MOLECULAR AND CELLULAR BIOLOGY, Nov. 1997, p. 6683–6692 Vol. 17, No. 11
0270-7306/97/$04.00+0

A Domain Shared by the *Polycomb* Group Proteins Scm and ph Mediates Heterotypic and Homotypic Interactions

AIDAN J. PETERSON,[1] MICHAEL KYBA,[2] DOUGLAS BORNEMANN,[1]
KELLY MORGAN,[1] HUGH W. BROCK,[2] AND JEFFREY SIMON[1]*

Department of Biochemistry and Department of Genetics and Cell Biology, University of Minnesota, St. Paul, Minnesota 55108[1] and Department of Zoology, University of British Columbia, Vancouver, British Columbia, V6T 1Z4 Canada[2]

Received 7 May 1997/Returned for modification 11 June 1997/Accepted 21 August 1997

The Sex comb on midleg (Scm) and polyhomeotic (ph) proteins are members of the *Polycomb* group (PcG) of transcriptional repressors. PcG proteins maintain differential patterns of homeotic gene expression during development in *Drosophila* flies. The Scm and ph proteins share a homology domain with 38% identity over a length of 65 amino acids, termed the SPM domain, that is located at their respective C termini. Using the yeast two-hybrid system and in vitro protein-binding assays, we show that the SPM domain mediates direct interaction between Scm and ph. Binding studies with isolated SPM domains from Scm and ph show that the domain is sufficient for these protein interactions. These studies also show that the Scm-ph and Scm-Scm domain interactions are much stronger than the ph-ph domain interaction, indicating that the isolated domain has intrinsic binding specificity determinants. Analysis of site-directed point mutations identifies residues that are important for SPM domain function. These binding properties, predicted α-helical secondary structure, and conservation of hydrophobic residues prompt comparisons of the SPM domain to the helix-loop-helix and leucine zipper domains used for homotypic and heterotypic protein interactions in other transcriptional regulators. In addition to in vitro studies, we show colocalization of the Scm and ph proteins at polytene chromosome sites in vivo. We discuss the possible roles of the SPM domain in the assembly or function of molecular complexes of PcG proteins.

The formation of body structures along the anterior-posterior (A-P) axis in *Drosophila* flies is controlled by the homeotic transcription factors (31, 35). The homeotic proteins are deployed in restricted domains along the embryonic A-P axis that correspond with their realms of function (11, 30, 60). Restricted homeotic expression requires a set of transcriptional repressors, collectively known as the *Polycomb* group (PcG) proteins. Mutations that inactivate individual PcG proteins trigger homeotic gene expression in inappropriate positions along the A-P axis (37, 55, 57).

So far, approximately 15 PcG repressors have been identified in *Drosophila* (see references 44 and 52 for reviews). Several of these PcG proteins localize at specific sites in chromosomes, including the homeotic loci (10, 19, 36, 47). However, little is known about the molecular mechanisms used by PcG proteins for transcriptional repression. Although the molecular cloning and sequence determination of seven of the fly PcG proteins have been completed (7, 9, 13, 24, 28, 36, 43, 49, 53), the predicted primary sequences provide few clues about precise biochemical functions. These proteins lack recognizable DNA-binding domains or catalytic domains. In addition, in vitro tests have failed to identify any single PcG protein that binds to DNA with sequence specificity.

Genetic studies show that at least 12 PcG repressors are each required for homeotic gene repression (37, 55, 56). The need for so many repressors may reflect their coordinated action as components of multimeric protein complexes. In agreement with this idea, the PcG proteins Polycomb (Pc), polyhomeotic (ph), and Polycomb-like (Pcl) show identical distributions at loci on polytene chromosomes, and the Posterior sex combs (Psc) protein distribution is largely overlapping (19, 36, 47). Cytological evidence for PcG protein associations in vivo derives from a heterochromatin-associated chimeric protein, containing the Pc chromodomain, that attracts ectopic accumulation of the Psc and Pc proteins in heterochromatin (45). The association of Psc with its normal sites and the ectopic association of Psc and Pc with heterochromatic sites both require the activity of another PcG protein, the Enhancer of zeste [E(z)] protein (45, 47). Biochemical evidence for PcG protein complexes is provided by coimmunoprecipitation experiments that show association of the Pc and ph proteins in embryonic extracts (19).

