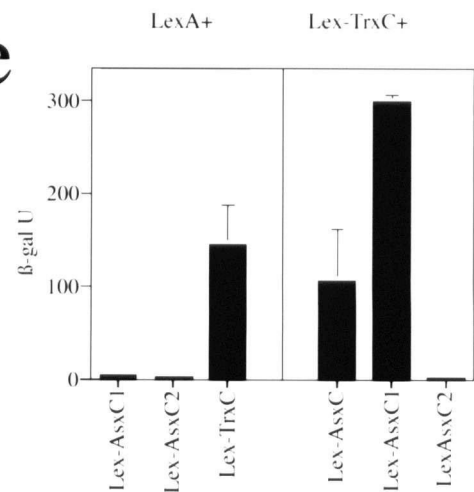


Figure 3.5

Asx-trx interactions. (a) Two-hybrid interactions between non-activating LexA-Asx fusion constructs and activator fusions to trx constructs. The graph shows the frequency of prototrophs on galactose (upper bar) and dextrose (lower bar). Dextrose shuts off transcription of the activator fusion. AD neg is the activator fusion plasmid with no insert. (b) Interactions between LexA fusions to AsxC2, which activates alone, and activator fusions to trx constructs. (c) Interactions between Lex-trxSET and Asx activator fusions. (d) *In vitro* co-affinity precipitation with glutathione agarose of reticulocyte lysate *in vitro* translated Asx constructs and bacterially produced GST fusions to trx constructs. (e) β -gal assays to monitor transcription in the presence to two different LexA fusions.

could measure the relative effect on transcription caused by the expression of the activation domain fusion. With LexA-fusions that activate transcription alone, the self-activation may be unaffected by the binding of the second fusion, in which case activation levels will go up because now there are two activation domains at the operator, or the self-activation may be blocked by the binding of the second fusion, in which case the activation levels may go down, depending on the relative contributions of the first activation domain (which is now blocked) and the second, which is brought in by the interaction. However one would not expect activation levels to drop to zero, because even if the activation activity of the LexA-fusion is completely blocked, the second fusion still carries its own activation domain.

Figure 3.5b shows the results of interactions with the Lex-AsxC2 construct. Activation caused by this fusion is actually reduced 18 fold by coexpression of the trxC activation fusion, and somewhat reduced (two fold) by coexpression of the SET domain activation fusion alone. This suggests that the AsxC2 can also interact with trxC, and that this interaction reduces the ability of AsxC2 to activate transcription. Whether the residual transcription in the presence of coexpressed trxC is due to residual activity of the Lex-AsxC2 construct, or due to the activation domain of the trxC activation fusion, or a combination of both, is impossible to tell. In any case these data suggest that AsxC2 can also interact with trxC, although the SET domain itself may not be sufficient for this interaction (the 50% reduction is not statistically significant).

Figure 3.5c shows the converse experiment, using activation domain fusions of Asx constructs and a LexA fusion of trxSET. In this conformation, trxSET-induced transcription is strongly enhanced with AsxC, as expected (upper panel) and also with AsxC2 (panel 3). The interaction with AsxC2 suggests that the statistically insignificant interaction from 5b, above, is the result of a real interaction, in other words that the AsxC2 domain can in fact interact with the SET domain. Unexpectedly, no enhancement or

suppression was seen between *trxSET* and *AsxC1* (panel 2). However a lack of interaction in this assay is not strong proof of the absence of a physical interaction, since there are still colonies that come up on leucine deficient medium. In other words, there is transcription at the reporter locus. Depending on the interaction between the two transcriptional activators at the locus, transcription with both present where one may be blocking the other need not necessarily be at a different level than transcription with only one present.

To ascertain whether these interactions could be due to direct protein-protein contact, I performed an *in vitro* GST-fusion protein binding assay. *In vitro* translated *AsxC* binds to GST-*trxC* coupled glutathione agarose, but not to glutathione agarose coupled to GST alone (Figure 3.5d, panel 1). *In vitro* translated *AsxC1* binds to both GST-*trxC* and GST-*trxSET* but not to GST alone (Figure 3.5d, panel 2) confirming that the minimally defined interaction, that between *AsxC1* and *trxSET* can occur directly.

The Transcriptional Consequence of an *Asx*-*trx* Interaction

The genetics of *Asx* suggest that the interaction demonstrated above may have transcriptional consequences. One *Asx* allele, *Asx^{P1}*, shows anterior homeotic transformations typical of the *trxG* in addition to posterior transformations typical of the *PcG* [19]. Recent work has now shown that this allele when homozygous, in addition to enhancing the posterior transformations of *Pc/+* flies, also enhances the anterior transformations of *trx/+* flies. In addition, several *Asx/trx* double heterozygotes show enhanced anterior transformations (Tom Milne, personal communication). This is in opposition to the usual behaviour of *PcG/trxG* double heterozygous allelic combinations, which is cosuppression. One could think of *Asx* as behaving as a member of the *PcG* with respect to interactions with other *PcG* genes, but as a member of the *trxG* with respect to interactions with *trx*. While the data from Figure 3.5b and c are a result of transcriptional interactions, because one of the constructs is fused to an exogenous activating sequence,

they do not isolate the influence of Asx and trxC sequences *per se* on one another. I therefore generated a second set of LexA fusions using a vector with a different selectable marker, allowing the coexpression of two LexA-fusions in the same cell. Since Lex-trxC activates transcription, coexpressing Lex-AsxC constructs gives a measure of the transcriptional interaction between Asx and trxC, together at one locus. Figure 3.5e shows that the AsxC1 and AsxC2 constructs have different effects on trxC-induced transcription. Lex-AsxC1 is capable of enhancing trxC-mediated transcription to levels approximately twice those seen with coexpression with LexA alone or Lex-AsxC, while Lex-AsxC2 abolishes trxC-mediated transcription completely. Since there are 5 LexA sites at the operator, the most likely interpretation of these data is that both proteins are present together at the operator, and that Asx is blocking or enhancing trxC through a physical association. This is backed up by the evidence from Figure 3.5b that AsxC2 can reduce trxC-induced transcription even when fused to an activation domain instead of a DNA-binding domain. However, an alternative explanation would be that the LexA fusions compete for the operator, and that Lex-AsxC2 always wins. Such a competition bias could be the result of trxC sequences occluding LexA, somewhat weakening its DNA binding ability, or it could be a result of a physical interaction between trxC and AsxC2.

Although the system is heterologous, these data are compelling evidence of a functional interaction between Asx and trxC in transcription. Considered together with the genetics of Asx, they suggest a model whereby Asx is a component of the system that integrates the repression signal of the PcG with the activation signal of the trxC. Depending on what inputs are received from other upstream proteins, Asx can either allow or suppress transcription mediated by trxC at the various homeotic loci, this allowance or suppression being dependent on different domains of Asx. Furthermore, they suggest a mechanism for the oncogenic transformation caused by ALL-1-transcriptional activator fusions. If the SET domain-containing C-terminus of ALL-1 is a transcriptional activation domain that is

subject to regulation by hAsx, then replacement of that domain with a different transcriptional activation domain that is not subject to regulation by Asx would generate a protein that activates transcription inappropriately at various loci, including loci that govern cell division.

Chapter IV: PcG Functional Interactions in Yeast

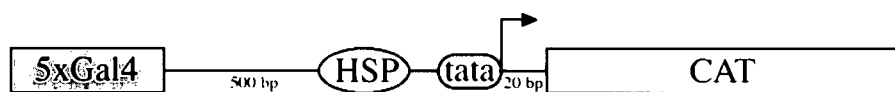
The PcG had been shown to be essential for transcriptional repression of the homeotic genes of the *BxC* and *AntC* by genetic means, but at the outset of this work there was no evidence that this repression was direct. It had been shown that *ph* and *Pc* were present on salivary gland chromosomes, but since they were present near genes that were repressed (the *BxC*) as well as near genes that were expressed (*ph* and *Pc*) it was not known whether their presence alone near a target gene was sufficient for transcriptional repression. Early on in this work, I demonstrated that *Pc*, *Psc*, and *ph* could repress transcription directly, simply by being targeted to a locus of transcription (see below). Having created a panel of PcG proteins for expression in yeast, I sought to use this panel to develop a functional assay system in yeast, with the long term goal of identifying non-PcG factors from yeast that are required for transcriptional repression by PcG proteins. The ability of these proteins to repress transcription in a heterologous system would mean either that they require no cofactors to repress, or that required cofactors are sufficiently conserved between flies and mammals to allow their functional interaction with the *Drosophila* PcG proteins. If such cofactors could be identified, the mechanism of silencing by the PcG could be characterized as similar to one or another of the well defined silencing or repression systems of yeast.

Pc, Psc, and ph Can Repress Transcription Directly

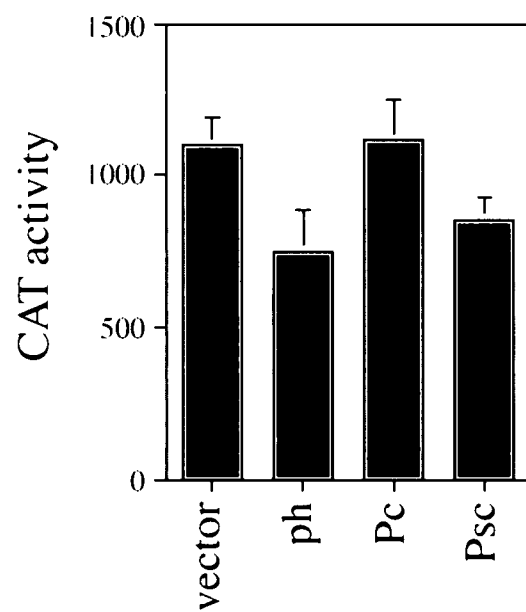
I set up a producer/reporter two plasmid system for use in the *Drosophila* tissue culture cell line SL2 [161]. The reporter plasmid consisted of the bacterial chloramphenicol acetyl transferase (*CAT*) gene under the control of the HSP70 promoter and upstream heat shock response elements, with 5 Gal4 binding sites 500 bp upstream (Figure 4.1a). Gal4 alone, Gal4-PcG fusions, and PcG nonfusions were expressed constitutively by the actin 5C promoter. SL2 cells were transiently transfected with both reporter and producer plasmids

a

SL2 CAT reporter



b



c

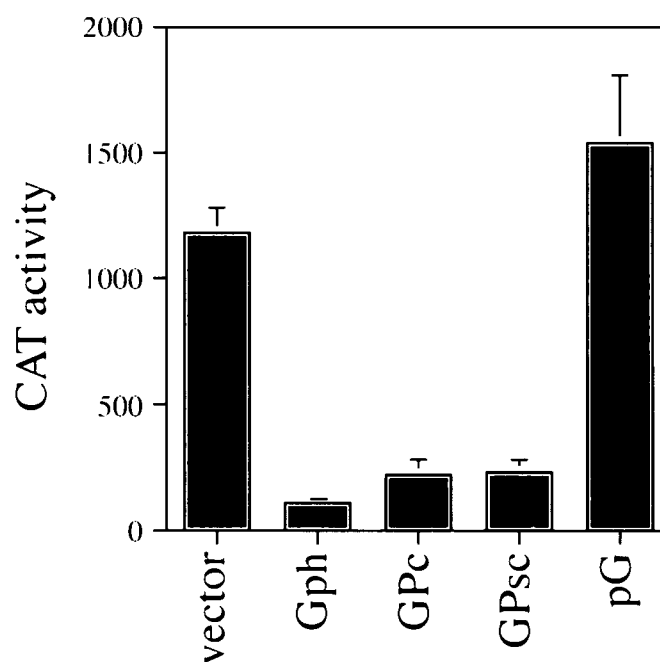


Figure 4.1

Transcriptional repression by PcG proteins in transiently transfected SL2 cells. (a) Schematic diagram of the reporter construct. (b) CAT activity in heat-shocked SL2 cells cotransfected with the reporter plasmid and producer plasmids that express only full length cDNAs of ph, Pc, and Psc. Vector: producer plasmid with no insert. (c) CAT activity in heat-shocked SL2 cells cotransfected with the reporter plasmid and producer plasmids that express Gal4 (pG), or Gal4-PcG fusions.

together, subjected to heat shock, and assayed for transcription of the *CAT* gene. As Figure 4.1b and c show, ph, Pc, and Psc are all able to repress heat-shock activated transcription when targeted upstream of the HSP promoter via Gal4 fusion, but not when expressed as nonfusions. Bunker and Kingston have since done similar experiments in mammalian cells [162] and shown that LexA fusions to Pc, Psc, and Su(z)2 can repress transcription induced by a variety of activators.

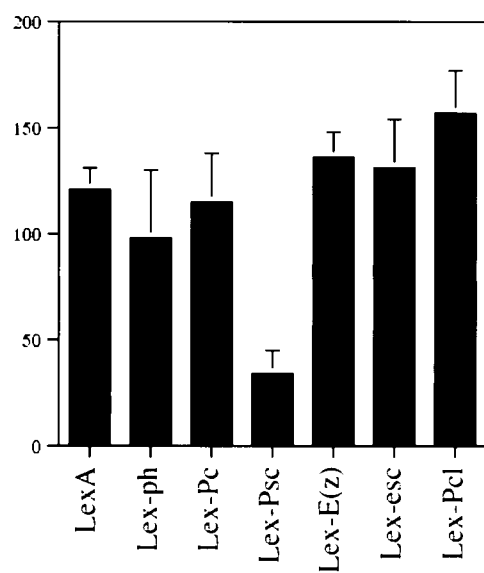
Psc is a Transcriptional Repressor In Yeast

Having yeast expression constructs for LexA-PcG fusions in hand, I tested them for transcriptional repression in yeast. The reporter plasmid in this case, JK1621 [163], has the *lacZ* gene under the control of the constitutive CYC1 promoter, with 5 LexA sites upstream. As shown in Figure 4.2a, Lex-Psc is able to repress transcription from the yeast CYC1 promoter, while Lex-ph, -Pc, -E(z), -esc, and -Pcl are not. Lex-Psc is also able to repress transcription from an integrated CYC1 reporter with upstream LexA sites (Figure 4.2b).

If Psc represses transcription in yeast in a manner homologous to its repression in the fly, then this demonstration of function in yeast could be an important step forward in understanding the mechanism of PcG transcriptional repression, the reason being that transcriptional repression and silencing are well studied and better understood in yeast. In an attempt to identify other factors necessary for repression by Psc, I repeated the two plasmid repression assay in a variety of mutant strains defective for components of known silencing or repression systems.

The products of the *SIR* genes are required for transcriptional silencing of the silent mating type loci [84] and telomeres [85]. When assayed in *sir*⁻ strains, Lex-Psc behaved as it did in wild type strains: transcription in the presence of Lex-Psc was always reduced relative to

a



b

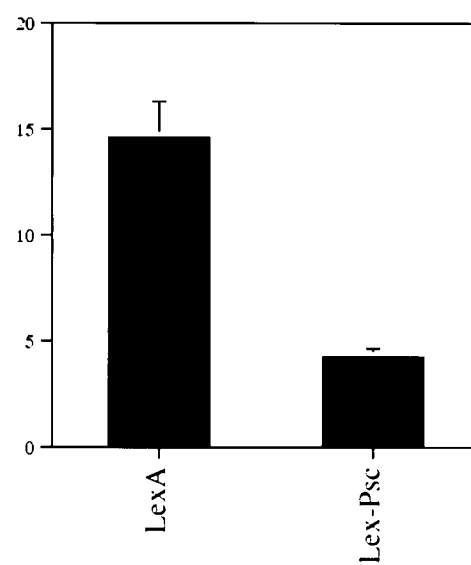
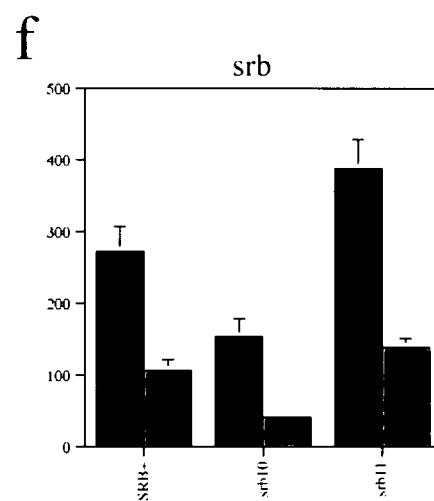
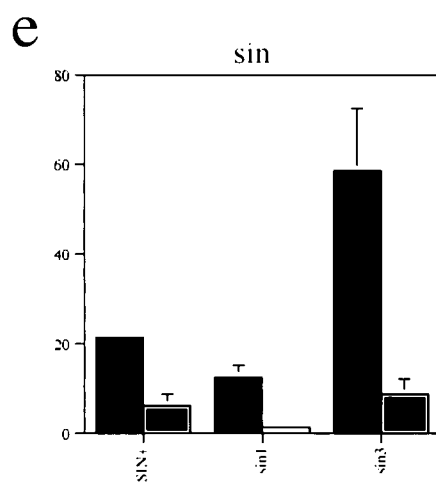
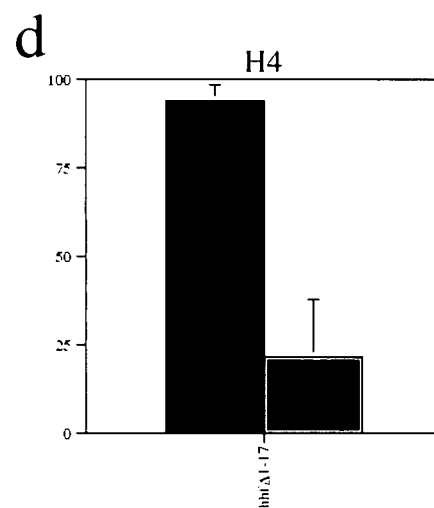
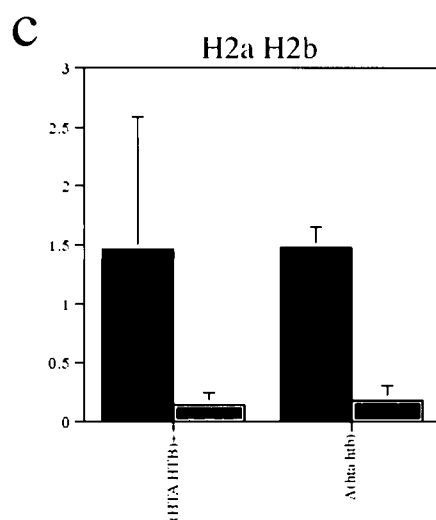
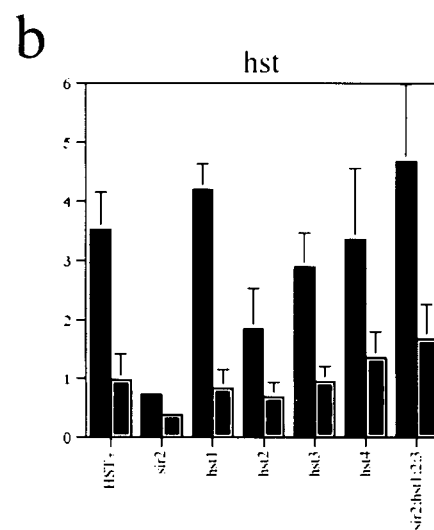
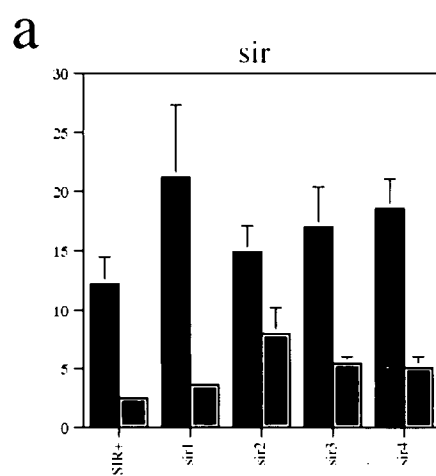


Figure 4.2

Transcriptional repression in yeast by Lex-Psc. (a) LexA and various full length PcG-LexA fusions monitored for their effect on transcription of a plasmid bearing a constitutive promoter with upstream LexA binding sites. (b) Transcription from a chromosomally integrated reporter gene driven by the CYC1 promoter with upstream LexA binding sites in the presence LexA or Lex-Psc.