Despite accumulating evidence for complexes of PcG proteins, many questions about the biochemical nature of such complexes remain unanswered. Since PcG complexes have not been purified from embryo extracts, their precise native molecular weights and the identity of subunit constituents are not known. Moreover, little is known about how PcG proteins directly contact each other and how specific protein domains are used to assemble or stabilize complexes.

One of the protein domains likely to contribute to PcG functions is present at the C termini of two PcG proteins, Scm and ph (7). This homology domain, termed the SPM domain, is 65 amino acids long, is shared with 38% identity between the two proteins, and is predicted to be largely α-helical. In addition to Scm and ph, two other proteins share this domain with 35 to 65% identity in pairwise comparisons (7). These proteins are a fly tumor suppressor protein, the product of the *lethal(3) malignant brain tumor* gene (61), and a mouse protein, Rae-28, which is likely the mouse homolog of fly ph (1, 41). Besides these four proteins, there are numerous proteins that contain a related domain with much lower overall identity (1, 46). These more distantly related proteins include members of the

* Corresponding author. Mailing address: Department of Biochemistry, University of Minnesota, 1479 Gortner Ave., St. Paul MN 55108. Phone: (612) 624-5361. Fax: (612) 625-5780. E-mail: simon@biosci.cbs.umn.edu.

Ets family of transcription factors (21, 29) and yeast proteins required for mating (46). We will refer to the high-homology domain subgroup that includes the Scm and ph versions as the SPM domain, and we will refer to the extended domain family as the SAM domain (46).

One of the more well-characterized SAM domains is present in the human TEL oncoprotein, an Ets class transcription factor, where it has been referred to as a helix-loop-helix (HLH) domain (23, 51). Recent studies have shown that this domain mediates self-binding and oligomerization of TEL protein and of TEL fusion protein derivatives (21, 29).

If the SPM domain in Scm and ph has similar biochemical properties to the related domain in TEL, then it might be used for binding interactions of these two PcG proteins. In this study, we used the yeast two-hybrid system and in vitro binding assays to test for direct protein-protein interactions between Scm and ph. We found that the SPM domain mediates both self-binding and Scm-ph cross-association and that the isolated domain is sufficient for binding interactions. Deletion mutations were used to assess the SPM domain role in full-length protein interactions, and site-directed point mutations were used to test how individual residues contribute to binding activity. Immunostaining experiments showed colocalization of the Scm and ph proteins on larval polytene chromosomes. Taken together, the in vitro and in vivo data suggest that the Scm and ph proteins participate in PcG complexes as direct binding partners.

MATERIALS AND METHODS

Generation of yeast two-hybrid constructs. Two-hybrid fusion protein constructs were made with the pEG202 bait and pJG4-5 prey vectors (20, 25) as follows. Scm constructs were derived from the Sc9 cDNA (7), and ph constructs were derived from the c4-11 cDNA (13), which corresponds to the proximal of the two ph genomic copies (12).

(i) lexScm767–877. For construct lexScm767–877, a 0.6-kb *Ngo*MI-*Nru*I cDNA fragment was inserted into the *Sma*I site of pBluescript KSII+ (Stratagene). A 0.6-kb *Eco*RI-*Not*I fragment was isolated from this clone and inserted into pEG202.

(ii) ACTScm767–877. For construct ACTScm767–877, a 0.6-kb *Eco*RI-*Xho*I fragment from lexScm767–877 was inserted into pJG4-5.

(iii) lexScm797–877. For construct lexScm797–877, PCR was used to create a 0.5-kb fragment, spanning the SPM domain and part of the adjacent 3′ untranslated region. This fragment, which contains a primer-derived *Eco*RI site preceding codon 797, was digested with *Eco*RI and *Not*I and inserted into pBluescript KSII+ to create pMinSPM. The same *Eco*RI-*Not*I fragment was then inserted into pEG202.

(iv) lexScm1–877. For construct lexScm1–877, PCR was used to create a 0.5-kb fragment spanning the N-terminal Scm coding region. This fragment, which contains a primer-derived *Eco*RI site immediately preceding the start codon, was inserted into pBluescript. The remainder of the Scm coding region was then inserted as a 2.8-kb *Nsi*I-*Sac*I fragment to construct pRS3.1, which contains the complete open reading frame. A 2.9-kb *Eco*RI-*Not*I fragment, derived from this reconstructed clone, was then inserted into pEG202.