LexA
Lex-Psc

Figure 4.3

Transcriptional repression by Lex-Psc in various mutant backgrounds. The left bar in each case is the level of reporter gene expression in the presence of LexA; the right bar is the level in the presence of Lex-Psc. (a) *sir* mutants and isogenic wild type. (b) *hst* (Homologous to Sir Two) mutants, including *sir2* and isogenic wild type. (c) Histone H2A,H2B deficiency (which reduces the dose of H2A and H2B by half) and isogenic wild type. (d) Histone H4 (One copy of *hhf* has been deleted, the other has been replaced by the silencing-defective mutant H4 Δ 4-29). (e) *sin* mutants and isogenic wild type. (f) *srb* mutants and isogenic wild type.

LexA alone (Figure 4.3a). However in the case of *sir2*, repression was not as strong as in the other *sir* mutants, or in the isogenic wild type strain.

SIR2 has four known homologues in *S. cerevisiae*, three of which are involved in telomeric silencing (HST1,3,4), and one of which (HST1) when overexpressed can complement *sir2* [164]. Because of the potential mitigation of Psc repression seen in *sir2*, I repeated the assay in *hst* mutant strains (Figure 4.3b). Individually, and in the multiple mutant, *sir2; hst1; hst2; hst3*, none of the *hst* mutant genotypes had significantly different levels of repression, relative to an isogenic wild type genotype. Intriguingly, repression in the *sir2* single mutant strain was again somewhat less severe than that seen in any other strain. The significance of this weak reduction is questionable however, given that it was not seen in the *sir2; hst1; hst2; hst3*, quadruple mutant.

Intact histones are required for silencing at the silent mating type loci and at telomeres, however neither a dose reduction of H2a/H2b nor a silencing-abolishing mutation of the amino tail of H4 prevented repression by Lex-Psc (Figure 4.3c,d). It would appear from these data that the repression due to Psc is not being mediated by the factors that are responsible for the maintenance of silent loci in *S. cerevisiae*. Since silenced loci in yeast are permanently devoid of transcription, it may not be surprising that Psc, which regulates loci that may be on or off, does not make use of silencing factors. Such silenced loci are more akin to transcriptionally inert heterochromatin than to the homeotic loci that are regulated by the PcG.

There are many factors that have been shown necessary for repression of regulated loci in *S. cerevisiae*. These fall into two broad classes: sequence-specific factors that are activated conditionally, and general factors that are recruited by activated sequence-specific factors. It is the latter class that is capable of shedding light on the mechanism of PcG silencing.

The products of the *SIN* genes are required for repression of the *HO* locus and other genes [165]. I tested Lex-Psc repression in *sin1* and *sin3* mutants, and found that neither of these mutants relieved repression by Psc (Figure 4.3e). Finally I tested two components of the pol II holoenzyme, SRB10 and SRB11, that mediate transcriptional repression [166; 167]. These also had no effect on repression by Psc (Figure 4.3f).

Before ruling out the involvement of yeast cofactors, several other general repressors would need to be tested, however until such a cofactor is found, it remains a possibility that Psc is repressing transcription without the help of yeast cofactors. In order to prevent transcription, Psc would need to do more than merely bind upstream of a transcribed gene. Perhaps Psc is able to assemble into a higher order homomultimer that occludes the binding of activators to the *CYC1* promoter.

Telomeric Effects of the PcG and *trxG* in Yeast

Yeast telomeres are transcriptionally silent, but are derepressed by mutations in histones, SIR genes, the telomeric repeat-binding protein RAP1, and select other factors. SIR3 enhances telomeric silencing when tethered to a telomere via LexA in a strain with telomeric LexA binding sites [168]. Telomeric silencing in this strain can be assayed by monitoring variegation of expression of two sub-telomeric reporter genes: *ADE2*, which when not transcribed gives rise to red colonies, and *URA3*, which when expressed causes toxic sensitivity to the purine analogue, 5-fluoroorotic acid (FOA). Using this same strain, I assayed telomeric silencing in the presence of my panel of LexA fusions. Despite the lack of a demonstrated interaction for Psc, if any protein from the panel interacted with silencing factors, such an interaction could enhance or disrupt (perhaps in a dominant negative manner) the function of these factors at telomeres.

The PcG proteins tested were ph, Pc, Psc, E(z), AsxA, AsxC1, and AsxC2. I also tested the *trx* constructs, *trx*C and *trx*SET. In the case of the PcG, none of the LexA fusions enhanced or disrupted telomeric silencing. The *trx* constructs, however, did have a measurable effect. Compared to LexA alone, Lex-*trx*SET enhanced telomeric silencing moderately, expanding the extent of red sectoring and increasing the resistance of the strain to FOA. The larger construct, *trx*C reduced telomeric silencing in both assays (Figure 4.4). The enhancement of silencing with targeted *trx*SET is significant in light of the fact that mutants in the yeast *SET1* gene cause derepression of telomeres and that this derepression is corrected by expression of the minimal SET domain of *SET1* [155]. It would appear from the behaviour of Lex-*trx*SET that the SET domain is capable of interacting with and enhancing the activity of silencing factors. The opposite behaviour of the larger construct, *trx*C, which contains the SET domain, demonstrates that silencing enhancement by the SET domain can be overcome by other sequences. It is possible that in this case, the other sequences are interacting erroneously, generating a dominant negative disruption of silencing. However, at the chromosomally integrated *LEU2* reporter gene, *trx*C acts as a strong activator (chapter 3). The blocking of silencing by *trx*C may then be due to its transcriptional activation potential overcoming the repressive environment of the telomere. Whatever the reason for *trx*C activating transcription and blocking silencing in yeast, in the fly *trx* is a bona fide activator, and as such, it is puzzling that it would contain a domain associated with silencing. It may be that the transcriptional activation activity of the C-terminus of *trx* is directed against nucleosomes or other higher order silencing complexes, and that the SET domain, with its ability to interact with silencing factors is the Trojan horse that brings *trx* to the silencing complexes that are slated for disruption.

E(z) has a SET domain, but had no effect on telomeric silencing. Since the minimal SET domain of E(z) was not tested, one can not directly compare the behaviour of *trx* with E(z). However if the SET domain is a protein module that recognizes silencing factors, E(z) may

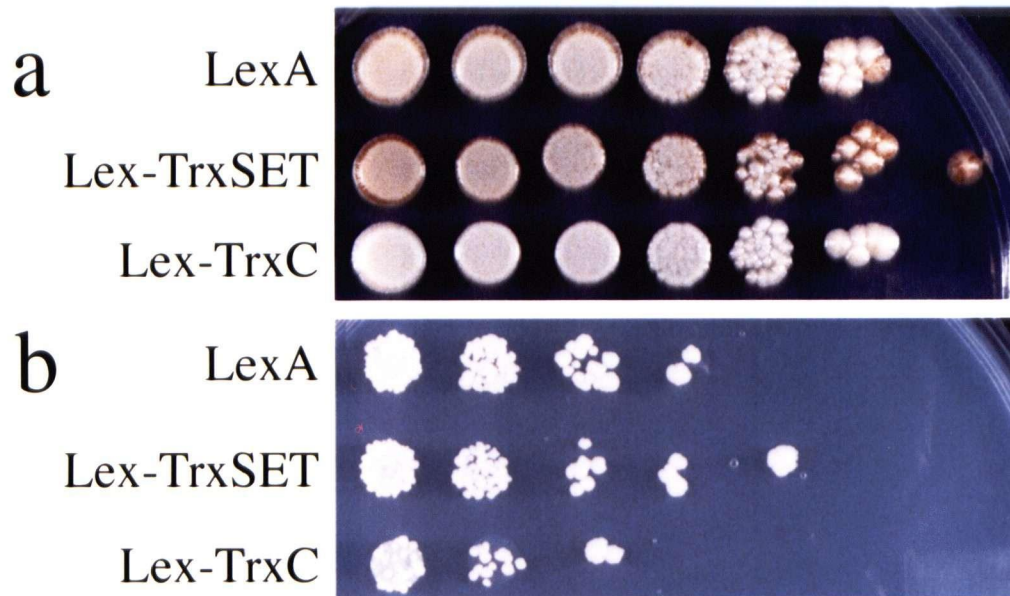


Figure 4.4

Telomeric silencing effects of *trx* sequences. (a) A 10x limiting dilution series of the telomeric reporter strain bearing various LexA fusions plated out on medium lacking FOA. Variegated repression of the telomere manifests as red colony sectoring. (b) The identical 10x limiting dilution series plated out on medium containing FOA. Colony growth is a measure of telomeric repression.

be using it to interact with these factors at the homeotic loci, helping to lock in the silenced state of this chromatin domain. In such a model of SET domain function, the genetic result that *E(z)* mutants enhance *ash-1* and show *trxG* phenotypes in some tissues at certain times [169] would mean that *E(z)* can also unlock a silenced domain using the SET domain as a key.

Implications for the Mechanism of PcG Action

The silencing assay and the repression assay appear to be giving two different results. In the silencing assay, one protein, *trx*, has the ability to interact with yeast silencing factors. In the other assay, *Psc*, which represses, does so independently of silencing factors. One interpretation would be that the mechanism of activation by the *trxG* is fundamentally different and unrelated to the mechanism of silencing by the PcG. To be rigorous though, the results tell us only that the mechanism of *trx* function in yeast is distinct from the mechanism of *Psc* function in yeast. If the paradigm is valid, these results actually invite the hypothesis that PcG/*trxG* repression/activation makes use of at least two distinct mechanisms, one involving histones, higher order chromatin structures, silencing factors and *trx*, the other which is mediated by *Psc*, not involving these. One group of PcG proteins may be involved in silencing, while another group may be involved in repression. The homeotic genes of the *BxC* and *AntC* do occupy a curious regulatory niche, requiring permanent silencing in certain tissues in order for determination to hold, yet not requiring ubiquitous silencing as do the silent mating type loci, which are never to be transcribed under any circumstances. In the years before the proliferation of PcG genes and the mass action/multimeric complex model of Locke *et al.* [28], the genetic interactions within the PcG were taken as evidence for multiple independent pathways to repression of the homeotic loci [74]. It may be time to revisit this interpretation.

Chapter V: Materials and Methods

Subcloning

I used linker PCR to generate appropriate restriction enzyme sites at the 5' and 3' ends of subcloned fragments. PCR products were digested, ligated into pBluescript, and sequenced to confirm the absence of PCR-induced mutations. They were then subcloned into pET28a (Novagen), pGEX-4T-1 (Pharmacia), pEG202 [110], pJG4-5 [170], and pBTM116 [171] as EcoRI/XhoI fragments. In the case of pBTM116, the XhoI site was ligated into a SalI site, destroying both. pBTM-ph, Pc, and Psc were exceptions, using instead EcoRI/BamHI digestion; see below. The standard PCR reaction contained the following: 1μg template, 0.5μL 10mg/mL acetylated BSA, 5μL 10x buffer (NEB), 0.7μL 25mM dNTPs, 1μL 100mM MgSO₄, 1μM each primer, 1μL (2 units) Vent polymerase (NEB), and H₂O to make 50μL final volume, overlaid with 50μL mineral oil. The temperature cycles were: 5 minutes at 95°C, 2x(1 minute at 4°C, 1 minute at 72°C, 1 minute at 95°C), 7x(1 minute at 45°C, 1 minute at 72°C, 1 minute at 95°C). I used a low number of cycles with a large amount of template (and an error-correcting polymerase) in order to minimize the chance of PCR-induced mutagenesis.

The template cDNAs used for these constructs were proximal *ph*: c4-11 [172], distal *ph*: c4-7 [99], *RAE28*: RAE-28[101] provided by Kazunori Shimada, *BEB1*: GSTboi2P[126] provided by Yasushi Matsui, *Pc*: Pc-12c [64] provided by Renato Paro, *Psc*: PscIIIA [104] provided by Paul Adler, *E(z)* e32 [173] provided by Richard Jones, *Asx*: Asxfl provided by Don Sinclair.

Pc: Primers Pc5 and Pc3 were used to generate the full length Pc EcoRI-ATG/BamHI fragment. This fragment was subcloned into EcoRI/BamHI digested pBTM116 to create pBTM-Pc. ΔchrPc was created using the primers Pc208f and Pc3. The minimal chromobox-containing fragment was generated with the primers chr5 and chr3.

Psc: I used the primers Psc5 and Psc3 to create the EcoRI-ATG/BamHI fragment Psc Δ B which contains amino acids 1-696. Full length Psc was created in all subsequent constructs by ligating the BamHI/BamHI fragment from the Psc cDNA into the BamHI site of Psc Δ B. Psc constructs designated Δ N were deleted for the 3' sequence following the NotI site (corresponding to amino acid 1460) by NotI digestion and religation, which liberated a NotI fragment, and those designated Δ S were deleted for the 3' sequence following the SalI site (corresponding to amino acid 205) in the same way. I created PscHD with the primers Psc748f and Psc1149r, ring with the primers Psc748f and Psc1005r, and HTH with the primers Psc1006f and Psc1149r.

ph: An EcoRI site was generated directly upstream of the first ATG of ph by PCR with the primers ph5 and ph255r. This EcoRI/XhoI fragment replaced the 5' Eco/Xho fragment of c4-11 (full length proximal ph cDNA.) ph contains a BamHI site 3 codons before the stop codon. ph was subcloned as an EcoRI/BamHI fragment. ph constructs designated Δ N retain amino acids 1-1418, and are deleted for the 3' sequence following the NcoI site by NcoI digestion and religation, which liberated an NcoI fragment. Those designated Δ S retain amino acids 1-522, and are deleted for the 3' sequence following the first SalI site in the same way. I created phHD using the primers phD5 and phD3. I created H1 by digesting phHD with NcoI, which liberates 3' sequence (corresponding to amino acid 1418 ff.) as an NcoI fragment, and recircularizing the plasmid. The construct H2 was created by cutting pETphHD with EcoRI and NcoI to remove the intervening sequence, filling in with Klenow, and religating, which regenerated the EcoRI site in the correct reading frame, as any fool can see. H2 Δ C was created by PCR with the primers H2f and MDL. dH2 was created by PCR with the primers H2f and H2r, but with the distal cDNA, c4-7. pSAM was created with the primers SAMf and H2r; dSAM was created using the same primers, but with the distal cDNA, c4-7.

Scm: The constructs Scm2 which encodes a GST fusion to amino acids 767-877 of Scm and sSAM which encodes a pET fusion to amino acids 797-877 of Scm were provided by Jeffrey Simon.

Pcl: The construct Lex-Pcl which encodes a fusion of full length Pcl to LexA, was provided by Rob Saint.

BEB1: The EcoRI/XhoI BEB1 SAM fragment was created with primers BEBf and BEBr.

RAE28: The EcoRI/XhoI RAE28 SAM fragment was created with primers RAEf and RAEr.

E(z): I subcloned the 2.5kb BglII/NotI fragment containing the entire E(z) orf into BamHI/NotI cut pBluescript. The pBluescript polylinker contains an EcoRI site upstream of the BamHI site, which was in the correct reading frame after the BamHI/BglII ligation, so E(z) was subcloned from this construct as an EcoRI/NotI fragment.

Asx: I created AsxA with the primers Asx5 and Asx1005r, AsxQ with the primers Asx1831f and Asx3414r and AsxC with the primers Asx3415f and Asx3.

Gal4 Fusion Plasmids: The first GAL4 fusion construct, pGPc was generated by subcloning the Eco RI-Bam HI Polycomb PCR product into Eco RI-Bam HI digested pM1 [174]. This generates a mammalian Gal4-Pc expression construct. Gal4-Pc was then subcloned as a Bgl II-Bam HI fragment into Bam HI digested phsNeoAct(Bam) [175]. This final construct, pGPc, has a single Eco RI site between Gal4 and Pc, and a single Bam HI site, at the 3' end of Pc. Upstream of Gal4 there is no restriction site as Bam HI has ligated into Bgl II. The constructs pGph and pGPsc were created by removing the Pc

sequence with Eco RI and Bam HI, and substituting ph or Psc DNA. The construct pG was created by digesting pGPc with Eco RI and Bam HI, blunting with Klenow and religating to remove the fragment containing Pc. The control non-fusions were created by cutting pHSNeoAct(Eco) [175] with Eco RI, and inserting the Eco RI cDNA of Pc, ph, or Psc.

CAT Reporter Plasmid: The reporter plasmid pG5hspCAT was created by inserting a Hind III-Bam HI fragment containing the hsp70 promoter and upstream P element sequence from the expression plasmid pNHT4 [176] into Xba I-Bam HI digested pG5BCAT [177]. The Hind III and Xba I overhangs were made compatible by filling in using only the nucleotides C and T for pG5BCAT (which leaves a T-C-5' overhang) and the nucleotides A and G for pNHT4 (which leaves a 3'-A-G overhang).

Mutagenesis

The mutants W1A and I62D were close enough to the 5' and 3' ends of the SAM sequence to be incorporated into their respective end primer-linkers (of the same names) and created by linker PCR. For all other mutants, a two step PCR protocol was used. Overlapping forward and reverse mutant primers were synthesized and used in separate reactions with the appropriate forward or reverse end primer linker. Vent polymerase (NEB) was used in the standard reaction through the following cycles: 5 min at 96°C; 2x(1 min at 96°C; 1 min at 4°C; 1 min at 72°C); 7x(1 min at 96°C; 1 min at 42°C; 1 min at 72°C). The products of the first step were purified by phenol/CHCl₃ extraction, 2% agarose TAE gel electrophoresis, and Qiaex gel extraction (Qiagen) and because their mutant ends overlapped, were used together as the template for the second step. The ligation of the two overlapping fragments was achieved with a single PCR using the forward and reverse end primer linkers using the following cycles: 2x(1 min at 96°C; 1 min at 4°C; 1 min at 72°C); 7x(1 min at 96°C; 1 min at 42°C; 1 min at 72°C). For the mutant construct H24L>4A, the

end primers were H2f and H2r, and the mutant primers were MLA and HLP. For all SAM domain mutants the end primers were SAMf and H2r. L33A used mutant primers L>Af and L>Ar. L41A used mutant primers L2>Af and L2>Ar. G52A used mutant primers G2>Af and G2>Ar.

Primers

Bold face denotes a restriction site

Pc5	5'-GGAGCGAATTCATGACTGGTCGAGGCAAGG-3'
Pc3	5'-GGGGGGGATCCCGACATTGTTTGGGTC-3'
Pc208f	5'-CCCATATGAATTCGACATCTACGAACAAACGAAC-3'
chr5	5'-CCCATATGAATTCGATCCAGTCGATCTAGTGTAC-3'
chr3	5'-GTGGGGATCCGATGAGGCGGCGATCCAGGAT-3'
Psc5	5'-GGAGCGAATTCATGATGACGCCAGAATCG-3'
Psc3	5'-AACGACTTGAGGAACTCCGAC-3'
Psc748f	5'-CGCATATGGAATTCAGGCCACGCCCCGTCCTTCTA-3'
Psc1149r	5'-CGCCGGATCCCTGGGGCGACTCATAAACACG-3'
Psc1005r	5'-GCGGCTCGAGTCATTCCCGTTCGTAAAGGCCCGG-3'
Psc1006f	5'-CCGCGAATTCCTGATGCGCAAAGGGCCTTC-3'
ph5	5'-GCGAATTCATGGATCGTCGTGCAT-3'
ph255r	5'-GGCCGCTCGAGCTGCTTGCCACCC-3'
phD5	5'-CCACGAATTCCTCAAGGCGATGATTAAG-3'
phD3	5'-GTGGGGATCCTCCTTAATGGACTCCACCTT-3'
H2f	5'-CCGCGAATTCATGGCTGAGGAGGAGAT-3'
H2r	5'-CCGCCTCGAGTCACTCCTTAATGGACTC-3'

Primers, cont.