(v) ACTScm1–877. For construct ACTScm1–877, a 2.4-kb *Eco*RI-*Cla*I fragment was isolated from pRS3.1 and was inserted into *Eco*RI-*Cla*I-cut ACTScm767–877.

(vi) lexScm1–797. For construct lexScm1–797, PCR was used to create a 150-bp *Bsg*I-*Sac*I fragment spanning Scm amino acids 746 to 797. The PCR introduced a TAA stop codon and *Xho*I and *Sac*I sites immediately after amino acid 797. This fragment was then inserted into *Bsg*I-*Sac*I-cut pRS3.1 to create plasmid pRSΔSPM. A 1.3-kb *Xho*I fragment, spanning amino acids 320 to 797, was isolated from pRSΔSPM and was inserted into *Xho*I-cut lexScm1–877 to create an Scm-LexA fusion protein lacking just the SPM domain.

(vii) ACTScm1–797. For construct ACTScm1–797, the same 1.3-kb *Xho*I fragment described above was inserted into *Xho*I-cut ACTScm1–877.

(viii) ACTph1–1589. For construct ACTph1–1589, PCR was used to create a 250-bp *Eco*RI-*Xho*I fragment spanning the ph N terminus and containing a primer-derived *Eco*RI site just upstream of the start codon. This fragment was then used to replace the N-terminal *Eco*RI-*Xho*I fragment in the c4-11 cDNA. A 4.7-kb *Eco*RI-*Bam*HI fragment spanning ph amino acids 1 to 1586 was isolated from this clone and inserted into pEG202 to create lexph1–1586. A 0.5-kb *Bam*HI fragment containing the C-terminal three ph codons and flanking 3′-UTR sequence was inserted into *Bam*HI-cut lexph1–1586 to make lexph1–1589. A 5.2-kb *Eco*RI fragment containing the complete ph coding region was isolated from this clone and inserted into pJG4-5.

(ix) ACTph1–1418. For ACTph1–1418, lexph1–1586 was cut with *Nco*I, which cleaves at codon 1418 and at a downstream polylinker site. After ligation, which creates a clone with a C-terminally truncated ph fusion protein, a 4.2-kb *Eco*RI-*Xho*I fragment was isolated and inserted into pJG4-5.

(x) ACTph1–522. For ACTph1–522, lexph1–1586 was cut with *Sal*I, which cleaves at codon 522 and at a downstream polylinker site. After ligation, a 1.6-kb *Eco*RI-*Xho*I fragment was isolated and inserted into pJG4-5.

(xi) ACTph1297–1589. For ACTph1297–1589, PCR was used to generate a 0.8-kb fragment with primer-derived *Eco*RI and *Bam*HI sites adjacent to ph codons 1297 and 1577, respectively. This *Eco*RI-*Bam*HI fragment was inserted into pEG202 to create lexph1297–1577. This clone was cut with *Xho*I, and a 1.2-kb *Xho*I fragment from cDNA c4-11, containing the ph C terminus, was inserted to create lexph1297–1589. A 1.3-kb *Eco*RI fragment was then isolated from this clone and inserted into pJG4-5.

PCR amplifications were performed with Vent polymerase (New England Biolabs). All DNA segments included in these constructs that resulted from PCR amplification were sequenced by dideoxy chain termination to verify wild-type Scm and ph sequences.

Yeast two-hybrid tests. Yeast strains for two-hybrid tests were constructed from the base strain EGY48 (*MATα his3 trp1 ura3 6lexAop-LEU2*) (20). Plasmids were introduced into yeast by lithium acetate transformation (27) with 1 to 5 μg of plasmid DNA in 45% polyethylene glycol–0.1 M LiAc–10 mM Tris (pH 8)–1 mM EDTA as the transformation buffer. Two-hybrid tests were performed by patching or streaking strains onto indicator plates supplemented with 2% galactose and 1% raffinose to induce prey fusion protein expression. Activation of *lacZ* from the pSH18-34 reporter plasmid (20) was assayed by scoring blue color on minimal medium lacking histidine, tryptophan, and uracil and containing 40 μg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) per ml. Activation of the chromosomal *LEU2* reporter (20) was assayed by scoring for growth on minimal medium lacking histidine, tryptophan, uracil, and leucine.