MDL	5'-CCGCCTCGAGTCACGCTTGGCCGTCGATCTCCT-3'
MLA	5'-ATCGACGGCCAAGCGGCTGCGGCGGCCAAGGAGAAGCA TTTGGTG-3'
HLP	5'-CGCTTGGCCGTCGATCTCCT-3'
SAMf	5'-GGCGGAATTCAGCAGCTGGAGTGTGGAC-3'
W>A	5'-CCGCGAATTCAGCAGCGGAGTGTGGACGATGTC-3'
I4>D	5'-CGCCCTCGAGTCACTTATCGGACTCCACCTTGGC-3'
L>Af	5'-GGCCAAGCGGCTCTGTTGCTCAAGGAG-3'
L>Ar	5'-GAGCAACAGAGCCGCTTGGCCGTCGAT-3'
L2>Af	5'-CTCAAGGAGAAGCATGCGGTGAACGCTATGGGC-3'
L2>Ar	5'-GCCCATAGCGTTCACCGCATGCTTCTCCTTGAG-3'
G2>Af	5'-GGCATGAAGCTGGCTCCAGCTCTTAAAATT-3'
G2>Ar	5'-AATTTTAAGAGCTGGAGCCAGCTTCATGCC-3'
RAEf	5'-CCGCGAATTCCTAGCCAGTGGAGC-3'
RAEr	5'-CGCCCTCGAGGGTCACTTAGGT-3'
BEBf	5'-CCGCGAATTCGCAGAGTTTTGGTCACCCGAA-3'
BEBr	5'-GCGGCTCGAGTCACTCTTTGATTTTTTCTATTTC-3'
Asx5	5'-CCGGGAATTCATGAAAACCATTACGCCG-3'
Asx1005r	5'-CCGGCTCGAGCTCGCCCCAGAAGGGCTC-3'
Asx1831f	5'-CCGGGAATTCATGATTTTCGTTTTCTCAG-3'
Asx3414r	5'-GGCCGGATCCTGAGATGATATTTAGTGA-3'
Asx3415r	5'-CCGGGAATTCATGACGCGTCCTGCCAAT-3'
Asx3	5'-CCGGGGATCCGTTATCCACCTCATCTA-3'

Primers, cont.

JGf 5'-CTGAGTGGAGATGCCTCC-3'

Sequencing

All sequencing was done on an automated sequencer (NAPS unit) using fluorescent dye termination. Sequence from PCR products was obtained after subcloning into pBluescript, and priming with T7 and T3 standard primers.

Sequencing of hAsx

The insert from cDNA 42515 of the Image Consortium (Genbank Accession T16795) was subcloned into pBluescript in three pieces: a 750 bp Hind/Hind fragment, a 550 bp Xho/Xho fragment, and a 420 Hind/Hind fragment. These were sequenced with T7 and T3 primers. The sequence of the three fragments as well as the end sequence from the ESTs were melded into a single 1505 bp contig.

Preliminary sequencing of the Asx-interacting two-hybrid clones

Preliminary forward sequence through the EcoRI restriction site was obtained using the primer JGf.

Complete sequencing of the z40 interacting clone

The entire insert was subcloned into pBluescript as an EcoRI/XhoI fragment and end sequence was obtained with T7 and T3 primers. This was augmented by sequencing a BamHI deletion derivative (BamHI site at 365 bp 3' of the EcoRI site). These sequences were melded into a single 1108 bp contig.

Complete sequencing of the z34 interacting clone

The entire insert was subcloned into pBluescript as an EcoRI/XhoI fragment and end sequence was obtained with T7 and T3 primers. Several 5' deletion derivatives were generated using the enzymes BamHI, BstXI, PstI, and XbaI, which cut at positions 919, 1184, 1490, and 1714 respectively. This sequence was augmented with sequence from a *Drosophila* EST, HL02032 5' (Genbank accession: AA567945) and melded into a single 2286 bp contig.

GST-fusion protein expression and purification

pGEX-4T-1 derivative plasmids were transformed into the bacterial strain AD202. Single colonies were grown to an OD₆₀₀ of 0.6 in 250mL LB at 37°C and induced with the addition of 250μL of 1M IPTG. Induction was carried out for 15 hours at 25°C. Cells were collected by centrifugation, resuspended in 15mL of 20mM Tris•Cl/100mM NaCl/1mg/mL lysozyme, and left at room temperature for 1 hour. 5μL of β-mercapto ethanol was added, and the resuspended cells were subjected to 6 cycles of freeze/thaw with N₂(l). The extract was cleared by centrifugation for 40 minutes at 12500 rpm (SS34) at 4°C, and filtered through miracloth.

***In vitro* co-affinity precipitations**

³⁵S-methionine labeled proteins were generated using the Promega TNT rabbit reticulocyte lysate transcription/translation kit according to the manufacturer's instructions. Templates were uncut plasmid DNA. cDNAs with appropriate initiator methionine codons were transcribed by T7 or T3 polymerase from pBluescript constructs, and inserts lacking an initiator methionine were transcribed by T7 from pET28a (Novagen) constructs which provided the initiator methionine. GST-fusion protein bound to glutathione agarose beads were prepared by incubating an aliquot of raw bacterial extract with 50μL of a 50% slurry of reduced glutathione agarose (Sigma) in 100mM NaCl/20mM Tris•Cl pH 7.5 (TBS), in

1mL of TBS/1%NP40/0.5%PMSF saturated isopropanol (PMSF) for 30 minutes with gentle rocking at 4°C. The amount of bacterial extract was normalized to give 1µg of fusion protein in each experimental tube. The bound beads were washed twice in TBS/1%NP40, and once in TBS. They were then blocked in a solution of 5% skim milk in TBS for 30 minutes at 4°C. The ³⁵S-labeled proteins from the *in vitro* translation reactions were precleared with the addition of GST-bound glutathione agarose in TBS, followed by incubation at 4°C with gentle rocking for 30 minutes. For each 200µL *in vitro* translation reaction, 100µL bed volume of glutathione agarose coupled to 10µg of GST in a volume of 500µL was used in the preclearing step. 70µL of precleared lysate and 5µL of 10% NP40 (to 0.1% final) was added to the blocking mixture in each experimental tube and these were incubated for 30 minutes at 4°C. The bound beads were washed twice in TBS/0.5% NP40, twice in 500mM NaCl/20mM Tris•Cl pH 7.5, and once in TBS, followed by elution in 30µL TBS/20mM reduced glutathione pH7.5. The eluate was analyzed by Tricine SDS PAGE [178] on a 10%/16% discontinuous gel for labeled minimal SAM domains and SDS PAGE on a 12% gel for the larger pHHD construct. 1/3 of the eluate was loaded in each experimental lane, and 2.5 µL of the pre-bound lysate was loaded in the control lane.

Binding to Ni-NTA agarose

100µL of *in vitro* translated ³⁵S-labeled sSAM was mixed with 50µL of Ni-NTA agarose (Qiagen) in 1mL TBS and rocked for 30 minutes at 4°C. The beads were then washed 3x in TBS, and eluted with 100mM imidazole/TBS pH 7.5. Equal volumes of pre-bound translation reaction and bound/eluted polypeptide were loaded and run on 10%/16% discontinuous Tricine SDS PAGE and visualized by autoradiography.

Transformation and Culturing of Yeast Strains

Yeast were grown nonselectively on YPD or selected on CM dropout medium lacking uracil, tryptophan, histidine, or leucine. For transformations, 50mL of fresh yeast culture at an OD₆₀₀ of 1.0 were collected by centrifugation for 5 minutes at 2000 rpm at room temperature on a tabletop centrifuge (Clay-Adams) and resuspended in 40mL of dH₂O. Cells were pelleted again and resuspended in 1.5mL of freshly prepared 100mM LiOAc/TE. 200μL of cells were added to glass tubes containing 1μg of the plasmid to be transformed plus 200μg of denatured herring sperm DNA as carrier. 1.2 mL of PEG solution (8 parts sterile 50%PEG/1 part 1M LiOAc/1 part 10x TE) was added and the tubes were set turning at 30°C for 30 min. A 15 minute heat shock at 42°C was applied and yeast were plated directly onto selective plates.

Yeast Strains

EGY48	MAT α ura3-52 his3 trp1 leu2::lexA-LEU2
JRY4012	MAT α can1-100 his3-11 leu2-3,112 lys2 Δ trp1-1 ura3-1 gal+
JRY4622	isogenic to JRY4012 sir1 Δ ::LEU2
JRY4588	isogenic to JRY4012 sir2 Δ ::LEU2
JRY4606	isogenic to JRY4012 sir3 Δ ::LEU2
JRY4581	isogenic to JRY4012 sir4 Δ ::LEU2
IH2534	MAT α ura3-52 leu2 his3 trp1 ade2 lys2 gal-
IH2536	isogenic to IH2534 sin3 Δ ::TRP1
IH2542	isogenic to IH2534 sin1 Δ ::TRP1
RMY202I	MAT α ade2-101 his3 Δ 200 lys2-801 trp1 Δ 901 ura3-52 hht1,hhf1::LEU2 HHT2,HHFA1-17
DY1571	MAT α ade2 can1 his3 leu2 trp1 ura3 lexA:UAS ^{cyc1} :lacZ:URA3
DY1609	MAT α ade2 can1 his3 leu2 trp1 ura3 UAS ^{cyc1} :lexA:lacZ:URA3

Yeast Strains, cont.

FY250	MAT α his3 Δ 200 leu2 Δ 1 ura3-52 trp1 Δ 63
FY604	MAT α his3 Δ 200 leu2 Δ 1 ura3-52 trp1 Δ 63 (hta2-htab2) Δ TRP1
YPH680	MAT α ura3-52 his3 Δ 200 leu2 Δ 1 trp1 Δ 63 lys2 Δ 202
YCB428	isogenic to YPH680 sir2 Δ 2
YCB532	isogenic to YPH680 hst1 Δ 3::TRP1
YCB494	isogenic to YPH680 hst2 Δ 2::TRP1
YCB424	isogenic to YPH680 hst3 Δ 3::TRP1
YCB644	isogenic to YPH680 hst4 Δ 1::TRP1
YCB483	MAT unknown ura3-52 his3 Δ 200 leu2 Δ 1 trp1 Δ 63 lys2 Δ 202 ade2 Δ sir2 Δ 2::TRP1 hst1 Δ 3::TRP1 hst2 Δ 2::TRP1 hst3 hst3 Δ 3::TRP1

These strains were the gifts of the following people: EGY48 Erica Golemis[110]; FY250, FY604 Fred Winston [179]; JRY4012, JRY4581, JRY4588, JRY4606, JRY4622 Jasper Rine [180]; IH2534, IH2536, IH2542 Ira Herskowitz; YPH680, YCB428, YCB532, YCB494, YCB424, YCB644, YCB483 Carrie Baker Brachmann [164]; RMY202I, Randall K. Mann [181]; DY 1571, DY1609 David Stillman

Two-Hybrid Interaction Assays

The strain EGY48 was transformed with derivatives of plasmids EG202 and JG4-5 [110] encoding the LexA- and activator-fusion proteins respectively. Three individual transformed colonies from each plate were streaked out on both dextrose and galactose plates containing complete minimal medium lacking uracil, histidine, and leucine. Growth was scored after 4 days: a strong interaction was deemed to have occurred if the colonies reached 1mm in diameter. Plates with slower-growing colonies were scored as weak interactions, and the absence of growth indicated no interaction. For the quantitative assays

used in Chapter 3, an overnight culture was grown in galactose medium supplemented with leucine but lacking histidine, tryptophan, and uracil (3⁻). Cells (0.5 ml) were pelleted and resuspended in H₂O (150 µL). A limiting dilution series (10x steps, 10 ml per sample) was spotted onto leucine supplemented 3⁻galactose, leucine deficient 3⁻galactose, and leucine deficient 3⁻dextrose plates. Colony number was counted in the spot with the most colonies that could reliably be scored, and extrapolated to the first sample. Frequency of prototrophs was taken to be (cfu from leucine deficient 3⁻ galactose) divided by (cfu from leucine supplemented 3⁻ galactose plates).

β-galactosidase assays

Yeast were grown in triplicate overnight in 2mL of selective medium, pelleted, and resuspended in 1mL Z-buffer (60mM Na₂HPO₄, 40mM NaH₂PO₄, 10mM KCl, 1mM MgSO₄, 50mM β-mercaptoethanol, pH 7). 100µL was diluted in 900µL dH₂O to a disposable cuvette for absorbance readings at 600nm. To the remaining 900µL, one drop (25µL) of 0.1%SDS and two drops (50µL) of CHCl₃ were added. Samples were vortexed 15 sec and left 15 min at 30°C to equilibrate. In 10 second intervals, 200µL of 4mg/mL ONPG was added to each tube, and mix by vortexing. Reactions were stopped in the same order in the same 10 sec intervals by adding 500µL of 1MNa₂CO₃. The samples were pelleted for 5 minutes at max. on a benchtop centrifuge, and the supernatant saved for absorbance readings at 420nm. Units of β-galactose were equal to:

$1000(OD_{420})/(t)(v)(OD_{600})$, where v is the volume (in mL) of yeast used (.9 in this protocol) and t is the time of the reaction in minutes. This protocol is modified from Breeden and Naysmith [182].

Co-immunoprecipitation from Kc nuclear extracts

Nuclear extracts were prepared from 2L of Kc cells at a cell density of 2x10⁶ cells/mL according to Heberlein *et al.* [183] and Parker & Topol [184] Antibody to Pc was kindly

provided by Dr. Jacob Hodgson. 2 μ L of pre-immune serum was added to 200 μ L of nuclear extract and incubated at 4°C with gentle rocking for 30 minutes. 80 μ L of a 50% slurry of proteinA sepharose in HEMG (25mM HEPES-K⁺ pH 7.6, 100mM KCl, 12.5mM MgCl₂, 0.1mM EDTA, 0.1mM EGTA, 15% glycerol, 1.5mM DTT) was added and the tube was rocked for a further 60 minutes. The beads were removed by centrifugation, and the cleared extract divided evenly between two tubes containing equal amounts of IgG, either 0.5 μ L of pre immune serum or 1 μ L of affinity-purified anti-Pc antibody. The antibody was bound for 60 minutes, 20 μ L of 50% proteinA beads were added and bound for 30 minutes. The bound beads were then washed 3x in HEMG and eluted with SDS PAGE loading buffer, run on an 8% gel, transferred to nitrocellulose and blocked in 3% BSA. The filter was then cut into high and low molecular weight pieces, and the bottom was probed with the same anti-Pc antibody, while the top was probed with anti-ph [172] and anti-Psc [33].

SL2 Transfections

SL2 cells were seeded into 6x3 mL well plates and grown to 50% confluency in Schneider's Medium + 10% FCS. Transfections were carried out with Lipofectin from Gibco (Cat. No. 18292-011) according to the manufacturer's directions. Each experiment was performed in triplicate. Each replicate contained 0.11 μ g of G5HSPCAT reporter, 1 μ g of producer plasmid.

CAT assay

24 hours after transfection, cells were exposed to a 1hr heat shock (37°C) and left to recover overnight at room temperature. The next day, cells were harvested, and CAT activity monitored using the Boehringer Mannheim CAT-ELISA kit (Cat. No. 1363 727) according to the manufacturer's directions.

Telomeric Variagation Assays

Triplicate 2mL cultures were grown overnight. A limiting dilution series (10x steps, 10 ml per sample, the same as for the two-hybrid quantitation, above) was spotted onto two sets of plates: one with 5-fluoroorotic acid (FOA) (1mg/mL), the other without. Colony number was counted in the spot with the most colonies that could reliably be scored, and extrapolated to the first sample. Frequency of FOA resistance was taken to be (cfu from the FOA plate) divided by (cfu from the control plates). To monitor ADE2 expression, plates with visible colonies were transferred to room temperature, and allowed to continue growing. Red colour appeared over a week.

Conclusion

The work described in this thesis has identified a network of protein-protein interactions within the PcG/trxG. Interprotein interactions, defined by domains that show interactions in one or more assays, can be described by a linear network: Scm-ph-Psc-Pc-z40-Asx-trx. E(z)-esc, and Pcl do not as yet fit into this network. In addition to the heterotypic interactions, I have demonstrated homotypic interactions for Scm, ph, Psc, and Asx. As more proteins are tested for interactions with the panel I have generated, the network may grow to include E(z)-esc and Pcl, and will probably change from a chain to a web of interactions.

The Assembly of PcG Complexes

The network is not to be taken to describe the members of a particular complex and their interactions, rather it is a network of interaction possibilities. The constitution of a particular complex will likely be governed by the availability of potential members, through the specificity of the various *cis* elements at the locus of complex assembly. The rather promiscuous interaction behaviour of the average PcG protein is consistent with the idea of higher order complexes brought about through multiple cooperative interactions. The constitution of a given complex will likely further be subject to the effects of covalent modifications of monomers, allosteric interactions, steric hindrance, and cooperative assembly.

The following is a model for assembly of multimeric PcG complexes (Figure 6.1) based on some of the observations made in this thesis: PcG proteins are not synthesized in an assembly-ready state, and must be modified in some way to enable assembly. This is suggested by the absence of, or weakening of association when many of the interacting

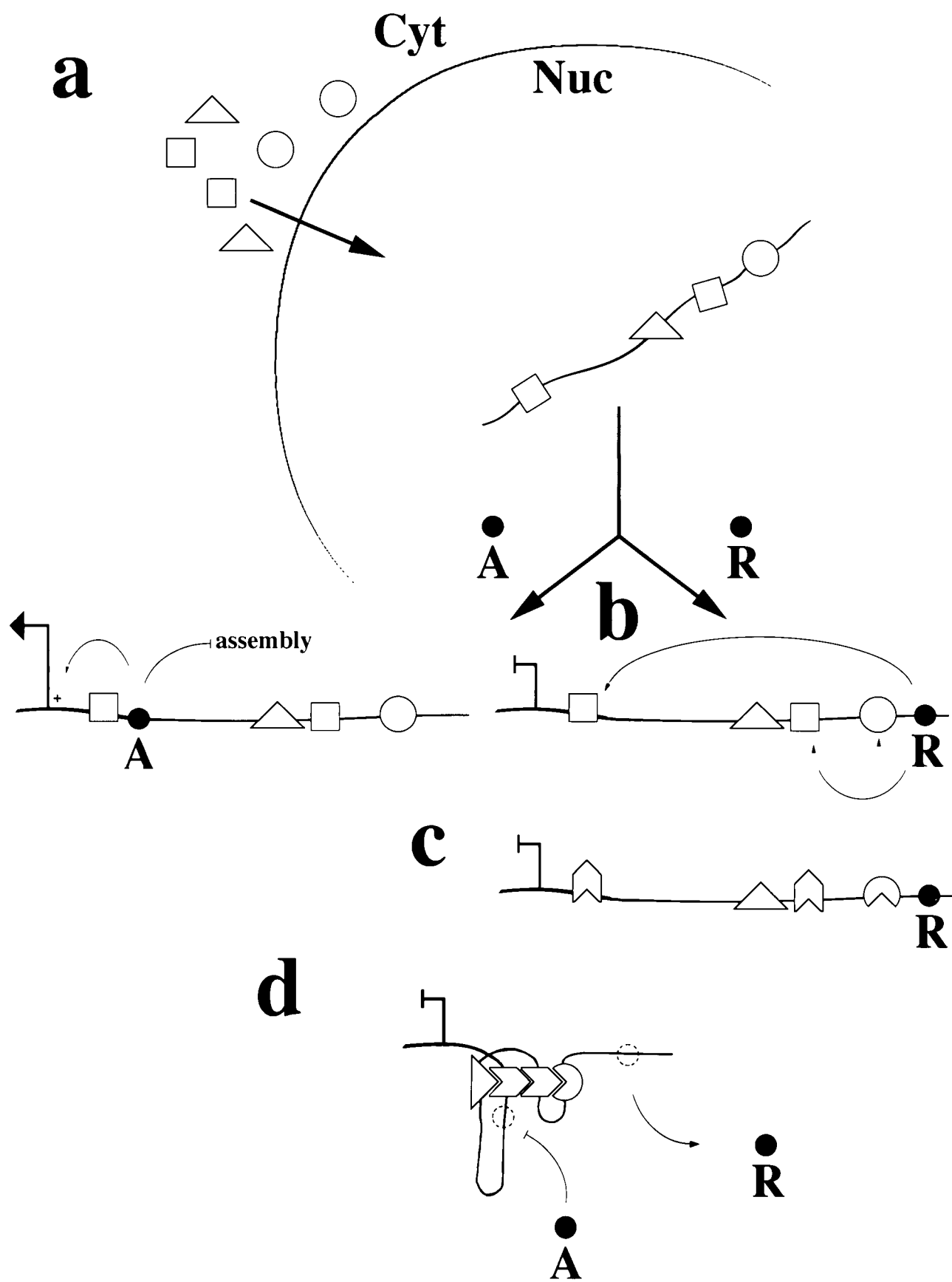


Figure 6.1

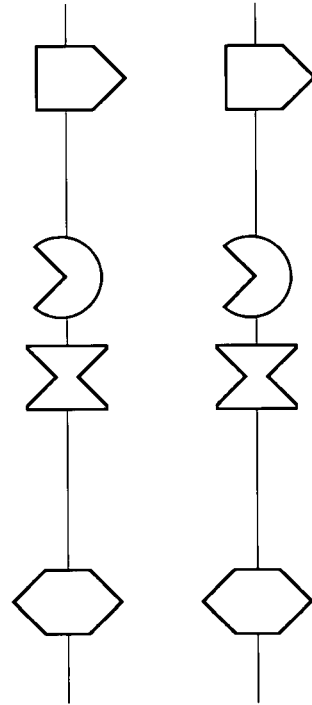
A model for regulated PcG complex assembly. (a) Newly synthesized proteins are not assembly-competent. Cyt (cytoplasm); Nuc (nucleus). After nuclear import, proteins bind to their chromosomal target sites irrespective of transcriptional activity at the site. (b) Two possibilities exist: (left) the locus is transcriptionally active, in which case a transcriptional activator "A" directly or indirectly prevents assembly of the PcG complex. (right) The locus is transcriptionally silent, in which case a repressor "R" induces some modification (c) which renders the PcG proteins competent for complex assembly. (d) Once a complex has assembled, the inducing repressor is no longer needed. In addition, activators are now inhibited, either by occlusion of their *cis* elements as shown, or by being prevented from interacting with the basal transcriptional complex while bound to their *cis* elements.