Control experiments were performed to test the expression and behavior of the bait fusion proteins. Expression of bait fusion proteins was verified on Western blots of crude yeast extracts prepared by glass bead disruption (14). Western blotting was performed with a rabbit polyclonal anti-LexA antibody (gift from E. Golemis and R. Brent) at a 1:5,000 dilution and goat anti-rabbit horseradish peroxidase (HRP)-conjugated secondary antibody (Bio-Rad) at 1:5,000. Blots were developed with a chemiluminescence-based detection system (ECL kit; Amersham). Each bait fusion protein was detected at approximately the predicted molecular weight. lexScm767–877 appeared reproducibly as a doublet at the expected molecular weight. Nuclear localization of each bait fusion protein was verified by its ability to partially repress *lacZ* expression from pJK101 (20). Each bait protein used failed to activate the pSH18-34 *lacZ* reporter in the absence of a prey protein. Each bait protein also did not repress *lacZ* from the reporter pJK100 (20), which has LexA binding sites upstream of UAS_{GAL}.

Generation of GST fusion constructs. Glutathione-*S*-transferase (GST)-Scm and GST-ph fusion protein constructs were generated with the pGEX series of vectors (Pharmacia) as follows.

(i) GSTScm1–877. For construct GSTScm1–877, a 2.7-kb *Eco*RV-*Nru*I fragment, containing the complete Scm open reading frame, was isolated from pRS3.1 (see above) and inserted into the *Sma*I site of pGEX-3X.

(ii) GSTScm1–797. For construct GSTScm1–797, a 2.4-kb *Eco*RV-*Sac*I fragment was isolated from pRSΔSPM (see above) and was inserted into the *Sma*I site of pGEX-3X.

(iii) GSTScm797–877. For GSTScm797–877, a 0.4-kb *Eco*RV-*Msc*I fragment from pMinSPM (see above) was inserted into the *Sma*I site of pGEX-3X.

(iv) GSTph1511–1576. For GSTph1511–1576, PCR was used to generate a 0.2-kb ph cDNA fragment containing a primer-derived *Eco*RI site immediately upstream of residue 1511 and a primer-derived TGA stop codon and *Xho*I site immediately after residue 1576. The resulting *Eco*RI-*Xho*I fragment was inserted into pGEX-4T-1.

Production of GST fusion proteins. Cultures of *Escherichia coli* DH5α harboring the fusion constructs described above were grown at 37°C in L broth plus 200 μg of ampicillin per ml. When cultures reached an optical density at 550 nm of 1, fusion protein expression was induced by addition of 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside). Unfused GST control protein was induced in cells containing pGEX-2T. After 45 min of further incubation at 37°C, cells were collected by centrifugation, washed once with 10% sucrose–20 mM Tris (pH 8.0)–25 mM EDTA, and resuspended in lysis buffer (10% sucrose, 40 mM Tris [pH 7.5], 0.2 mM EDTA, 2 mM phenylmethylsulfonyl fluoride). The cell suspensions were flash frozen in liquid nitrogen, thawed, and then treated with 0.4 mg of lysozyme per ml for 1 h on ice. The cell lysates were sonicated twice for 30 s and then subjected to an additional freeze-thaw cycle. Insoluble material was pelleted by centrifugation at 16,000 × *g* for 30 min after addition of NaCl to 0.5 M and sodium dodecyl sulfate (SDS) to 0.03%. SDS-polyacrylamide gel electrophoresis (PAGE) analysis of the soluble and insoluble fractions showed that greater than 75% of the smaller fusion proteins (GST, GSTScm797–877, and GSTph1511–1576) was present in the soluble fraction. Approximately 40% of the larger GSTScm1–877 and GSTScm1–797 proteins was distributed in the soluble fraction.

GST fusion proteins were bound to glutathione-agarose beads (Sigma) as

follows. Soluble cell extracts were supplemented with 1% Triton X-100 and 1 mM dithiothreitol (DTT) and then mixed with prehydrated beads for 15 min at 4°C. A 10:1 ratio of extract to slurry of beads was typically used. The beads were washed four times in bead buffer (150 mM NaCl, 20 mM phosphate [pH 7.5], 1% Triton X-100, 1 mM DTT) and then stored in this buffer at 4°C. SDS-PAGE was used to assess the purity and integrity of the fusion proteins attached to the beads. The smaller fusion protein preparations (GST, GSTScm797–877, and GSTph1511–1576) each consisted of one prominent species at the expected molecular weight. Preparations of the larger GSTScm1–877 and GSTScm1–797 proteins contained the expected full-length species and a series of smaller species. The patterns of smaller bands are identical for GSTScm1–877 and GSTScm1–797. For this reason, and since the N-terminal GST moiety is needed for attachment of beads, these smaller species are likely C-terminally truncated degradation products of the Scm fusion proteins.