domains that I have identified are expressed in the context of their full length proteins as opposed to in the absence of external sequences. Such suppression may be necessary to avoid the formation of complexes in the cytoplasm which would inhibit the nuclear import of the subunits. The modifications that allow assembly could happen before, after, or during (as a consequence of) recognition of DNA. Accepting for the moment Pirrotta's notion of a complex at the PRE inducing the recruitment of smaller, distant *cis* elements, complex assembly would happen after DNA recognition. The local presence of factors that are correlated with transcriptional activation (transcriptional activators, the holoenzyme, basal transcription factors, trxG proteins, or RNA) then block activation of assembly, while factors that are associated with repression (transient repressors such as hunchback protein) induce the activation of assembly. The activation event itself could be a covalent modification such as phosphorylation or dephosphorylation (given the presence of multiple serine/threonine rich domains in several PcG proteins) which would allow the factors present at the various *cis* elements to engage one another, catalysing the further inclusion of distant *cis* elements. This model allows PcG proteins to be present at a locus without necessarily repressing transcription there, as is seen for several loci in larval salivary gland nuclei. It also resolves the apparent inconsistency of the ubiquitous presence of PcG proteins with the limited domains of expression and repression of their target genes. Finally it makes two predictions. One, that some PcG proteins should be present in multiple isotypic forms, and two, that there should exist modifiers of PcG proteins, which if they are specific to PcG proteins, should themselves have PcG mutant phenotypes. The corollary of this latter prediction is that some PcG genes will encode enzymatic modifiers of PcG proteins. This has been suggested for E(Pc) [151].

Dimerization

Crosstalk (transvection) between homologous chromosomes requires that a protein located at a particular site on one chromosome can bind to a protein located at the homologous site

a



b

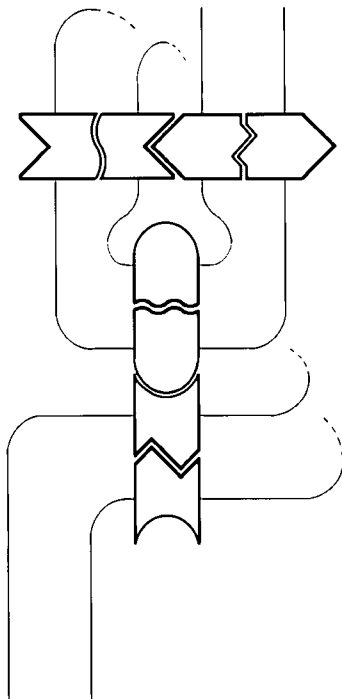
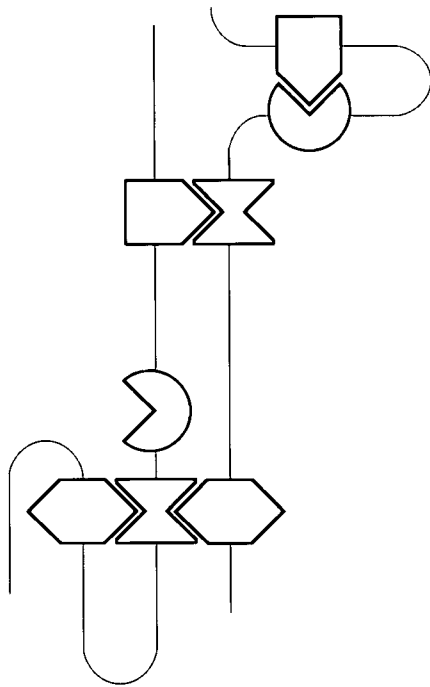
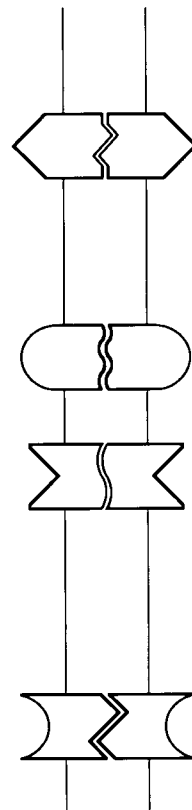


Figure 6.2

Self-association of PcG proteins may facilitate inter-homologue interactions.

(a) A locus bound by several proteins capable of heteromultimerizing, but incapable of homodimerizing. Protein interactions occur and are inimical to homologue synapsis.

(b) A locus bound by proteins capable both of homodimerizing and of heteromultimerizing. Dimerization leads to homologue synapsis, which could enable transvection at the locus.

Higher order complexes are also possible, with a lower degree of freedom than in case (a) due to obligatory synapsis.

on the homologous chromosome. If, as sequence homology makes likely, the same protein is bound at the same site on both homologues, then the ability of such a protein to dimerize will facilitate crosstalk. *zeste*, which mediates transvection, can bind DNA [185], and can self-aggregate [186]. I have shown that *Asx*, *Psc*, *Scm*, and *ph* have dimerization (or multimerization) modules. This means that the DNA target sites of these three proteins are potential homologue synapsis points. It invites the prediction that mutations that affect the dimerization of these proteins will also affect transvection at the BxC. Certain *Psc* and *Scm* mutant alleles are already known to affect transvection at the *white* gene. Perhaps these genes affect *zeste-white* transvection because of a nearby PRE to which their products are bound, while the lack of effect on transvection at *white* by other PcG mutants means that this particular PRE is not a target site for these products.

PcG mutations cause a phenotype similar to gain of function mutations of the homeotics, so their requirement for transvection should be fairly easy to detect. Since the absence of transvection leads to loss of function at a locus showing pairing-dependent complementation, PcG mutations that suppress transvection should suppress this complementation, causing reduced expression, rather than their usual effect of causing increased expression of a given homeotic gene. An alternative test would be to examine transvection-suppressing rearrangements of the BxC for phenotypic enhancement of PcG heterozygotes. Finding an effect on transvection at the BxC would confirm the role of interhomologue interactions in PcG function at the BxC. It would mean that the *Su(z)* and *E(z)* phenotypes of PcG mutants and derepression of the homeotics are different manifestations of the same phenomenon.

Silencing or Repression?

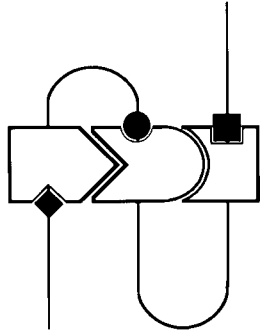
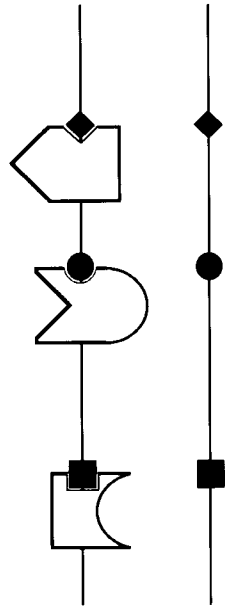
In chapter 4 it was suggested that some PcG/trxG factors might be involved in setting up a silenced chromatin conformation while others might be involved in conventional

transcriptional repression. In accordance with the model described earlier, two phases of PcG action could exist: the first competing directly with the transcriptional apparatus, the second locking in the silenced state. Silencing and repression could therefore both be occurring together at the homeotic loci. Given the large number of PcG proteins, it would not be surprising at all if the PcG restricted expression of target genes via multiple mechanisms. One should not rule out the possibility that different components of a multimeric complex, or even different domains within a given protein, are themselves functioning in different pathways to repression or silencing.

The Large Membership of the PcG

Notwithstanding the possibility of multiple mechanisms of repression or silencing, why are so many different PcG proteins needed? Jurgens [16] and Landecker *et al.* [187] estimate that there are around 40 *Pc*-enhancing loci. Assuming that the untested PcG proteins behave similarly to the ones that I have tested, what can the sheer multitude of members of this group tell us about how the group functions? Multiple factors must be interacting to i) initiate silencing, ii) maintain silencing, or iii) reverse silencing. Beginning with the least likely of the three, if it is true that maintenance of silencing requires a large number of factors, the mechanism must involve more than simple occlusion of enhancer or promoter binding sites. Even so, if silencing can be achieved at some loci without the presence of certain PcG proteins, why is it that these are needed at other loci? Using multiple factors where only a few are necessary seems unparsimonious. It is most likely that the multiplicity of PcG members is related instead to initiation (and reversal, at loci where it occurs) of silencing. If one assumes that PcG silencing activity at a locus is dependent on the presence of other transcription factors, and that direct interactions with these factors either initiate silencing or prevent it, then the sheer multitude of transcription factors capable of acting on PcG-regulated loci would invoke a large number of PcG proteins. Every locus has its own particular constellation of transcription factors, therefore every locus capable of

a



b

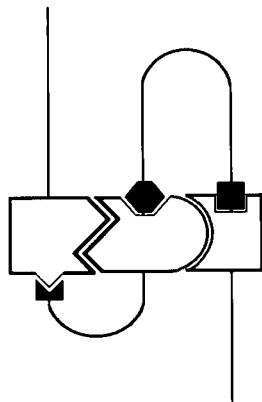
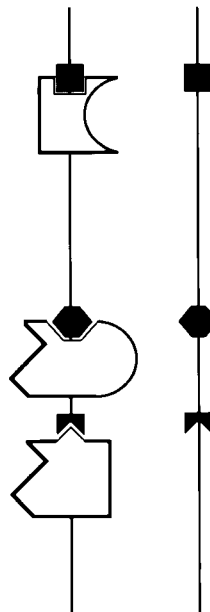


Figure 6.3

Different loci require different PcG proteins for silencing. Transcription factors are shown in black. A potential functional interaction between a given transcription factor and a given PcG protein is designated by an indentation in the PcG protein of the same shape as the transcription factor. Loci (a) and (b) share one transcription factor in common but each have two which are unique. In this case, each transcription factor is specific to a different PcG protein, hence each share one PcG protein in common, but have two which are unique. The interacting transcription factor and PcG protein are shown adjacent to one another, suggestive of direct binding, but need not be, especially if the functional interaction were something other than physical binding. The relationship between which transcription factors bind to a locus and which PcG proteins do is governed by the functional consequences of potential interactions. Different loci have different patterns of upstream transcription factor binding, hence different loci have different requirements for PcG proteins, although in the final analysis the activity accomplished by the PcG proteins may be the same from locus to locus (shown here as aggregation occurring at both loci).

being silenced by the PcG would have its own corresponding constellation of PcG proteins (Figure 6.3). One of the predictions of this explanation is that PcG proteins, in addition to interacting with each other, should also interact with other transcription factors. This seems to be the case for Asx, given its interactions with the zinc-finger transcription factor, Bowel and with *trx*.

It is clear that identifying interactions within the PcG is just the beginning of understanding how PcG-mediated repression works. It will be necessary to continue the search for protein interactions beyond the limits of the PcG, perhaps through such means as were used in chapter 3 for Asx, until a connection (direct or indirect) is made with the transcriptional apparatus itself.

Nomenclature

ALL-1	Acute Lymphoblastic Leukemia locus 1
AntC	Antennapedia Complex
AP	anterior-posterior
Asx	Additional sex combs
BxC	Bithorax Complex
BEB1	BEM1-binding protein 1
bmi-1	B-cell-specific Molony murine leukemia virus insertion site 1
brm	brahma
co-IP	co-immunoprecipitate
CAT	chloramphenicol acetyl transferase
CDK	cyclin dependent kinase
cfu	colony-forming units
crm	cramped
CTD	carboxyl terminal repeat domain of pol II
E(Pc)	Enhancer of Polycomb
E(var)	Enhancer of PEV
E(z)	Enhancer of zeste
en	engrailed
esc	extra sex combs
EST	expressed sequence tag
FOA	5-fluoroorotic acid
GST	glutathione-S-transferase
HEPES	N-[2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)]
hE(z)	human E(z)
hph	human ph

Nomenclature, cont.

HST	homologous to SIR2
HTH	helix-turn-helix
IP	immunoprecipitate
mel-18	melanoma-specific cDNA 18
mph	mouse ph
mxo	multi sex combs
Ni-NTA	nickel-nitrilotriacetic acid
NURF	nucleosome remodeling factor
Pc	Polycomb
PcG	Polycomb Group
Pcl	Polycomblake
PEG	polyethylene glycol
PEV	Position Effect Variegation
ph	polyhomeotic
pho	pleiohomeotic
PMSF	phenylmethanesulfonyl fluoride
pol II	RNA polymerase II
PP2A	protein phosphatase 2A
PRE	PcG Response Element
Psc	Posterior sex combs
RAE28	retinoic acid elevated cDNA 28
RRM	RNA Recognition Motif
SAM	Self-Association Motif (formerly Sterile Alpha Motif)
Sce	Sex combs extra
Scm	Sex combs on midleg

Nomenclature, cont.

Scr	Sex combs reduced
SET	Su(var)3-9-E(z)-Trx
SIN	SWI-independent
SIR	Silent Information Regulator
SL2	Schneider Line 2
SPM	Scm-ph-mbt
SRB	Suppressor of RNA polymerase B
Su(var)	Suppressor of PEV
Su(z)2	Suppressor two of zeste
sxc	super sex combs
TBS	100mM NaCl/20mM Tris•Cl ph 7.5
Trl	Trithoraxlike
trx	trithorax
trxG	trithorax Group
Ubx	Ultrabithorax

Bibliography

1. **Lewis, E.B.**, A gene complex controlling segmentation in *Drosophila*, *Nature*, 276, 565-570, 1978.
2. **Kaufman, T.C., Lewis, R.A. and Wakimoto, B.T.**, Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homeotic gene complex in polytene chromosome interval 84A-B., *Genetics*, 94, 115-133, 1980.
3. **Harding, K. and Levine, M.**, Gap genes define the limits of *Antennapedia* and *bithorax* gene expression during early development in *Drosophila*., *EMBO J.*, 7, 205-214, 1988.
4. **Zhang, C.-C. and Bienz, M.**, Segmental determination in *Drosophila* conferred by *hunchback*, a repressor of the homeotic gene *Ultrabithorax*, *Proc. Natl. Acad. Sci. USA*, 89, 7511-7515, 1992.
5. **Struhl, G. and Brower, D.**, Early role of the *esc*⁺ gene product in the determination of segments in *Drosophila*., *Cell*, 31, 285-292, 1982.
6. **McKeon, J. and Brock, H.W.**, Interactions of the *Polycomb* group of genes with homeotic loci of *Drosophila*, *Roux's Arch. Dev. Biol.*, 199, 387-396, 1991.
7. **Simon, J., Chiang, A. and Bender, W.**, Ten different *Polycomb* group genes are required for spatial control of the *abd-A* and *Abd-B* homeotic products, *Development*, 114, 493-505, 1992.
8. **Soto, M.C., Chou, T.-B. and Bender, W.**, Comparison of germ-line mosaics of genes in the Polycomb group of *Drosophila melanogaster*., *Genetics*, 140, 231-243, 1995.
9. **Struhl, G. and Akam, M.E.**, Altered distribution of *Ultrabithorax* transcripts in *extra sex combs* mutant embryos of *Drosophila*, *EMBO J.*, 4, 3259-3264, 1985.
10. **Jones, R.S. and Gelbart, W.M.**, Genetic analysis of the *Enhancer of zeste* locus and its role in gene regulation in *Drosophila melanogaster*, *Genetics*, 126, 185-199, 1990.
11. **Struhl, G.**, A gene product required for correct initiation of segmental determination in *Drosophila*, *Nature*, 293, 36-41, 1981.
12. **Duncan, I.M.**, *Polycomblike*: a gene that appears to be required for the normal expression of the bithorax and Antennapedia complexes of *Drosophila melanogaster*, *Genetics*, 102, 49-70, 1982.
13. **Sato, T., Russell, M.A. and Denell, R.E.**, Homoeosis in *Drosophila*: a new enhancer of Polycomb and related homoeotic mutations, *Genetics*, 105, 357-370, 1983.
14. **Ingham, P.W.**, A gene that regulates the bithorax complex differently in larval and adult cells of *Drosophila*, *Cell*, 37, 815-823, 1984.

15. **Dura, J.-M., Brock, H.W. and Santamaria, P.,** *Polyhomeotic*: a gene in *Drosophila melanogaster* required for correct expression of segment identity, *Mol. Gen. Genet.*, 198, 213-220, 1985.
16. **Jurgens, G.,** A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*, *Nature*, 316, 153-155, 1985.
17. **Breen, T.R. and Duncan, I.M.,** Maternal expression of genes that regulate the bithorax complex of *Drosophila melanogaster*, *Dev. Biol.*, 118, 442-456, 1986.
18. **Adler, P.N., Martin, E.C., Charlton, J. and Jones, K.,** Phenotypic consequences and genetic interactions of a null mutation in the *Drosophila Posterior Sex Combs* gene, *Dev. Genet.*, 12, 349-361, 1991.
19. **Sinclair, D.A.R., Campbell, R.B., Nicholls, F., Slade, E. and Brock, H.W.,** Genetic analysis of the *Additional sex combs* locus of *Drosophila melanogaster*, *Genetics*, 130, 817-825, 1992.
20. **Adler, P.N., Charlton, J. and Brunk, B.,** Genetic interactions of the *Suppressor 2 of zeste* region genes, *Dev. Genet.*, 10, 249-260, 1989.
21. **Girton, J.R. and Jeon, S.H.,** Novel embryonic and adult homeotic phenotypes are produced by *pleiohomeotic* mutations in *Drosophila*, *Dev. Biol.*, 161, 393-407, 1994.
22. **Docquier, F., Saget, O., Forquignon, F., Randsholt, N.B. and Santamaria, P.,** The *multi sex combs* gene of *Drosophila melanogaster* is required for proliferation of the germline, *Roux's Arch. Dev. Biol.*, 205, 203-214, 1996.
23. **Yamamoto, Y., Girard, F., Bello, B., Affolter, M. and Gehring, W.,** The cramped gene of *Drosophila* is a member of the Polycomb-group, and interacts with mus209, the gene encoding Proliferating Cell Nuclear Antigen., *Development*, 124(17), 3385-3394, 1997.
24. **Cheng, N.N., Sinclair, D.A.R., Campbell, R.B. and Brock, H.W.,** Interactions of *polyhomeotic* with *Polycomb* group genes of *Drosophila melanogaster*., *Genetics*, 138, 1151-1162, 1994.
25. **Campbell, R.B., Sinclair, D.A.R., Couling, M. and Brock, H.W.,** Genetic interactions and dosage effects of *Polycomb* group genes of *Drosophila*., *Mol. Gen. Genet.*, 246, 291-300, 1995.
26. **Kennison, J.A. and Russell, M.A.,** Dosage-dependent modifiers of homeotic mutations in *Drosophila melanogaster*., *Genetics*, 116, 75-86, 1987.
27. **Phillips, M.D. and Shearn, A.,** Mutations in *polycombeotic*, a *Drosophila Polycomb* group gene, cause a wide range of maternal and zygotic phenotypes, *Genetics*, 125, 91-101, 1990.
28. **Locke, J., Kotarski, M.A. and Tartof, K.D.,** Dosage-dependent modifiers of position-effect variegation in *Drosophila* and a mass action model that explains their effect, *Genetics*, 120, 181-198, 1988.