In experiments that compared the binding properties of wild-type and mutant GSTph1511–1576 proteins (see Fig. 5B), bead preparations with equivalent concentrations of each GST fusion protein were used. Ratios of cell extract to beads in the protein attachment step were adjusted to correct for different yields of these proteins from *E. coli.*

In vitro synthesis of radiolabelled proteins. Radiolabelled proteins were produced in vitro by coupled transcription-translation with the TNT kit (Promega) and [^{35}S]methionine (Amersham). Full-length Scm protein was produced from the cDNA Sc9 (7) with SP6 RNA polymerase. Mutant Scm proteins were produced from Sc9 derivatives carrying site-directed mutations in the SPM domain. Radiolabelled Scm1–797 was produced from the template clone Sc9ΔSPM. This clone was constructed by insertion of a 1.3-kb *Sal*I-*Sac*I fragment from pRSΔSPM (see above), which contains an engineered stop codon after residue 797, into *Sal*I-*Sac*I-cut Sc9. Radiolabelled Scm797–877 was synthesized as a fusion to the HIS_6 peptide tag of pET28-a (Novagen) with T7 RNA polymerase. This template plasmid was generated by insertion of the 0.5-kb *Eco*RI-*Not*I fragment from pMinSPM (see above) into pET28-a. Full-length ph protein was produced from a pBlueScript clone, pT3ph, with T3 polymerase. pT3ph contains the 5.7-kb *Eco*RI fragment from the c4-11 ph cDNA (13). ph1–1417 was synthesized from the clone phT3ΔSPM. This clone was constructed by *Nco*I digestion of pT3ph, generation of blunt ends with Klenow polymerase, and religation. This procedure introduces a stop codon three residues downstream of ph residue 1417. Radiolabelled ph1511–1576 was synthesized as a HIS_6 fusion in pET28-a. This template plasmid was made by insertion of the same *Eco*RI-*Xho*I fragment as in GSTph1511–1576 into pET28-a.

GST pulldown assays. In vitro translation extracts containing Scm or ph radiolabelled proteins were precleared by incubation with GST-bound glutathione-agarose beads before use in binding assays. Five microliters of in vitro translation reaction product was mixed with 250 μl of binding buffer (20 mM Tris [pH 7.5], 200 mM NaCl, 0.1% Triton X-100, 0.25% bovine serum albumin [BSA], 2 mM phenylmethylsulfonyl fluoride). This mixture was incubated with 20 μl of GST-beads in binding buffer for 1 h at 4°C on a rotator, and the unbound radioactive material was collected. Prior to use in the binding reaction, GST-fusion protein beads were blocked in 0.25% BSA for 30 min at 4°C. GST-fusion protein samples contained approximately 1 μg of total protein per μl of beads. Binding reactions were performed by mixing 250 μl of precleared radioactive protein with 20 μl of GST-fusion protein beads, followed by incubation on a rotator for 1 h at 4°C. The pelleted bead samples were then washed five times each with 500 μl of binding buffer. Bound radioactive protein in the final pellets was resuspended in 25 μl of 2× SDS sample buffer. The entire pellet sample (P) and 4% of the unbound material from the first supernatant (S) were analyzed by SDS-PAGE. Radioactive proteins larger than 30 kDa were electrophoresed on 10% SDS gels. Proteins smaller than 10 kDa were run on 14% SDS gels with 2× resolving gel and running gel buffers (42). Gels were fixed for 1 h in 45% methanol–10% acetic acid, rinsed for 30 min in water, and then soaked for 1 h in 1 M sodium salicylate. Gels were dried, and radiolabelled proteins were detected by autoradiography at −70°C.

Generation of site-directed mutations. Site-directed mutations in Scm were generated with the Altered Sites II in vitro mutagenesis system (Promega). A 0.9-kb *Pst*I-*Sac*I cDNA fragment containing the SPM domain coding region was inserted into pALTER-1 (Promega) to produce the mutagenesis template pALTER-SPM. The oligonucleotides used were 5′-TAATAAAGCTTTACTA TCGATTTCGTG-3′ for the G31S mutation, 5′-TCTCTGAATTTAGGGATG ATAAAGCTTTACCA-3′ for the L35S L36S double mutation, and 5′-GGCT GGTCCTAGTGCCAGGCCCATGTA-3′ for the K49A mutation. Clones containing the desired mutations were identified by DNA sequencing. A 0.2-kb *Cla*I-*Nru*I fragment containing the Scm C terminus and adjacent 3′-noncoding DNA was isolated from each of the mutant pALTER-SPM derivatives and inserted into *Cla*I-*Nru*I-cut Sc9 cDNA. This created a full-length mutant Scm cDNA, which was used as a template for in vitro transcription and translation.