29. **Franke, A., DeCamillis, M.A., Zink, D., Cheng, N., H.W. Brock and Paro, R.,** *Polycomb* and *polyhomeotic* are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*, *EMBO J.*, 11, 2941-2950, 1992.
30. **Lonie, A., D'Andrea, R., Paro, R. and Saint, R.,** Molecular characterisation of the *Polycomblike* gene of *Drosophila melanogaster*, a trans-acting negative regulator of homeotic gene expression., *Development*, 120, 2629-2636, 1994.
31. **Carrington, E.A. and Jones, R.S.,** The *Drosophila Enhancer of zeste* gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution, *Development*, 122(12), 4073-4083, 1996.
32. **Rastelli, L., Chan, C.S. and Pirrotta, V.,** Related chromosome binding sites for *zeste*, suppressor of *zeste* and *Polycomb* group protein and their dependence on *Enhancer of zeste* function, *EMBO J.*, 12, 1513-1522, 1993.
33. **Martin, E.C. and Adler, P.N.,** The *Polycomb* group gene *Posterior sex combs* encodes a chromosomal protein, *Development*, 117, 641-655, 1993.
34. **van der Lugt, N., Alkema, M., Berns, A. and Deschamps, J.,** The *Polycomb*-group homologue Bmi-1 is a regulator of murine Hox expression, *Mech. Dev.*, 58, 153-164, 1996.
35. **Akasaka, T., Kanno, M., Balling, R., Mieza, M.A., Taniguchi, M. and Koseki, H.,** A role for *mel-18*, a Polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton, *Development*, 122, 1513-1522, 1996.
36. **Takahara, Y., Tomotsune, D., Shirai, M., Katoh-Fukui, Y., Nishii, K., Motaleb, M., Nomura, M., Tsuchiya, R., Fujita, Y., Shibata, Y., Higashinakagawa, T. and Shimada, K.,** Targeted disruption of the mouse homologue of the *Drosophila polyhomeotic* gene leads to altered anteroposterior patterning and neural crest defects, *Development*, 124(9), 3673-3682, 1997.
37. **Coré, N., Bel, S., Gaunt, S.J., Aurrand-Lions, M., Pearce, J., Fisher, A. and Djabali, M.,** Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice., *Development*, 124, 721-729, 1997.
38. **Alkema, M.J., Lugt, N.M.T.v.d., Bobeldijk, R.C., Berns, A. and Lohuizen, M.v.,** Transformation of axial skeleton due to overexpression of *bmi-1* in transgenic mice, *Nature*, 374, 724-727, 1995.
39. **Muller, J., Gaunt, S. and Lawrence, P.A.,** Function of the Polycomb protein is conserved in mice and flies., *Development*, 121, 2847-2852, 1995.
40. **Busturia, A. and Bienz, M.,** Silencers in *Abdominal-B*, a homeotic *Drosophila* gene, *EMBO J.*, 12, 1415-1425, 1993.
41. **Chan, C.-S., Rastelli, L. and Pirrotta, V.,** A *Polycomb* response element in the *Ubx* gene that determines an epigenetically inherited state of repression., *EMBO J.*, 13, 2553-2564, 1994.

42. **Zink, B., Engstrom, Y., Gehring, W. and Paro, R.,** Direct interaction of the *Polycomb* protein with *Antennapedia* regulatory sequences in polytene chromosomes of *Drosophila melanogaster*., *EMBO J.*, 10, 153-162, 1991.
43. **Gindhart, J.G. and Kaufman, T.C.,** Identification of *Polycomb* and *trithorax* group responsive elements in the regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*., *Genetics*, 139, 797-814, 1995.
44. **Simon, J., Chiang, A.C., Bender, W., Shimell, M.J. and O'connor, M.,** Elements of the *Drosophila* bithorax complex that mediate repression by *Polycomb* group products., *Dev. Biol.*, 158, 131-144, 1993.
45. **Fauvarque, M.-S. and Dura, J.-M.,** *polyhomeotic* regulatory sequences induce developmental regulator-dependent variegation and targeted *P*-element insertions in *Drosophila*., *Genes Dev.*, 7, 1508-1520, 1993.
46. **Kassis, J.A.,** Unusual properties of regulatory DNA from the *Drosophila engrailed* gene: three "pairing-sensitive" sites within a 1.6kb region., *Genetics*, 136, 1025-1038, 1994.
47. **Strutt, H., Cavalli, G. and Paro, R.,** Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression., *EMBO J.*, 16(12), 3621-3632, 1997.
48. **Zink, D. and Paro, R.,** *Drosophila* Polycomb-group regulated chromatin inhibits the accessibility of a *trans*-activator to its target DNA., *EMBO J.*, 14, 5660-5671, 1995.
49. **Sigrist, C. and Pirrotta, V.,** Chromatin insulator elements block the silencing of a target gene by the *Drosophila* polycomb response element (PRE) but allow trans interactions between PREs on different chromosomes., *Genetics*, 147(1), 209-221, 1997.
50. **Kennison, J.A. and Tamkun, J.W.,** Dosage-dependent modifiers of *Polycomb* and *Antennapedia* mutations in *Drosophila*, *Proc. Natl. Acad. Sci. USA*, 85, 8136-8140, 1988.
51. **Ingham, P.W.,** Differential expression of *bithorax complex* genes in the absence of the *extra sex combs* and *trithorax* genes, *Nature*, 306, 591-593, 1983.
52. **Breen, T. and Harte, P.,** *trithorax* regulates multiple homeotic genes in the bithorax and Antennapedia complexes and exerts different tissue-specific, parasegment-specific and promoter-specific effects on each, *Development*, 117, 119-134, 1993.
53. **Shearn, A.,** The *ash-1*, *ash-2* and *trithorax* genes of *Drosophila melanogaster* are functionally related, *Genetics*, 121, 517-525, 1989.
54. **Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C. and Kennison, J.A.,** *brahma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2., *Cell*, 68, 561-572, 1992.
55. **Peterson, C.L. and Herskowitz, I.,** Characterization of the yeast *SWI1*, *SWI2*, and *SWI3* genes, which encode a global activator of transcription., *Cell*, 68, 573-583, 1992.

56. **Hirschhorn, J., Brown, S., Clark, C. and Winston, F.,** Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure., *Genes Dev.*, 6(12(A)), 2288-2298, 1992.
57. **Dingwall, A., Beek, S., McCallum, C., Tamkun, J., Kalpana, G., Goff, S. and Scott, M.,** The *Drosophila* snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex., *Mol Biol Cell*, 6(7), 777-791, 1995.
58. **Farkas, G., Guasz, J., Galloni, M., Reuter, G., Gyurkovics, H. and Karch, F.,** The *trithorax-like* gene encodes the *Drosophila* GAGA factor., *Nature*, 371, 806-808, 1994.
59. **Lee, H., Kraus, K., Wolfner, M. and JT, L.,** DNA sequence requirements for generating paused polymerase at the start of hsp70., *Genes Dev*, 6(2), 284-295, 1992.
60. **Croston, G., Kerrigan, L., Lira, L., Marshak, D. and Kadonaga, J.,** Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription., *Science*, 251(4994), 643-649, 1991.
61. **Tsukiyama, T., Becker, P.B. and Wu, C.,** ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor, *Nature*, 367, 525-532, 1994.
62. **Tripoulas, N., LaJeunesse, D., Gildea, J. and Shearn, A.,** The *Drosophila ash1* gene product, which is localized at specific sites on polytene chromosomes, contains a SET domain and a PHD finger, *Genetics*, 143, 913-928, 1996.
63. **Adamson, A. and Shearn, A.,** Molecular genetic analysis of *Drosophila ash2*, a member of the trithorax group required for imaginal disc pattern formation., *Genetics*, 144(2), 621-633, 1996.
64. **Paro, R. and Hogness, D.S.,** The *Polycomb* protein shares a homologous domain with a heterochromatin-associated protein in *Drosophila*, *Proc. Natl. Acad. Sci. USA*, 88, 263-267, 1991.
65. **Paro, R.,** Imprinting a determined state into the chromatin of *Drosophila*, *Trends Genet.*, 6, 416-421, 1990.
66. **Paro, R.,** Mechanisms of heritable gene repression during development of *Drosophila*., *Curr. Op. Cell Biol.*, 5, 999-1005, 1993.
67. **Schlossherr, J., Eggert, H., Paro, R., Cremer, S. and Jack, R.S.,** Gene inactivation in *Drosophila* mediated by the *Polycomb* gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fibre., *Mol. Gen. Genet.*, 243, 453-462, 1994.
68. **Pirrotta, V. and Rastelli, L.,** *White* gene expression, repressive chromatin domains, and homeotic gene regulation in *Drosophila*., *Bioessays*, 16(8), 549-556, 1994.
69. **Pirrotta, V.,** PcG complexes and chromatin silencing, *Curr Opin Genet Dev*, 7(2), 249-258, 1997.

70. **Gutjahr, T., Frei, E., Spicer, C., Baumgartner, S., White, R.A.H. and Noll, M.,** The Polycomb-group gene, *extra sex combs* encodes a nuclear member of the WD-40 repeat family., *EMBO J.*, 14, 4296-4306, 1995.
71. **Messmer, S., Franke, A. and Paro, R.,** Analysis of the functional role of the *Polycomb* chromo domain in *Drosophila melanogaster*., *Genes Dev.*, 6, 1241-1254, 1992.
72. **Buchenau, P., Hodgson, J., Strutt, H. and Arndt-Jovin, D.,** The distribution of Polycomb-group proteins during cell division and development in *Drosophila* embryos: impact on models for silencing., *J. Cell Biol.*, in press, 1998.
73. **Alkema, M.J., Bronk, M., Verhoeven, E., Otte, A., Veer, L.J.v.t., Berns, A. and Lohuizen, M.v.,** Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian Polycomb complex, *Genes and Development*, 11, 226-240, 1997.
74. **Struhl, G.,** Role of the *esc*⁺ gene product in ensuring the selective expression of segment-specific homeotic genes in *Drosophila*, *J. Embryol. Exp. Morphol.*, 76, 297-331, 1983.
75. **Grigliatti, T.,** Position-effect variegation--an assay for nonhistone chromosomal proteins and chromatin assembly and modifying factors., *Methods Cell Biol.*, 35, 587-627, 1991.
76. **Reuter, G. and Spierer, P.,** Position-effect variegation and chromatin proteins., *Bioessays*, 14, 605-612, 1992.
77. **Sinclair, D.A.R., J., C.N., Antonchuk, J., Milne, T.A., Stankunas, K., Ruse, C., Grigliatti, T.A., Kassis, J.A. and Brock, H.W.,** *Enhancer of Polycomb* is a Suppressor of Position-Effect Variegation in *Drosophila melanogaster*, *Genetics*, 148, 211-220, 1998.
78. **Laible, G., Wolf, A., Dorn, R., Reuter, G., Nislow, C., Lebersorger, A., Popkin, D., Pillus, L. and Jenuwein, T.,** Mammalian homologues of the Polycomb-group gene *Enhancer of zeste* mediate gene silencing in *Drosophila* heterochromatin and at *S. cerevisiae* telomeres., *EMBO J*, 16(11), 3219-3232, 1997.
79. **Wu, C.,** Transvection, nuclear structure, and chromatin proteins., *J Cell Biol.*, 120(3), 587-590, 1993.
80. **Lewis, E.,** The theory and application of a new method of detecting chromosomal rearrangements in *Drosophila melanogaster*., *Am. Nat.*, 89, 73-89, 1954.
81. **Wu, C.-t. and Howe, M.,** A genetic analysis of the *Suppressor 2 of zeste* complex of *Drosophila melanogaster*, *Genetics*, 140, 139-181, 1995.
82. **Wu, C.-T., Jones, R.S., Lasko, P.F. and Gelbart, W.M.,** Homeosis and the interaction of *zeste* and *white* in *Drosophila*, *Mol. Gen. Genet.*, 218, 559-564, 1989.

83. **Kalisch, W.E. and Rasmuson, B.,** Changes of *zeste* phenotype induced by autosomal mutations in *Drosophila melanogaster*, *Hereditas*, 78, 97-104, 1974.
84. **Rine, J. and Herskowitz, I.,** Four genes responsible for a position effect on expression from *HML* and *HMR* in *Saccharomyces cerevisiae*, *Genetics*, 116, 9-22, 1987.
85. **Aparicio, O., Billington, B. and Gottschling, D.,** Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*., *Cell*, 66(6), 1279-1287, 1991.
86. **Laurenson, P. and Rine, J.,** Silencers, silencing and heritable transcriptional states, *Microbiol Rev*, 56, 543-560, 1992.
87. **Moretti, P., Freeman, K., Coodly, L. and Shore, D.,** Evidence that a complex of SIR proteins interacts with the silencer and telomere binding protein RAP1, *Genes Dev*, 8, 2257-2269, 1994.
88. **Varanasi, U., Klis, M., Mikesell, P. and Trumbly, R.,** The Cyc8 (Ssn6)-Tup1 corepressor complex is composed of one Cyc8 and four Tup1 subunits., *Mol Cell Biol*, 16, 6707-6714, 1996.
89. **Smith, R., Redd, M. and Johnson, A.,** The tetratricopeptide repeats of Ssn6 interact with the homeo domain of $\alpha 2$., *Genes Dev*, 8, 2857-2867, 1995.
90. **Treitel, M. and Carlson, M.,** Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein, *Proc Natl Acad Sci*, 92, 3132-3136, 1995.
91. **Cooper, J., Roth, S. and Simpson, R.,** The global transcriptional regulators SSN6 and TUP1, play distinct roles in the establishment of a repressive chromatin structure., *Genes Dev*, 8, 1400-1410, 1994.
92. **Edmondson, D., Smith, M. and Roth, S.,** Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4, *Genes Dev*, 10, 1247-1259, 1996.
93. **Herschbach, B., Arnaud, M. and Johnson, A.,** Transcriptional repression directed by the yeast $\alpha 2$ protein *in vitro*, *Nature*, 370, 309-311, 1994.
94. **Thompson, C., Koleske, A., Chao, D. and Young, R.,** A multisubunit complex associated with the RNA polymerase CTD and TATA-binding protein in yeast, *Cell*, 73, 1361-1375, 1993.
95. **Kim, Y.-J., Bjorklund, S., Li, Y., Sayre, M. and Kornberg, R.,** A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II, *Cell*, 77, 599-608, 1994.
96. **Carlson, M.,** GENETICS OF TRANSCRIPTIONAL REGULATION IN YEAST: Connections to the RNA Polymerase II CTD, *Annu Rev Cell Dev Biol*, 13, 1-23, 1997.

97. **Gunster, M.J., Satijn, D.P.E., Hamer, K.M., Blaauwen, J.L.d., Bruijn, D.d., Alkema, M.J., Lohuizen, M.v., Driel, R.v. and Otte, A.P.,** Identification and Characterization of Interactions between the Vertebrate Polycomb-Group Protein BMI1 and Human Homologues of Polyhomeotic, *Mol. Cell. Biol.*, 17(4), 2326-2335, 1997.
98. **Reijnen, M.J., Hamer, K.M., Blaauwen, J.L.d., Lambrechts, C., Schoneveld, I., Driel, R.v. and Otte, A.P.,** Polycomb and *bmi-1* homologs are expressed in overlapping patterns in *Xenopus* embryos and are able to interact with each other, *Mech. Dev.*, 53, 35-46, 1995.
99. **Hodgson, J.W., Cheng, N.N., Sinclair, D.A.R., Kyba, M., Randsholt, N.B. and Brock, H.W.,** The *polyhomeotic* locus of *Drosophila melanogaster* is transcriptionally and post-transcriptionally regulated during embryogenesis, *Mech Dev.*, 66(1-2), 69-81, 1997.
100. **Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W. and Schreiber, S.L.,** Structural basis for the binding of proline-rich peptides to SH3 domains., *Cell*, 76(Cell 76 (5): 933-945), 933-945, 1994.
101. **Nomura, M., Takihara, Y. and Shimada, K.,** Isolation and characterization of retinoic acid-inducible cDNA clones in F9 cells: one of the early inducible clones encodes a novel protein sharing several highly homologous regions with a *Drosophila polyhomeotic* protein., *Differentiation*, 57, 39-50, 1994.
102. **Bornemann, D., Miller, E. and Simon, J.,** The *Drosophila Polycomb* group gene *Sex comb on midleg (Scm)* encodes a zinc finger protein with similarity to polyhomeotic protein, *Development*, 122, 1621-1630, 1996.
103. **Ponting, C.P.,** SAM: a novel motif in yeast sterile and *Drosophila* polyhomeotic proteins, *Protein Sci*, 4, 1928-1930, 1995.
104. **Brunk, B.P., Martin, E.C. and Adler, P.N.,** *Drosophila* genes *Posterior sex combs* and *Suppressor two of zeste* encode proteins with homology to the murine *bmi-1* oncogene, *Nature*, 353, 351-353, 1991.
105. **van Lohuizen, M., Frasch, M., Wientjens, E. and Berns, A.,** Sequence similarity between the mammalian *bmi-1* proto-oncogene and the *Drosophila* regulatory genes *Psc* and *Su(z)2*, *Nature*, 353, 353-355, 1991.
106. **Tagawa, M., Sakamoto, T., Shigemoto, K., Matsubara, H., Tamura, Y., Ito, T., Nakamura, I., Okitsu, A., Imai, K. and Taniguchi, M.,** Expression of novel DNA-binding protein with zinc finger structure in various tumor cells., *J Biol Chem*, 265(32), 20021-20026, 1990.
107. **Irminger-Finger, I. and Nöthiger, R.,** The *Drosophila melanogaster* gene *lethal(3)Ah* encodes a ring finger protein homologous to the oncoproteins MEL-18 and BMI-1, *Gene*, 163, 203-208, 1995.
108. **Pearce, J.H.H., Singh, P.B. and Gaunt, S.J.,** The mouse has a *Polycomb*-like chromobox gene, *Development*, 114, 921-929, 1992.

109. **Strutt, H. and Paro, R.**, The polycomb group protein complex of *Drosophila melanogaster* has different compositions at different target genes., *Mol. Cell Biol.*, 17(12), 6773-6783, 1997.
110. **Golemis, E.A., Gyuris, J. and Brent, R.**, Two hybrid systems/interaction traps, in *Current protocols in molecular biology*, Ausubel, F.M., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J.A. and Struhl, K., Eds., New York, John Wiley & Sons, 1993, 13.14.1-13.14.17.
111. **Estojak, J., Brent, R. and Golemis, E.A.**, Correlation of two-hybrid affinity data with in vitro measurements, *Mol Cell Biol*, 15, 5820-5829, 1995.
112. **Peterson, A., Kyba, M., Borneman, D., Morgan, K., Brock, H. and Simon, J.**, A Domain Shared by the *Polycomb* Group Proteins Scm and ph Mediates Heterotypic and Homotypic Interactions., *Mol. Cell Biol.*, 17(11), 6683-6692, 1997.
113. **Wismar, J., Loffler, T., Habtermichael, N., Vef, O., Geissen, M., Zirwes, R., Altmeyer, W., Sass, H. and Gateff, E.**, The *Drosophila melanogaster* tumor suppressor gene *lethal(3)malignant brain tumor* encodes a proline-rich protein with a novel zinc finger., *Mech Dev.*, 53, 141-154, 1995.
114. **Rhodes, N., Connell, L. and Errede, B.**, STE11 is a protein kinase required for cell-type-specific transcription and signal transduction in yeast., *Genes Dev*, 4, 1862-1874, 1990.
115. **Wang, Y., Xu, H.P., Riggs, M., Rodgers, L. and Wigler, M.**, *byr2*, a *Schizosaccharomyces pombe* gene encoding a protein kinase capable of partial suppression of the *ras1* mutant phenotype., *Mol. Cell. Biol.*, 11, 3554-3563, 1991.
116. **Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Smaldon, N., Smith, A., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohldman, P.**, 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*, *Nature*, 368(6466), 32-38, 1994.
117. **Schultz, J., Ponting, C.P., Hofmann, K. and Bork, P.**, SAM as a protein interaction domain involved in developmental regulation., *Protein Sci*, 6, 249-253, 1997.
118. **Barr, M.M., Tu, H., Aelst, L.v. and Wigler, M.**, Identification of Ste4 as a Potential Regulator of Byr2 in the Sexual Response Pathway of *Schizosaccharomyces pombe*, *Mol. Cell. Biol.*, 16(10), 5597-5603, 1996.
119. **Boulukos, K., Pognonec, P., Rabault, B., Begue, A. and Ghysdael, J.**, Definition of an Ets1 protein domain required for nuclear localization in cells and DNA-binding activity *in vitro*., *Mol. Cell. Biol.*, 9, 5718-5721, 1989.