Site-directed mutations in ph were made by one- or two-step PCR strategies. These strategies used a wild-type forward end primer, 5′-GGCGGAATTCAGC AGCTGGAGTGTGGAC-3′, containing an *Eco*RI site adjacent to ph codon 1511 and a wild-type reverse end primer, 5′-CCGCCTCGAGTCACTCCTTAA TGGACTC-3′, containing ph codon 1576 followed by a TGA stop codon and an *Xho*I site. The W1A mutation was constructed by one-step PCR with the forward end primer 5′-CCGCGAATTCAGCAGCGCGAGTGTGGACGATGTC-3′ and the wild-type reverse end primer. The I63D mutation was made by one-step PCR with the wild-type forward end primer and the reverse end primer 5′-CG CCCTCGAGTCACTTATCGGACTCCACCTTGGC-3′. The L34A, L42A, and G51A mutations were made with overlapping forward and reverse primers containing the mutations. Each mutant primer was first used in a PCR with the appropriate wild-type forward or reverse end primer to generate two PCR products that overlap at the site of the mutation. Pairs of products were gel purified, combined, and used as a template in a second PCR with the wild-type forward and reverse end primers. The mutagenic primers used to make these three mutations are as follows: L34A, 5′-GGCCAAGCGGCTCTGTTGCTCA AGGAG-3′ and 5′-GAGCAACAGAGCCGCTTGGCCGTCGAT-3′; L42A, 5′-CTCAAGGAGAAGCATGCGGTGAACGCTATGGGC-3′ and 5′-GCCCAT AGCGTTCACCGCATGCTTCTCCTTGAG-3′; and G51A, 5′-GGCATGAAG CTGGCTCCAGCTCTTAAAATT-3′ and 5′-AATTTTAAGAGCTGGAGCC AGCTTCATGCC-3′. The mutant 0.2-kb PCR products were each inserted as *Eco*RI-*Xho*I fragments into pGEX-4T-1 for use in binding assays.

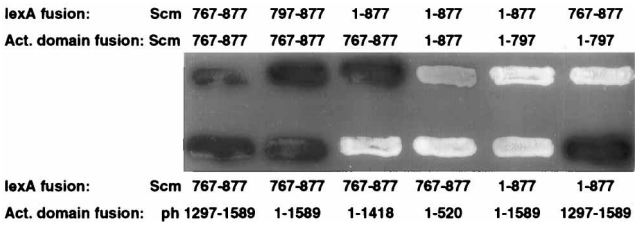

FIG. 1. Yeast two-hybrid tests for Scm and ph protein interactions. Yeast strains containing the indicated LexA and activation (Act.) domain fusion constructs are shown after growth for 48 h on X-Gal indicator medium. The top row shows tests for Scm-Scm self-binding, and the bottom row shows tests for Scm-ph interactions. The full-length proteins are Scm1–877 and ph1–1589. Maps of the constructs are shown in Fig. 2.

Immunostaining of polytene chromosomes. Affinity-purified rabbit polyclonal Scm antibody was generated with a GST fusion protein encompassing Scm amino acids 323 to 877 as an immunogen. The production and characterization of this Scm antibody will be presented elsewhere (7b). The specificity of the antibody for Scm protein is demonstrated by (i) detection of a single, major band of approximately the expected molecular mass (95 kDa) on Western blots of embryo extracts, (ii) elimination of this reacting species in extracts from *Scm* null mutant embryos, and (iii) depletion experiments that show complete loss of chromosome staining after preincubation of the antibody with the Scm immunogen. In parallel depletion experiments, chromosome immunostaining is retained after preincubation with fusion proteins containing the SPM domain from either Scm (GSTScm767–877) or ph (GSTph1511–1576). The affinity-purified rabbit polyclonal ph antibody used for chromosome staining has been described previously (13, 19).