120. **Klammt, C.**, The *Drosophila* gene *pointed* encodes two ETS-like proteins which are involved in the development of the midline glial cells., *Development*, 117, 163-176, 1993.
121. **Golub, T., Barker, G., Bohlander, S., Hiebert, S., Ward, D., Bray-Ward, P., Morgan, E., Raimondi, S., Rowley, J. and Gilliland, D.**, Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia., *Proc Natl Acad Sci U S A*, 92(11), 4917-4921, 1995.
122. **Carroll, M., Tomasson, M., Barker, G., Golub, T. and Gilliland, D.**, The TEL/platelet-derived growth factor beta receptor (PDGF beta R) fusion in chronic myelomonocytic leukemia is a transforming protein that self-associates and activates PDGF beta R kinase-dependent signaling pathways., *Proc Natl Acad Sci U S A*, 93(25), 14845-14850, 1996.
123. **Jousset, C., Carron, C., Boureux, A., Quang, C.T., Oury, C., Dusanter-Fourt, I., Charon, M., Levin, J., Bernard, O. and Ghysdael, J.**, A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerization interface essential to the mitogenic properties of the TEL-PDGFR β oncoprotein, *EMBO J.*, 16(1), 69-82, 1997.
124. **Dura, J.-M., Randsholt, N.B., Deatrick, J., Erk, I., Santamaria, P., Freeman, J.D., Freeman, S.J., Weddell, D. and Brock, H.W.**, A complex genetic locus, *polyhomeotic*, is required for segmental specification and epidermal development in *D. melanogaster*, *Cell*, 51, 829-839, 1987.
125. **Bender, L., Lo, H.S., Le, H., Kokojan, V., Peterson, V. and Bender, A.**, Associations among PH and SH3 domain-containing proteins and Rho-type GTPases in Yeast, *J. Cell Biol.*, 133, 879-894, 1996.
126. **Matsui, Y., Matsui, R., Akada, R. and Toh-e, A.**, Yeast src homology region 3 domain-binding proteins involved in bud formation, *J. Cell Biol.*, 133, 865-878, 1996.
127. **Choi, K., Satterberg, B., Lyons, D.M. and Elion, E.A.**, Ste5 Tethers Multiple Protein Kinases in the MAP Kinase Cascade Required for Mating in *S. cerevisiae*, *Cell*, 78, 499-512, 1994.
128. **Kranz, J.E., Satterberg, B. and Elion, E.A.**, The MAP kinase Fus3 associates with and phosphorylates the upstream signaling component Ste5, *Genes Dev*, 8, 313-327, 1994.
129. **Marcus, S., Polverino, A., Barr, M. and Wigler, M.**, Complexes between STE5 and components of the pheromone-responsive mitogen-activated protein kinase module, *Proc. Natl. Acad. Sci. USA*, 91, 7762-7766, 1994.
130. **Printen, J.A. and Sprague, G.F.J.**, Protein-Protein Interactions in the Yeast Pheromone Response Pathway: Ste5p Interacts With All Members of the MAP Kinase Cascade, *Genetics*, 138, 609-619, 1994.

131. **Paroush, Z., Finley, R.L., Kidd, T., Wainwright, S.M., P. W. Ingham, R., Brent and Ish-Horowicz, D.,** Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins, *Cell*, 79, 805-815, 1994.
132. **Hart, M., Wang, L. and Coulter, D.,** Comparison of the Structure and Expression of *odd-skipped* and Two Related Genes That Encode a New Family of Zinc Finger Proteins in *Drosophila*, *Genetics*, 144, 171-182, 1996.
133. **Wang, L. and Coulter, D.,** bowel, an odd-skipped homolog, functions in the terminal pathway during *Drosophila* embryogenesis., *EMBO J.*, 15(12), 3182-3196, 1996.
134. **Stolow, D. and Haynes, S.,** Cabeza, a *Drosophila* gene encoding a novel RNA binding protein, shares homology with EWS and TLS, two genes involved in human sarcoma formation., *Nucleic Acids Res.*, 23(5), 835-843, 1995.
135. **Sinclair, D., Milne, T., Hodgson, J., Shellard, J., Salinas, C., Kyba, M., Randazzo, F. and Brock, H.,** The *Additional sex combs* gene of *Drosophila* encodes a chromatin protein that binds to shared and unique Polycomb group sites on polytene chromosomes., *Development*, 125(7), 1207-1216, 1998.
136. **Jeffrey, P., Russo, A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J. and Pavletich, N.,** Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex, *Nature*, 376, 313, 1995.
137. **Horne, M., Goolsby, G., Donaldson, K., Tran, D., Neubauer, M. and Wahl, A.,** Cyclin G1 and Cyclin G2 Comprise a New Family of Cyclins with Contrasting Tissue-specific and Cell Cycle-regulated Expression., *J. Biol. Chem.*, 271(11), 6050-6061, 1996.
138. **Tamura, K., Kanaoka, Y., Jinno, S., Nagata, A., Ogiso, Y., Shumizu, K., Hayakawa, T., Nojima, H. and Okayama, H.,** Cyclin G: a new mammalian cyclin with homology to fission yeast Cig1, *Oncogene*, 8, 2113-2118, 1993.
139. **Okamoto, K. and Beach, D.,** Cyclin G is a transcriptional target of the p53 tumor suppressor protein, *EMBO J.*, 13(19), 4816-22, 1994.
140. **Rechsteiner, M. and Rogers, S.,** PEST sequences and regulation by proteolysis, *Trends Biochem. Sci.*, 21(7), 267-271, 1996.
141. **Kanaoka, Y., Kimura, S., Okazaki, I., Ikeda, M. and Nojima, H.,** GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain., *FEBS Lett.*, 402, 73-80, 1997.
142. **Okamoto, K., Kamibayashi, C., Serrano, M., Prives, C., Mumby, M. and Beach, D.,** p53-Dependent Association between Cyclin G and the B' Subunit of Protein Phosphatase 2A, *Mol. Cell. Biol.*, 16(11), 6593-6602, 1996.
143. **Tehrani, M., Mumby, M. and Kamibayashi, C.,** Identification of a novel protein phosphatase 2A regulatory subunit highly expressed in muscle., *J. Biol. Chem.*, 271, 5164-5170, 1996.

144. **Arion, D., Meijer, L., Brizuela, L. and Beach, D.,** cdc2 is a component of the M phase-specific histone H1 kinase: evidence for identity with MPF., *Cell*, 55(2), 371-378, 1988.
145. **Lu, M., Mpoke, S., Dadd, C. and Allis, C.,** Phosphorylated and dephosphorylated linker histone H1 reside in distinct chromatin domains in Tetrahymena macronuclei., *Mol Biol Cell*, 6(8), 1077-1087, 1995.
146. **Shiekhhattar, R., Mermelstein, F., Fisher, R., Drapkin, R., Dynlacht, B., Wessling, H., Morgan, D. and Reinberg, D.,** Cdk-activating kinase complex is a component of human transcription factor TFIIH., *Nature*, 374(6519), 283-287, 1995.
147. **Serizawa, H., Makela, T., Conaway, J., Conaway, R., Weinberg, R. and Young, R.,** Association of Cdk-activating kinase subunits with transcription factor TFIIH., *Nature*, 374(6519), 280-282, 1995.
148. **Liao, S., Zhang, J., Jeffery, D., Koleske, A., Thompson, C., Chao, D., Viljoen, M., van Vuuren, H. and Young, R.,** A kinase-cyclin pair in the RNA polymerase II holoenzyme., *Nature*, 374(6518), 193-196, 1995.
149. **Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G.,** The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3-9* combines domains of antagonistic regulators of homeotic gene complexes., *EMBO J.*, 13, 3822-3831, 1994.
150. **Dombradi, V., Axton, J., Barker, H. and Cohen, P.,** Protein phosphatase 1 activity in *Drosophila* mutants with abnormalities in mitosis and chromosome condensation., *FEBS Lett.*, 275(1-2), 39-43, 1990.
151. **Kennison, J.,** The Polycomb and trithorax group proteins of *Drosophila*: trans-regulators of homeotic gene function., *Annu Rev Genet.*, 29:, 289-303, 1995.
152. **Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B. and Evans, G.,** A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias., *Nat Genet*, 2(2), 113-118, 1992.
153. **Gu, Y., Nakamura, T., Adler, H., Prasad, R., Canaani, O., Cimino, G., Croce, C.M. and Canaani, E.,** The t(4;11) chromosome translocation of human acute leukemias fuses the *ALL-1* gene, related to *Drosophila trithorax*, to the *AF-4* gene, *Cell*, 71, 701-708, 1992.
154. **Tkachuk, D.C., Kohler, S. and Cleary, M.L.,** Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias, *Cell*, 71, 691-700, 1992.
155. **Nislow, C., Ray, E. and Pillus, L.,** SET1, A Yeast Member of the Trithorax Family, Functions in Transcriptional Silencing and Diverse Cellular Processes., *Mol Biol Cell*, 8(12), 2421-2436, 1997.
156. **Cimino, G., Lo Coco, F., Biondi, A., Elia, L., Luciano, A., Croce, C., Masera, G., Mandelli, F. and Canaani, E.,** ALL-1 gene at chromosome 11q23 is consistently altered in acute leukemia of early infancy., *Blood*, 82(2), 544-456, 1993.

157. **Bernard, O. and Berger, R.**, Molecular basis of 11q23 rearrangements in hematopoietic malignant proliferations., *Genes Chromosomes Cancer*, 13, 75-85, 1995.
158. **Prasad, R., Yano, T., Sorio, C., Nakamura, T., Rallapalli, R., Gu, Y., Leshkowitz, D., Croce, C. and Canaani, E.**, Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia, *Proc. Nat. Sci. USA*, 91, 12160-12164, 1995.
159. **Rubnitz, J., Morrissey, J., Savage, P. and Ceary, M.**, ENL, the gene fused with HRX in t(11;19) leukemias, encodes a nuclear protein with transcriptional activation potential in lymphoid and myeloid cells, *Blood*, 84, 1747-1752, 1994.
160. **Slany, R., Lavau, C. and Cleary, M.**, The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX., *Mol Cell Biol*, 18(1), 122-129, 1998.
161. **Schneider, I.**, Cell lines derived from late embryonic stages of *Drosophila melanogaster*., *J Embryol Exp Morphol*, 27(2), 353-365, 1972.
162. **Bunker, C.A. and Kingston, R.E.**, Transcriptional repression by *Drosophila* and mammalian *Polycomb* group proteins in transfected mammalian cells., *Mol. Cell Biol.*, 14, 1721-1732, 1994.
163. **Keleher, C., Redd, M., Schultz, J., Carlson, M. and Johnson, A.**, Ssn6-Tup1 is a general repressor of transcription in yeast., *Cell*, 68(4), 709-719, 1992.
164. **Brachmann, C., Sherman, J., Devine, S., Cameron, E., Pillus, L. and Boeke, J.**, The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability., *Genes Dev*, 9(23), 2888-2902, 1995.
165. **Sternberg, P., Stern, M., Clark, I. and Herskowitz, I.**, Activation of the yeast HO gene by release from multiple negative controls., *Cell*, 48(4), 567-577, 1987.
166. **Kuchin, S., Yeghiayan, P. and Carlson, M.**, Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8 contribute to transcriptional control in yeast., *Proc Natl Acad Sci U S A*, 92(9), 4006-4010, 1995.
167. **Wahi, M. and Johnson, A.**, Identification of Genes Required for $\alpha 2$ Repression in *Saccharomyces cerevisiae*, *Genetics*, 140, 79-90, 1995.
168. **Lustig, A., Liu, C., Zhang, C. and Hanish, J.**, Tethered Sir3p Nucleates Silencing at Telomeres and Internal Loci in *Saccharomyces cerevisiae*, *Mol Cell Biol*, 16(5), 2483-2495, 1996.
169. **LaJeunesse, D. and Shearn, A.**, *E(z)*: a polycomb group gene or a trithorax group gene?, *Development*, 122, 2189-2197, 1996.
170. **Gyuris, J., Golemis, E., Chertkov, H. and Brent, R.**, Cdi1, a Human G1 and S Phase Protein Phosphatase That Associates with Cdk2, *Cell*, 75, 791-803, 1993.
171. **Bartel, P. and Fields, S.**, Analyzing protein-protein interactions using two-hybrid system., *Methods Enzymol*, 254, 241-263, 1995.

172. **DeCamillis, M.A., Cheng, N., Pierre, D. and Brock, H.W.,** The *polyhomeotic* gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome binding sites with *Polycomb*, *Genes Dev.*, 6, 223-232, 1992.
173. **Jones, R.S. and Gelbart, W.M.,** The *Drosophila* Polycomb-group gene *Enhancer of zeste* shares a domain of sequence similarity with *trithorax*, *Mol. Cell Biol.*, 13(10), 6357-66, 1993.
174. **Sadowski, Bell, B., Broad, P. and Hollis, M.,** GAL4 fusion vectors for expression in yeast or mammalian cells., *Gene*, 118(1), 137-141, 1992.
175. **Thummel, C., Boulet, A. and Lipshitz, H.,** Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection, *Gene*, 74, 445-456, 1988.
176. **Schneuwly, S., Klemenz, R. and Gehring, W.J.,** Redesigning the body plan of *Drosophila* by ectopic expression of the homeotic gene *Antennapedia*, *Nature*, 325, 816-818, 1987.
177. **Lin, Y., Carey, M., Ptashne, M. and Green, M.,** GAL4 derivatives function alone and synergistically with mammalian activators in vitro., *Cell*, 54(5), 659-664, 1988.
178. **Schägger, H. and Jagow, G.v.,** Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa., *Anal. Biochem.*, 166, 368-379, 1987.
179. **Hirschhorn, J., Bortvin, A., Ricupero-Hovasse, S. and Winston, F.,** A new class of histone H2A mutations in *Saccharomyces cerevisiae* causes specific transcriptional defects in vivo., *Mol. Cell. Biol.*, 15(4), 1999-2009, 1995.
180. **Pillus, L. and Rine, J.,** Epigenetic Inheritance of Transcriptional States in *S. cerevisiae*, *Cell*, 59, 637-647, 1989.
181. **Mann, R. and Grunstein, M.,** Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL1 gene in vivo., *EMBO J*, 11(9), 3297-3306, 1992.
182. **Breeden, L. and Nasmyth, K.,** Regulation of the yeast HO gene, *Cold Spring Harb Symp Quant Biol*, 50, 643-650, 1985.
183. **Heberlein, U., England, B. and Tjian, R.,** Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene., *Cell*, 41(3), 965-977, 1985.
184. **Parker, C. and Topol, J.,** A *Drosophila* RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity., *Cell*, 36(2), 357-369, 1984.
185. **Benson, M. and Pirrotta, V.,** The product of the *Drosophila* *zeste* gene binds to specific DNA sequences in white and Ubx., *EMBO J*, 6(5), 1387-1392, 1987.
186. **Bickel, S. and Pirrotta, V.,** Self-association of the *Drosophila* *zeste* protein is responsible for transvection effects., *EMBO J*, 9(9), 2959-2967, 1990.

187. Landecker, H.L., Sinclair, D.A.R. and Brock, H.W., A screen for enhancers of *Polycomb* and *Polycomblike* in *Drosophila melanogaster*., *Dev. Genet.*, 15, 425-434, 1994.

Appendix A: Interactor Preliminary Sequence

z1

```

*               *               *
GAATTCGGCACGAGGCGGTCCGACA AGGACTCGATGCCATGTTACGGCAG TGACTTCCAGATCACCACATCGGCG
CTTAAGCCGTGCTCCGCCAGGCTGT TCCTGAGCTACGGTACAATGCCGTC ACTGAAGGTCTAGTGGTGTAGCCGC
E F G T R R S D K D S M P C Y G S D F Q I T T S A>

100
*               *               *
CAGTGTGACGAGCGAAAGCTGTATG CCCGCAAGGAGGACATTCTGCATGA AGTACTGAACATGCTGCCTCTGCTG
GTCACACTGCTCGCTTTTCGACATAC GGGCGTTTCCTCCTGTAAGACGTACT TCATGACTTGTACGACGGAGACGAC
Q C D E R K L Y A R K E D I L H E V L N M L P L L>

200
*               *               *
AAGCCGGGCAATGAGGAGGCCAAGC TTATCTACCTGACCCTCATAACAGT TGCCGTCAAGGACACCATGCAGCAA
TTCCGGCCCGTTACTCCTCCGGTTCG AATAGATGGACTGGGAGTATGGTCA ACGGCAGTTCCTGTGGTACGTCGTT
K P G N E E A K L I Y L T L I P V A V K D T M Q Q>

300
*               *               *
ATTGTGCCCCACGAGTTGGTGCAGC AGATCTTCTCGTACCTACTCATCCA TCCAGCTATCACCAGCGAGGACAGA
TAACACGGGTGCCTCAACCACGTCG TCTAGAAGAGCATGGATGAGTAGGT AGGTCGATAGTGGTCGCTCCTGTCT
I V P T E L V Q Q I F S Y L L I H P A I T S E D R>

400
*               *               *
CGTTGCTCAACATTTGGCTGCGTC ACTTGGAGGATCATATCCAAGCGGG TTGTGGCGGCCTGACAAATCGCAG
GCAAGCGAGTTGTAAACCGACGCAG TGAACCTCCTAGTATAGGTTTCGCCC AACACCGCCCGGACTGTTTACGCTC
R S L N I W L R H L E D H I Q A G C G G P D K S Q>

500
*               *               *
TTACTTCCTGCAGCCCTCGCCGCAA CTGGTCGCTGGTGGGTAGCTCAACA GGCAGTGGTAGCTTGTTCCTCTTCC
AATGAAGGACGTCGGGAGCGGCGTT GACCAGCGACCACCCATCGAGTTGT CCGTCACCATCGAACAAGGAGAAGG
L L P A A L A A T G R W W V A Q Q A V V A C S S S>

600
*               *               *
GGGGGACCANCTCTTCCGACAGGAT CCTGTTCCGTGGGTGGCCTCATCC TCGGTTGGTGCCCCGCAAGCGGGAG
CCCCCTGGTNGAGAAGGCTGTCTTA GGACAAGGCAGCCACCGGAGTAGG AGCCAACCACGGGGCGTTTCGCCCTC
G G P X L P T G S C S V G W P H P R L V P R K R E>

700
*               *               *
CCGTTCTCTAACGCACCAACGACTG GCAAACGGATCGCCCCGCCAGGAA GCAATTGGAAAACAAATTGGCCGGT
GGCAAGGAGTTGCGTGGTTGCTGAC CGTTTGCCTAGCGGGGCGGGTCCTT CGTTAACCTTTTGTTTAACCGGCCA
P F L N A P T T G K R I A P P R K Q L E N K L A G>

GATTGG
CTAACC
D W>