Polytene chromosomes were isolated from third instar larvae prior to spiracle eversion. Immunostaining with Scm antibody was performed with a modified version of protocols described previously (47, 62). Chromosomes were fixed in 45% acetic acid–1% formaldehyde for 2 to 4 min. After chromosome spreading and squashing, the slides were washed in phosphate-buffered saline (PBS; 0.15 M NaCl, 10 mM NaP_i [pH 7.2]) and then incubated in blocking buffer (PBS plus 5% BSA, 5% nonfat dry milk, 0.4% Tween 20) for 30 min at room temperature. Slides were incubated with anti-Scm antibody (1/50) in blocking buffer overnight at 4°C. Signals were developed with the secondary antibody and HRP detection system from the Vectastain ABC kit (Vector Laboratories). The signal was intensified by addition of 0.008% $NiCl_2$ and 0.008% $CoCl_2$ to the HRP reaction mixture. Immunostaining of chromosomes with ph antibody was performed as described previously (13).

RESULTS

Tests for Scm self-binding in the two-hybrid system. A series of two-hybrid fusion protein constructs were generated with the LexA DNA binding domain in the vector pEG202 and the acidic activation domain in the vector pJG4-5 (20, 25). Results of selected pairwise tests on X-Gal indicator plates are shown in Fig. 1, and a compilation of the two-hybrid data is presented in Fig. 2. The coding region for a C-terminal portion of the Scm protein was inserted into the bait and prey vectors to create constructs lexScm767–877 and ACTScm767–877, respectively. This portion of Scm consists of 65 amino acids of the SPM domain plus 45 amino acids of the region immediately upstream. As shown in Fig. 1, a yeast strain harboring both of these plasmids and a *lexAop-lacZ* reporter gene pro-

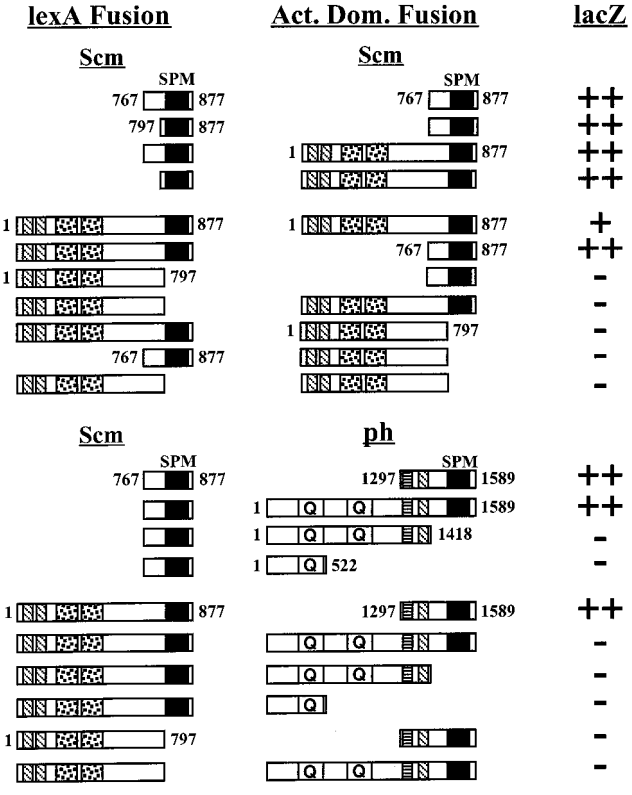

FIG. 2. Summary of Scm and ph two-hybrid constructs and results. Maps of the Scm and ph fusion proteins used in two-hybrid tests are shown. Each line shows a pairwise test performed with the indicated LexA fusion protein and activation (Act.) domain fusion protein. *lacZ* expression was assayed by scoring for blue color on X-Gal media. ++ indicates intense blue signal visible after 16 h of incubation at 30°C, + indicates blue signal visible after 36 to 48 h of incubation, and − indicates absence of blue color after 72 h of incubation. The extents of the Scm and ph coding regions are indicated by numbers on the maps. Full-length proteins are Scm1–877 and ph1–1589. Approximate positions of homology domains and motifs in Scm and ph proteins are indicated by boxes on the maps, which are not drawn precisely to scale. Solid boxes represent the SPM domain, dotted boxes represent mbt repeats (7), diagonally slashed boxes represent Cys_2-Cys_2 zinc fingers (7, 13), horizontally slashed boxes represent the H-I homology region shared by ph and Rae-28 (41), and Q represents glutamine-rich regions.

duces substantial β-galactosidase activity. The accumulation of X-Gal product is easily detectable within 16 h of growth; this signal is at least as strong as activation of the same *lacZ* reporter by a control LexA-GAL4 activator fusion protein. These results indicate that there is self-binding through this portion of Scm protein. The two-hybrid signal intensity suggests a qualitatively strong interaction (16).