```


* * *
 TTCCTTATGATGNCCCAGATATATG CCTCTCCCGACTTCGGCACGAGGCG GATCACCGAGCGTGAGAAGAACAAG
 AAGGAATACTACNGGGTCTATATAC GGAGAGGGCTGAAGCCGTGCTCCGC CTAGTGGCTCGCACTCTTCTTGTTC
 F L M M X Q I Y A S P D F G T R R I T E R E K N K>

100

* * *
 AAGCGCGATGTGCGCGGCTGGTATG AGCCAACGATCGCCCCGGAAGGAGT CGTGGATCACAGACACCAGGAGGTG
 TTCGCGCTACACGCGCCGACCATACTCGGTTGCTAGCGGGCCCTTCCTCA GCACCTAGTGTCTGTGGTCTCTCCAC
 K R D V R G W Y E P T I A R E G V V D H R H Q E V>

200

* * *
 CCAACGGACGTGGAGCGCGGCGACA TTCCCGTTCTGAATGGCGATTGCGA AGACGCCCTCGCACGATCGCTCAGC
 GGTGCGCTGCACCTCGCGCCGCTGT AAGGGCAAGACTTACCGCTAACGCT TCTGCGGGAGCGTGCTAGCGAGTCG
 P T D V E R G D I P V L N G D C E D A L A R S L S>

300

* * *
 GATTTACTGGCTCTGGTGAAGCTGC TCCGCGAAGACGTGCGCCACCAGCG CCAGGAGATTGCCTACCTGCGTATG
 CTAATGACCGAGACCACTTCGACG AGGCGCTTCTGCAGCGGTGGTGC GGTCTCTAACGGATGGACGCATAC
 D L L A L V K L L R E D V A H Q R Q E I A Y L R M>

* * *
 CTCCTGGAGAACTGTGCCGGCTGCA AGAATCCCCTCACCACCGATAACCA ACTGCGCATCGAGCCCGACTGCCGT
 GAGGACCTCTTGACACGCGCGACGT TCTTAGGGGAGTGGTGGCTATTGGT TGACGCGTAGCTCGGGCTGACGGCA
 L L E N C A G C K N P L T T D N Q L R I E P D C R>

400

* * *
 TCCGCCAATCCCTGTTATCCTGGAG TGGAGTGCCTTGGACTCGGCGGNCGG TCCCCGATGTGGCACTTGTCCCCTT
 AGGCGGTTAGGGACAATAGGACCTC ACCTCACGAACCTGAGCCGCCNGCC AGGGGCTACACCGTGAACAGGGGAA
 S A N P C Y P G V E C L D S A X G P R C G T C P L>

500

* * *
 GGCTTCATTGGCGATGGNAAGAGCT GCAAGNCGGNGTTACCTGCGCCCA TCACATGTGCTATCCAGGCGTCAAG
 CCGAAGTAACCGCTACCNTTCTCGA CGTTCNGCCCNCAATGGACGCGGGT AGTGTACACGATAGGTCCGCGAGTTC
 G F I G D G K S C K X G V T C A H H M C Y P G V K>

600

* * *
 TGTCACGATACCGTGAATTGGAGCC AGTGTGATCCTGTCCAGCCGGCTAC GAGGGTGATGGACGACATGTCCGGTA
 ACAGTGCTATGGCACTTAACCTCGG TCACACTAGGACAGGTGCGCCGATG CTCCCACTACCTGCTGTACAGCCAT
 C H D T V N W S Q C D P V Q P A T R V M D D M S V>

* * *
 CGTATCCTGGCTGGAACACCGTGCC CTCAGGACTCTTTGGTGCCTCGATT ATGTGAACATATCAGAGGGAGGCCA
 GCATAGGACCGACCTTGTGGCACGG GAGTCCTGAGAAACCACGGAGCTAA TACACTTGTATAGTCTCCCTCCGGT
 R I L A G T P C P Q D S L V P R L C E H I R G R P>

z3

```

      *                               *                               *
GAATTCGGCACGAGGCGCCTAATTC TGGACGGAATCATTCACTGCACGGC CAGGGATGAGTTCTCGTACCAGGAG
CTTAAGCCGTGCTCCGCGGATTAAG ACCTGCCTTAGTAAGTCACGTGCCG GTCCCTACTCAAGAGCATGGTCCTC
E F G T R R L I L D G I I Q C T A R D E F S Y Q E>

```

100

```

      *                               *                               *
ATGATATCTTTCTCTGCCGCTCTGCG CCCATCCCAATCCCAAAAAGGTCCT GATCGTGGGCGGTGGTGTATGGCGGC
TACTATAGAAAAGGACGGCGAGACGC GGGTAGGGTTAGGGTTTTTCCAGGA CTAGCACCCGCCACCACTACCGCCG
M I S F L P L C A H P N P K K V L I V G G G D G G>

```

200

```

      *                               *                               *
GTTGCTCGCGAGGTGGTAAAGCATC CACTGGTTCGAGGAAGTGCATCAGGT GGAAATTGACGACCGTGTCTCGAG
CAACGAGCGCTCCACCATTTCTAG GTGACCAGCTCCTTCACGTAGTCCA CCTTTAACTGCTGGCACAGCAGCTC
V A R E V V K H P L V E E V H Q V E I D D R V V E>

```

300

```

      *                               *                               *
CTGTCCAAGCAATATCTCCCAGCGA TGGCCTGTGGTTTCGCCAACGAGAA GTTGAAGCTTACCATTGGCGATGGA
GACAGGTTCTGTATAGAGGGTCGCT ACCGGACACCAAAGCGGTTGCTCTT CAACTTCGAATGGTAACCGCTACCT
L S K Q Y L P A M A C G F A N E K L K L T I G D G>

```

```

      *                               *                               *
TTCGACTATATGAAGAAACACAAGA ACGAATTTGATGTCATCATCACCGA CAGCTCGGATCCCATTGGTCCGGCA
AAGCTGATATACTTCTTTGTGTTCT TGCTTAAACTACAGTAGTAGTGGCT GTCGAGCCTAGGGTAACCAGGCCGT
F D Y M K K H K N E F D V I I T D S S D P I G P A>

```

400

```

      *                               *                               *
GTGAGCCTGTTTCAGGAAAGCTACT ACGAGCTAATGAAACACGCGCTGAA GGATGACGGAATCGTGTGCTCCAG
CACTCGGACAAAGTCCTTTCGATGA TGCTCGATTACTTTGTGCGCGACTT CCTACTGCCTTAGCACACGAGGGTC
V S L F Q E S Y Y E L M K H A L K D D G I V C S Q>

```

500

```

      *                               *                               *
GGCGGTAGCTTCTGGCTGGACCTGG ACTACATCAAGAAGACCATGTCCGG NTGCAAGGAGCACTTGGTTAGGTGG
CCGCCATCGAAGACCGACCTGGACC TGATGTAGTTCTTCTGGTACAGGCC NACGTTCTCGTGAACCAATCCACC
G G S F W L D L D Y I K K T M S G C K E H L V R W>

```

```

      *
CCTATGCCGTACCTCCGTCCGTCC TATCCCTGGGCACATTG
GGATACGGCAGTGGAGGCAGGCAGG ATAGGGACCCGTGTAAC
P M P S P P S V L S L G T L>

```

```

      *                               *                               *
GAATTCGGAACGAGGCGGAAAATGG TCCAACGTCTGACGCTCCGGAGACG CCTGTCCTACAACACACGCTCCAAC
CTTAAGCCTTGCTCCGCCTTTTACC AGGTTGCAGACTGCGAGGCCTCTGC GGACAGGATGTTGTGTGCGAGGTTG
E F G T R R K M V Q R L T L R R R L S Y N T R S N>

```

100

```

      *                               *                               *
AAGCGGCGCATTGTTCGCACGCCCG GTGGTCGTCTGGTTTACCAGTATGT GAAGAAGAACCCACCGTGCCCCGT
TTCGCCGCGTAACAAGCGTGCGGGC CACCAGCAGACCAAATGGTCATACA CTTCTTCTTGGGGTGGCACGGGGCA
K R R I V R T P G G R L V Y Q Y V K K N P T V P R>

```

200

```

      *                               *                               *
TGCGGNCAGTGCAAGGAGAAGTTGA AGGGTATCACCCCTCCCGCCCCAG CGAGCGCCCCCGCATGTCCAAGCGC
ACGCCNGTCACGTTCCTCTTCAACT TCCCATAGTGGGGAGGGCGGGGTC GCTCGGGGGGCGTACAGGTTGCGG
C G Q C K E K L K G I T P S R P S E R P R M S K R>

```

300

```

      *                               *                               *
CTGAAGACCGTGTCAGGACCTACG GTGGAGTGCTGTGCCACAGCTGTCT GCGCGAGCGTNTCGTGCGCGCCTTC
GAGTTCTGGCACAGGTCTCGGATGC CACCTCACGACACGGTGTGACAGA CGCGCTCGCANAGCACGCGCGGAAG
L K T V S R T Y G G V L C H S C L R E R X V R A F>

```

```

      *                               *                               *
CTCATCGAGGAGCAGAAGATCGTCA AGGCCCTGAAGAGCCAGCGNGAGGC GCTCGTCAAGCCGGTGTAAAGCCCC
GAGTAGCTCCTCGTCTTCTAGCAGT TCCGGGACTTCTCGGTGCGNCTCCG CGAGCAGTTCGGCCACATTCCGGGG
L I E E Q K I V K A L K S Q R E A L V K P V *

```

400

```

      *                               *                               *
AAGGNCAAGCCCCGAGACCAAGAAGA AGCCCGCTGCTGGAGCCAAGGGAAC CAAGGGCGGTGNCGGTAAGGTCANC
TTCCNGTTCGGGCTCTGGTTCTTCT TCGGGCGACGACCTCGGTTCCTTG GTTCCCGCCACNGCCATTCCAGTNG

```

500

```

      *                               *                               *
ANGGGTGGGTGCTGGCGCAAGGGA GNCGCTGGNAAGAAGCCCGNCAGA AGCCAGCCGCTTGGAAGCCAGGAA
TNCCCCACCGACCGCGGTTCCCT CNGCGACNTTCTTTCGGGCCNGTCT TCGGTGCGCGAACCTTTCGGTCTCTT

```

600

```

      *                               *                               *
GTNAACAGCCACGCAACGAGTCGGG NGTNTTGNTAANTAATTTTNAATA ATTGGGTTTTTCCACTTGGAATAA
CANTTGTGCGGTGCGTTGCTCAGCCC NCANAACNATTNATTAATAANTTTAT TAACCCAAAAAAGGTGAACCTTTTT

```

```

AAAAAAAAAACTCGAG
TTTTTTTTTTTGAGCTC

```

z11

```

*                               *                               *
GAATTCGGCACGAGGCGGTTCCGCG AGGGCACCTCCGAGTACGACCTGAA GCGGCGGCCAGCCTGGACGGATCGG
CTTAAGCCGTGCTCCGCCAAGGCGC TCCCGTGGAGGCTCATGCTGGACTT CGCCGCCGGTCGGACCTGCCTAGCC
E F G T R R F R E G T S E Y D L K R R P A W T D R>

100
*                               *                               *
ATAATGTACGCCGTGCAGCCACTGA ACCGGCAGCCCGGCATGCAGCTATC CATTGAGCAATGCTCGTATAAGTCC
TATTACATGCGGCACGTCCGGTGA CTGGCCGTGCGGCCGTACGTCGATAG GTAACCTCGTTACGAGCATATTCAGG
I M Y A V Q P L N R Q P G M Q L S I E Q C S Y K S>

200
*                               *                               *
CATCCCCTGTACACCATCAGTGATC ACAAGCCGGTGACCAGTGACTTTAC CATCAAGCTCTACCCGAATGTACGG
GTAGGGGACATGTGGTAGTCACTAG TGTTCGGCCACTGGTCACTGAAATG GTAGTTCGAGATGGGCTTACATGCC
H P L Y T I S D H K P V T S D F T I K L Y P N V R>

300
*                               *                               *
GCGCCCGGCGTGGTGTCTCGCCTC TGTGCTCTTGAAGATTGGGGACGA GAACACGGTGGAGTATCACAAGCAG
CGCGGGCCGCACCACAAGAGCGGAG ACAGCGAGACCTTCTAACCCCTGCT CTTGTGCCACCTCATAGTGTTCGTC
A P G V V F S P L S L W K I G D E N T V E Y H K Q>

400
*                               *                               *
GCAGAGTTCGACGAGGGGTCCAACG ACTGGATTGGNATCTTTCCGTCGGA GTACGCCAGTTTGGCGGATTACGTA
CGTCTCAAGCTGCTCCCCAGGTTGC TGACCTAACCNTAGAAAGGCAGCCT CATGCGGTCAAACCGCCTAATGCAT
A E F D E G S N D W I G I F P S E Y A S L A D Y V>

500
*                               *                               *
GCCTACGAGTATGTCAATCAGGNTG AGTCGGCCTCATCCTCGGACTCCAA TCACCAACCGGGATCCGTTTGGAGA
CGGATGCTCATACAGTTAGTCCNAC TCAGCCGGAGTAGGAGCCTGAGGTT AGTGGTTGGCCCTAGGCAAACCTCT
A Y E Y V N Q X E S A S S S D S N H Q P G S V W R>

600
*                               *                               *
CGGCCTCGCATCATCGAAGGGGGTC GGGCATCATACAGGAATCGCCATGC GACAGGTCGCCATCAGGAGGCTAAT
GCCGGAGCGTAGTAGCTTCCCCCAG CCCGTAGTATGTCTTAGCGGTACG CTGTCCAGCGGTAGTCTCCGATTA
R P R I I E G G R A S Y R N R H A T G R H Q E A N>

700
*                               *                               *
GCCCCAAGAGTTGGTGC GGCTAGATT TCGCCGACGATGTGGAAC TCGGTCA CGGCGAGCAATACCTGTTGATATAT
CGGGTTCTCAACCACGCCGATCTAA AGCGGTGCTACACCTTGACGCAGT GCCGCTCGTTATGGACAAC TATATA
A Q E L V R L D F A D D V E L R H G E Q Y L L I Y>

800
*                               *                               *
TTCCGCAGCACCGGAGTCCGGGGCG TGACCAGTTTGGCCGGCGTCAGTGG TGTCTTTGTGGCGGAGAAGCGGCAC
AAGCGTCGTGGCCTCAGGCCCCGC ACTGGTCAAACCGGCCGAGTCACC ACAGAAACACCGCCTCTTCGCCGTG
F R S T G V R G V T S L A G V S G V F V A E K R H>

```

* * *

GAATTCGGCACGAGGCGGAAACCCCT TCAAATGCACGGAATGCGGCAAGGG ATTTTGCCAATCGAGAACCTTGGCT
CTTAAGCCGTGCTCCGCCTTTGGGA AGTTTACGTGCCTTACGCCGTTCCT TAAAACGGTTAGCTCTTGGAACCGA
E F G T R R K P F K C T E C G K G F C Q S R T L A>

100

* * *

GTCCACAAGATCCTGCACATGGAGG AATCACCCCAAGTGCCCCGTCTG CAGTCGATCATTCATCAGCGCTCC
CAGGTGTTCTAGGACGTGTACCTCC TTAGTGGGGTGTTCACGGGGCAGAC GTCAGCTAGTAAGTTAGTCGCGAGG
V H K I L H M E E S P H K C P V C S R S F N Q R S>

200

* * *

AACCTGAAGACCCATCTGCTCACCC ACACGGATCACAAGCCCTACGAGTG CTCTTCATGCGGCAAAGTTTTCGGC
TTGGACTTCTGGGTAGACGAGTGGG TGTGCCTAGTGTTCGGGATGCTCAC GAGAAGTACGCCGTTCAAAAGGCG
N L K T H L L T H T D H K P Y E C S S C G K V F R>

300

* * *

CGTAAC TCGATCTACGACGCCATG NCTTGACCCATGCAGTGGGTGAGGT CAACTCCGGGGACTATGTGGATGTG
GCATTGACGCTAGATGCTGCGGTAC NGAAC TGGGTACGTACCCACTCCA GTTGAGGCCCTGATACACCTACAC
R N C D L R R H X L T H A V G E V N S G D Y V D V>

* * *

GGCGAAGAGGATGAGGCCAGAAATT TANGTGGCGACGAGGAGGATTCGTT GCTGGAAGTGGA CTGCCCCGCCAG
CCGCTTCTCCTACTCCGGTCTTTAA ATNCACCGCTGCTCCTCCTAAGCAA CGACCTTCACCTGAGCGGGGCGGTC
G E E D E A R N L X G D E E D S L L E V D S P R Q>

400

* * *

TCGCCAGTTCACAACTTGGGCGAGT CTGGTGGATCGGGTGAGAAATCTGA GTCCGAAAGAATGAGACTCAAGCGC
AGCGGTCAAGTGTGAACCCGCTCA GACCACCTAGCCCACTCTTTAGACT CAGGCTTTCTTACTCTGAGTTCGCG
S P V H N L G E S G G S G E K S E S E R M R L K R>

500

* * *

AAGGCAGNCATCGATCATGAGGAAA GCGAAGAGGAGTTCGATGACTTCGA CGAGGAAGAGGAATGCAGGGATCTT
TTCCGTCNGTAGCTAGTACTCCTTT CGCTTCTCCTCAAGCTACTGAAGCT GCTCCTTCTCCTTACGTCCCTAGAA
K A X I D H E E S E E E F D D F D E E E E C R D L>

* * *

GAATTCGGCACGAGGCGGGCCTGCG AGAAGGCTTGGCGCGATTTTATTAT TGCAAAGATGACCCCCAAGCCGCC

CTTAAGCCGTGCTCCGCCCGGACGC TCTTCCGAACCGCGCTAAAATAATA ACGTTTCTACTGGGGGTTTCGGCGGG

E F G T R R A C E K A W R D F I I A K M T P K P P>

100

* * *

CGTATTCACCAGGTGGAGATGGGTT CGGAGCCAATGGATATCAACGAGGA TGAGGCCGATGCACCGGATGATGAT

GCATAAGTGGTCCACCTCTACCCAA GCCTCGGTTACCTATAGTTGCTCCT ACTCCGGCTACGTGGCCTACTACTA

R I H Q V E M G S E P M D I N E D E A D A P D D D>

200

* * *

CTGCCCATGTTGAATCTGGCCTCGT TTGCCATCTACAAGCTGTTTCGCGGA GTGGGAACGGGAGGGCTATGTCGTG

GACGGGTACAACTTAGACCGGAGCA AACGGTAGATGTTTCGACAAGCGCCT CACCCTTGCCCTCCCGATACAGCAC

L P M L N L A S F A I Y K L F A E W E R E G Y V V>

300

* * *

CCCGAGATGCACCCTTCGGCCAATG CTGCCCCAACAGGCGGGAGGGGATGC CGGAACCTCCAGTTCCCCCGTGCCG

GGGCTCTACGTGGGAAGCCGTTAC GACGGGTTGTCCGCCCTCCCTACG GCCTTGAGGTCAAGGGGGGCACGGC

P E M H P S A N A A Q Q A G G D A G T P V P P V P>

* * *

AAGGAGCCAAAGAAGCCGCCAGTGC GCACCGAGCTACCTCTGGCTGGGA GACCATGCACCCGGCGACCATTTCTT

TTCTTCGGTTTCTTCGGCGGTCACG CGTGGCTCGATGGGAGACCGACCCT CTGGTACGTGGGCCGCTGGTAAGAA

K E P K K P P V R T E L P S G W E T M H P A T I L>

400

* * *

TGNATTATGCGTCCGGGACTCAACT ACGTGGGACTACGGGTCATCTGGCG ACAAGANCAACGGCATGCAGCATCT

ACNTAATACGCAGGCCCTGAGTTGA TGCACCCTGATGCCAGTAGACCGC TGTTCTNGTTGCCGTACGTCGTAGA

X I M R P G L N Y V G L R V I W R Q X Q R H A A S>

500

* *

GGGAATCATGGTGGACAACAGGAG TCCACGCCAACGGGAGATCAAAGNA

CCCTTAGTACCACCTGTTGGTCCTC AGGTGCGGTTGCCCTCTAGTTTCNT

G N H G G Q P G V H A N G R S K X

* * *

ATGAGAAACGATGGAGGCATGCGCN ATCGCGGAGGAAGCGGTGGCGGTAA TGGAGGCGGTGGCGGCGGACGCTAC
TACTCTTTGCTACCTCCGTACGCGN TAGCGCCTCCTTCGCCACCGCCATT ACCTCCGCCACCGCCGCTGCGATG
M R N D G G M R X R G G S G G G N G G G G G G R Y>

100

* * *

GATCGCGGAGGAAGCGGTGGTGGTG GCGGCGGCGGTGGCAACNTNCANCC CCGTGATGGTGACTGGAAATGCAAC
CTAGCGCCTCCTTCGCCACCGCCAC CGCCGCCGCCACCGTTGNANGTNGG GGCCTACCACTGACCTTTACGTTG
D R G G S G G G G G G G G N X X P R D G D W K C N>

200

* * *

AGCTGTAATAACACCAACTTCGCCT GCGCAACGAATGCAATAGATGTNA GACTCCCAAGGCGACGACGAGGGC
TCGACATTATTGTTGAAGCGGA CCGCGTTGCTTACGTTATCTACANT CTGAGGGTTCCCGCTGCTGCTCCCG
S C N N T N F A W R N E C N R C X T P K G D D E G>

300

* * *

TCTAGCGGAGGTGGTGAAGCGGCG GCTACCGCGGCGGTGGTGGCGGAGG AGGCTACGACCGAGGAAATGATCGT
AGATCGCCTCCACCACCTTCGCCGC CGATGGCGCGGCCACCGCCCTCC TCCGATGCTGGCTCCTTTACTAGCA
S S G G G G S G G Y R G G G G G G G Y D R G N D R>

* * *

GGATCCGCGGCGGTGGATATCACA ACAGAGATCGCGGTGGCAACTCGCA GGGAGGCGAAGGCGGCGGCGGCGGT
CCTAGGCCGCGCCACCTATAGTGT TGTCTCTAGCGCCACCGTTGAGCGT CCCTCCGCTTCCGCCGCGCGGCCA
G S G G G G Y H N R D R G G N S Q G G E G G G G G G>

400

* * *

GGTGGTGGCTACTCCCGCTTCNATG ACNACNATGGCNGAAGACGCCGTGG CCCTTGAAGTGGTGGCGGCAATCCC
CCACCACCGATGAGGGCGAAGNTAC TGNTGNTACCGNCTTCTGCGGCACC GGGAACTTCACCACCGCGTTAGGG
G G G Y S R F X D X X G X R R R G P *

* *

CGTGATTGTGGACCGATGAGAAACC ATGGAGGCNTGCGC
GCACTAACACCTGGCTACTCTTTGG TACCTCCGNACGCG

GAATTCCGNNACGAGGCGGTCGGCTG CCGAGGAGTACCAGAAGTACATTAA TGCGGATAAGACGACCGTAGCTCTA
CTTAAGCCNTGCTCCGCCAGCCGAC GGCTCCTCATGGTCTTCATGTAATT ACGCCTATTCTGCTGGCATCGAGAT
E F G T R R S A A E E Y Q K Y I N A D K T T V A L>

100

TTCGCCGCCGAATGGGCAGAGCAAT GCGGTCAGGTGAAAGACGCGCTGGA GGAGCTGGCCAAGATTACTGGCGAA
AAGCGGCGGCTTACCCGTCTCGTTA CGCCAGTCCACTTTCTGCGCGACCT CCTCGACCGGTTCTAATGACCGCTT
F A A E W A E Q C G Q V K D A L E E L A K I T G E>

200

AAACTGCAGTTCATCAGCCTAAACG CTGAACAATTTCCCGAGATTTCCAT GAAACATCAGATCGAGGCCGTGCC
TTTGACGTCAAGTAGTCGGATTTGC GACTTGTTAAAGGGCTCTAAAGGTA CTTTGTAGTCTAGCTCCGGCACGGG
K L Q F I S L N A E Q F P E I S M K H Q I E A V P>

300

ACAGTCATATTCTTCGCCAAGGGCT CCGCCGTTGACCGTGTCTGATGGTGT AGACATCGCCGCCATAAGCGCCAAA
TGTCAGTATAAGAAGCGGTTCCCGA GCGGCAACTGGCACAGCTACCACA TCTGTAGCGCGGTATTTCGCGGTTT
T V I F F A K G S A V D R V D G V D I A A I S A K>

TCCAAAAAGTTGGCCGAAAACGCAA GCAGCGCGCGGCAACAGGACAAAC GTTGAGGAACGCCTAAAGGCCCTA
AGGTTTTTCAACCGGCTTTTGCGTT CGTCGCGCGCGCGTGTCTGTTTG CAACCTCCTTGCGGATTTCCGGGAT
S K K L A E N A S S A A A T G Q T L E E R L K A L>

400

ATCAATACAGCTCCGCTGATGATAT TCATGAAGGGCGACCGAAATGGACC GCGTTGCGGATTCTCCAAGCAGCTC
TAGTTATGTCGAGGCGACTACTATA AGTACTTCCCGCTGGCTTTACCTGG CGCAACGCCTAAGAGGTTCTGTCGAG
I N T A P L M I F M K G D R N G P R C G F S K Q L>

500

ATCGGCATTGTGAACGAAACCAACT TGCCGTACGAGACATTTGACATCCT CGGCGACGAAGAAGTGCGTCAAGGC
TAGCCGTAACACTTGCTTTGGTTGA ACGGCATGCTCTGTAAACTGTAGGA GCCGCTGCTTCTTCACGCAGTTCCG
I G I V N E T N L P Y E T F D I L G D E E V R Q G>

600

CTGGTGAAAATACTCCGACTGGCC ACATATCCAGGTTTACGTCAAGGG TGAACCTATCGGCGGCTCGATATTA
GACCACTTTTGATGAGGCTGACCGG TGTATAGGGTCCAAATGCAGTTCCC ACTTGAAATAGCCGCCGAGCTATAAT
L V K T T P T G H I S Q V Y V K G E L I G G S I L>

Appendix B: Interactor Sequence Comparisons

z1 no high scoring matches

z2

	High Score	Smallest Sum Probability P(N)	N
Sequences producing High-scoring Segment Pairs:			
sp P49746 TSP3_HUMAN THROMBOSPONDIN 3 PRECURSOR /pir A57...	57	0.69	2
sp P49746 TSP3_HUMAN THROMBOSPONDIN 3 PRECURSOR pir A57121 thrombospondin 3 precursor - human gi 886299 (L38969) thrombospondin 3 [Homo sapiens] Length = 956			
Score = 57 (26.5 bits), Expect = 1.2, Sum P(2) = 0.69 Identities = 7/17 (41%), Positives = 13/17 (76%)			
Query: 123 DCRSANPCYPGVECLDS 139			
+C A+PC+PG C+++			
Sbjct: 319 ECAHADPCFPGSSCINT 335			
Score = 51 (23.7 bits), Expect = 8.0, Sum P(2) = 1.0 Identities = 7/13 (53%), Positives = 11/13 (84%)			
Query: 126 SANPCYPGVECLD 138			
S NPC+ GV+C++			
Sbjct: 279 SPNPCFRGVDCME 291			
Score = 41 (19.0 bits), Expect = 1.2, Sum P(2) = 0.69 Identities = 7/24 (29%), Positives = 14/24 (58%)			
Query: 85 REDVAHQEQEIAYLRMLLENCAGC 108			
R+D+ Q +E++ +R + C C			
Sbjct: 246 RDDIRDQVKEMSLIRNTIMECQVC 269			

z3

gi|309502 (L19311) spermidine synthase [Mus musculus] gi|1061192
 (Z67748) spermidine synthase [Mus musculus] prf||2113276A
 spermidine synthase [Mus musculus]
 Length = 302

Score = 601 (278.9 bits), Expect = 8.1e-80, P = 8.1e-80
 Identities = 111/163 (68%), Positives = 137/163 (84%)

Query: 7 LILDGIIQCTARDEFSYQEMISFLPLCAHPNPKKVLIVGGDGGVAREVVKHPLVEEVHQ 66
 L+LDG+IQCT RDEFSYQEMI+ LPLC+HPNP+KVLII+GGGDGGV REVVKHP VE V Q

Sbjct: 63 LVLDGVIQCTERDEFSYQEMIANLPLCSHPNPRKVLII+GGDGGVLRREVVKHPSVESVVQ 122

Query: 67 VEIDDRVVELSKQYLPAMACGFANERKLKLTIGDGFDMKKHKNFEDVIITDSSDPGPAV 126
 EID+ V+E+SK++LP MA GF++ KL L +GDGF++MK++++ FDVIITDSSDP+GPA

Sbjct: 123 CEIDEDVIEVSKKFLPGMAVGFSKKLTLHVGDFEFMKQNQDAFDVIITDSSDPMGPAAE 182

Query: 127 SLFQESYYELMKHALKDDGIVCSQGGSFWLDLDYIKKTMGCK 169
 SLF+ESYY+LMK ALK+DGI+C QG WL LD IK+ CK

Sbjct: 183 SLFKESYYQLMKTALKEDGILCCQGEQWLHLDLIKEMRHFCK 225

z7

Sequences producing High-scoring Segment Pairs:

	Score	P(N)	N
sp P45842 RL34_AEDAL 60S RIBOSOMAL PROTEIN L34 (L31) /pir...	396	1.7e-50	1

sp|P45842|RL34_AEDAL 60S RIBOSOMAL PROTEIN L34 (L31) pir||S47637
 ribosomal protein L31 - forest day mosquito gi|506631 (U03871)
 ribosomal protein L31 [Aedes albopictus]
 Length = 130

Score = 396 (182.7 bits), Expect = 1.7e-50, P = 1.7e-50
 Identities = 76/99 (76%), Positives = 84/99 (84%)

Query: 25 MVQRLTLRRRLSYNTRSNKRRIVRTPGGRLVYQYVKNPTVPRCGQCKEXLKGITPSXPS 84
 MVQRLTLRRRLSYNT+SNKRR+VRTPGGRLVY YVKK TVP+CGQCKE L GI PS PS
 Sbjct: 1 MVQRLTLRRRLSYNTKSNKRRVVRTPGGRLVYLYVKKQRTVPKCGQCKEKLSGIKPSRPS 60

Query: 85 ERPRMSKRLXTVSRTXGGVLCHSXLRRXVRASLIEEQR 123
 ERPRM +RL TV+RT GGVLCH LR R +RA LI+EQ+
 Sbjct: 61 ERPRMCRRLLKTVTRTFGGVLCHRLRERIIRAFLIDEQK 99

z11

Sequences producing High-scoring Segment Pairs:

	High Score	Smallest Sum Probability	P(N)	N
gi 1399101 (U45973) phosphatidylinositol (4,5)b...	177	3.0e-22		2

gi|1399101 (U45973) phosphatidylinositol (4,5)bisphosphate
 5-phosphatase homolog; has similarity to motifs conserved in
 phosphatidylinositol (4,5)bisphosphate 5-phosphatases, Swiss-Prot
 Accession Number Q01968, and the product of GenBank Accession Numb...
 Length = 329

Score = 177 (81.6 bits), Expect = 3.0e-22, Sum P(2) = 3.0e-22
 Identities = 34/93 (36%), Positives = 49/93 (52%)

Query: 16 KRRXAWTDRIMYAVQ 30
 KR+ AWTDRI++ ++
 Sbjct: 143 KRKPAWTDRIWLRLK 157

Query: 36 PGMQLSIEQCSYKSHPLYTISDHKPVTSDFTIKLYPNVRAPGVVFSPLSLWKIGDENTVE 95
 P S+ Y SH Y ISDHKPV+ F ++L P V AP +V P LW + ++ V
 Sbjct: 170 PASHFSLSLRGYSSHMTYGISDHKPVSGTFDLELKLPLVSAPLIVLMPEDLWTVENDMMVS 229

Query: 96 YHKQAEFDEGSNDWIGIFPSEYASLADYVAYEY 128
 Y ++F DWIG++ + DYV+Y +
 Sbjct: 230 YSSTSDFPSPWDWIGLYKVGLRDVNDYVSYAW 262

Score = 50 (23.0 bits), Expect = 3.0e-22, Sum P(2) = 3.0e-22
 Identities = 8/15 (53%), Positives = 13/15 (86%)

z28

Sequences producing High-scoring Segment Pairs:				Smallest Sum	High Score	Probability P(N)	N
gi 1388166	(U58282) Bowel [Drosophila melanogas...	894	1.4e-121	2			
gi 1480194	(U62004) Sob protein [Drosophila mel...	435	1.2e-55	2			
sp P23803 ODD_DROME	ODD-SKIPPED PROTEIN /gi 296793 (X574...	263	1.9e-29	1			
gb AA141582 AA141582	CK02065.3prime Drosophila Embryo... +2	421	8.3e-72				

Score = 894 (411.7 bits), Expect = 1.4e-121, Sum P(2) = 1.4e-121
Identities = 164/174 (94%), Positives = 168/174 (96%)

Query: 4 TRRKPFKCTECGKGFCQSRTLAVHKILHMEESPHKCPVCSRSFNQRSNLKTHLLTHTDHK 63
++ KPFKCTECGKGFCQSRTLAVHKILHMEESPHKCPVCSRSFNQRSNLKTHLLTHTDHK
Sbjct: 289 SKEKPFKCTECGKGFCQSRTLAVHKILHMEESPHKCPVCSRSFNQRSNLKTHLLTHTDHK 348

Query: 64 PYECSSCGKVFRNCDLRRHXLTHAVGEVNSGDYVDVGEEDDEARNLXGDEEDSLLEVDSP 123
PYECSSCGKVFRNCDLRRH LTHAVGEVNSGDYVDVGEEDDEARNL GDEEDSLLEVDSP
Sbjct: 349 PYECSSCGKVFRNCDLRRHALTHAVGEVNSGDYVDVGEEDDEARNLSGDEEDSLLEVDSP 408

Query: 124 RQSPVHNLGESGGSGEKSESERMRLKRKAXIDHEESEEFEFDDFDEEECDLAK 177
RQSPVHNLGESGGSGEKSESERMRLKRKA IDHEESEEFEFDDFDEEEE +DL +
Sbjct: 409 RQSPVHNLGESGGSGEKSESERMRLKRKAIDHEESEEFEFDDFDEEEELQDLPR 462

gi|1480194 (U62004) Sob protein [Drosophila melanogaster]
Length = 577

Score = 435 (200.3 bits), Expect = 1.2e-55, Sum P(2) = 1.2e-55
Identities = 75/84 (89%), Positives = 79/84 (94%)

Query: 4 TRRKPFKCTECGKGFCQSRTLAVHKILHMEESPHKCPVCSRSFNQRSNLKTHLLTHTDHK 63
++ KPFKC ECGKGFCQSRTLAVHKILHMEESPHKCPVC+RSFNQRSNLKTHLLTHTD K
Sbjct: 445 SKEKPFKCAECGKGFCQSRTLAVHKILHMEESPHKCPVCNRSFNQRSNLKTHLLTHTDIK 504

Query: 64 PYECSSCGKVFRNCDLRRHXLTH 87
PY C+SCGKVFRNCDLRRH LTH
Sbjct: 505 PYNCA SCGKVFRNCDLRRHSLTH 528

sp|P23803|ODD_DROME ODD-SKIPPED PROTEIN gi|296793 (X57480) odd gene
product [Drosophila melanogaster]
Length = 392

Score = 263 (121.1 bits), Expect = 1.9e-29, P = 1.9e-29
Identities = 42/63 (66%), Positives = 54/63 (85%)

Query: 4 TRRKPFKCTECGKGFCQSRTLAVHKILHMEESPHKCPVCSRSFNQRSNLKTHLLTHTDHK 63
++ KPFKC++CGKGFCQSRTLAVHK+ H+EE PHKCP+C RSFNQR+NLK+HL +H++
Sbjct: 271 SKDKPFKCDSCGKGFCQSRTLAVHKVTHLEEGPHKCPICQRSFNQRANLKSHLQSHSEQS 330

Query: 64 PYE 66
E
Sbjct: 331 TKE 333

z40

no high scoring matches

z46

		High	Smallest Sum	Probability
Sequences producing High-scoring Segment Pairs:		Score	P(N)	N
pir S54729	RNA-binding protein cabeza - fruit f...	763	2.2e-102	1

pir||S54729 RNA-binding protein cabeza - fruit fly (Drosophila
melanogaster) gi|532788 (U13178) RNA binding protein [Drosophila
melanogaster] gi|567106 (L37083) RNA binding protein [Drosophila
melanogaster]
Length = 404

Score = 763 (353.5 bits), Expect = 2.2e-102, P = 2.2e-102
Identities = 129/129 (100%), Positives = 129/129 (100%)

Query: 1 NVQPRDGDWKCNSCNNTNFAWRNECNRCKTPKGDDEGSSGGGGGGGYGGGGGGGYDRGN 60
NVQPRDGDWKCNSCNNTNFAWRNECNRCKTPKGDDEGSSGGGGGGGYGGGGGGGYDRGN
Sbjct: 276 NVQPRDGDWKCNSCNNTNFAWRNECNRCKTPKGDDEGSSGGGGGGGYGGGGGGGYDRGN 335

Query: 61 DRGSGGGGYHNRDRGGNSQGGGGGGGGGYSRFNDNNGGGRGGGGGNRRDGGPMRN 120
DRGSGGGGYHNRDRGGNSQGGGGGGGGGYSRFNDNNGGGRGGGGGNRRDGGPMRN
Sbjct: 336 DRGSGGGGYHNRDRGGNSQGGGGGGGGGYSRFNDNNGGGRGGGGGNRRDGGPMRN 395

Query: 121 DGGMRSRPY 129
DGGMRSRPY
Sbjct: 396 DGGMRSRPY 404

z60

		High	Smallest
		Score	Sum
Sequences producing High-scoring Segment Pairs:		High	Probability
		Score	P(N)
			N
pir S51247	thioredoxin homolog YDR098c - yeast ...	201	4.1e-27
<p>pir S51247 thioredoxin homolog YDR098c - yeast (Saccharomyces cerevisiae) gi 633632 (Z47746) probable thioredoxin [Saccharomyces cerevisiae] Length = 285</p> <p>Score = 53 (24.1 bits), Expect = 4.1e-27, Sum P(3) = 4.1e-27 Identities = 10/27 (37%), Positives = 13/27 (48%)</p> <p>Query: 19 DKTTVALFAAEWAEQCGQVKDALEELA 45 DK V F WAE C +K E ++</p> <p>Sbjct: 57 DKLIVLYFHTSWAEPCKALKQVFEAIS 83</p> <p>Score = 77 (35.0 bits), Expect = 4.1e-27, Sum P(3) = 4.1e-27 Identities = 18/72 (25%), Positives = 36/72 (50%)</p> <p>Query: 48 TGEKLQFISLNAEQFPEISMKHQIEAVPTVIFFAKGSVDRVDGVDAIAISAKSKKLAEN 107 + + F+S++A++ EIS +I AVP I KG+ + + G D + + +</p> <p>Sbjct: 87 SNSNVSFSLIDADENSEISELFEISAVPYFIIHKGTILKELSGADPKEYVSLLEDCKNS 146</p> <p>Query: 108 ASSAAATGQTLE 119 +S ++ T+E</p> <p>Sbjct: 147 VNSGSSQHTME 158</p> <p>Score = 201 (91.4 bits), Expect = 4.1e-27, Sum P(3) = 4.1e-27 Identities = 39/89 (43%), Positives = 57/89 (64%)</p> <p>Query: 114 TGQTLERLKAINTAPLMIFMKGDRNGPRCGFSKQLIGIVNETNLPYETFDILGDEEVR 173 T + + RL L+N AP+M+FMKG + P+CGFS+QL+GI+ E + + FDIL DE VR</p> <p>Sbjct: 181 TEEQINARLTKLVNAAPVMLFMKGSPSEPKCGFSRQLVGILREHQVRFGFFDILRDESVR 240</p> <p>Query: 174 QGLVKTTPTGHISQVYVKGELIGGSILLR 202 Q L K + Q+Y+ GE GG +++</p> <p>Sbjct: 241 QNLKKFSEWPTFPQLYINGEFQGGLDIIK 269</p>			