The SPM homology domain spans amino acids 806 to 870 in Scm protein (7). To more precisely test whether self-interaction is mediated by the SPM domain, we pared down the LexA fusion protein to contain only Scm amino acids 797 to 877. As shown in Fig. 1, lexScm797–877 also produces a strong positive signal in the two-hybrid test.

Additional constructs were generated to assess the role of the SPM domain in the context of full-length Scm protein (Fig. 2). Figure 1 shows that the isolated SPM domain interacts with full-length Scm protein and that full-length Scm protein also interacts with itself. The signal obtained when full-length Scm is present in both the bait and prey contexts is weaker than when isolated SPM domains are used. This may simply reflect the greater tendency for the smaller fusion proteins to enter yeast nuclei and occupy LexA binding sites (data not shown). Alternatively, the SPM domain may be less accessible for interaction in the full-length protein context.

The lexScm1–797 and ACTScm1–797 fusion proteins contain full-length Scm lacking only the SPM domain. Two-hybrid tests of these truncated proteins against either full-length Scm or isolated SPM domain do not show interaction (Fig. 1 and 2). Taken together, the two-hybrid data show that Scm protein is capable of self-interaction and they suggest that the SPM domain is necessary and sufficient for this interaction.

Tests for Scm-ph interaction in the two-hybrid system. The SPM domain in polyhomeotic protein spans amino acids 1513 to 1576 at the extreme C terminus of the protein (7, 13). To test for cross-interaction between Scm and ph through their C-terminal regions, the construct ACTph1297–1589 (Fig. 2) was generated. Figure 1 shows that there is a strong positive signal when ACTph1297–1589 is tested for interaction with lexScm767–877. In addition, lexScm767–877 interacts with full-length ph protein, and it fails to interact with deletion derivatives of ph protein that lack the C terminus (Fig. 1 and 2). These data, as well as additional tests summarized in Fig. 2, are consistent with physical interaction mediated through the respective SPM domains of the Scm and ph proteins.

In contrast to the strong interactions observed with many Scm-ph combinations (Fig. 1 and 2), we found that there was little or no signal when the two full-length proteins were tested for interaction on X-Gal indicator plates. This result could not be easily explained by failure of production or nuclear entry of either full-length protein, since each showed a strong interaction when tested against smaller SPM-containing fusion proteins (Fig. 2). We considered whether this result might reflect the inability of an Scm-ph multimer, with a minimum molecular mass of about 300 kDa, to specifically activate the *lacZ* reporter used in these assays. Examples of promoter bias exhibited by particular bait-prey combinations in the two-hybrid system have been reported (16). To address this possibility, we tested the full-length lexScm1–877 and ACTph1–1589 fusion constructs for activation of a *lexAop-LEU2* reporter. This *LEU2* reporter contains a different promoter and it is integrated into the yeast chromosome rather than present on a high-copy plasmid (20). We found that this full-length protein combination does activate the *LEU2* reporter as measured by growth on LEU^- media (data not shown). In addition, the reciprocal full-length bait-prey combination, with lexph1–1589 and ACTScm1–877, also yielded growth on LEU^- media. Growth was galactose dependent in these tests, indicating that it required expression of the respective prey proteins. We suggest that the full-length Scm and ph proteins do interact in yeast but that the resulting fusion protein complex is not configured for efficient activation of the *lacZ* reporter. The full-length Scm-ph protein interaction was further tested with an in vitro protein binding assay (see below).

In addition to ph interaction, Scm protein was tested for possible interaction with several other PcG proteins in the two-hybrid system. We found that both full-length and C-terminal Scm fusion proteins fail to interact with the Pc, esc, Psc, or Su(z)2 proteins (data not shown).

Scm and ph protein interactions in vitro. The two-hybrid data provide evidence for specific interactions between domains of the Scm and ph proteins. However, we cannot exclude the possibility that endogenous yeast proteins contribute to the interactions detected in the system. We also wished to perform independent tests of the full-length protein interactions, since the two-hybrid signals with these larger proteins were among the weakest detected. Consequently, we tested for direct interactions by using an in vitro protein binding assay: GST pull-

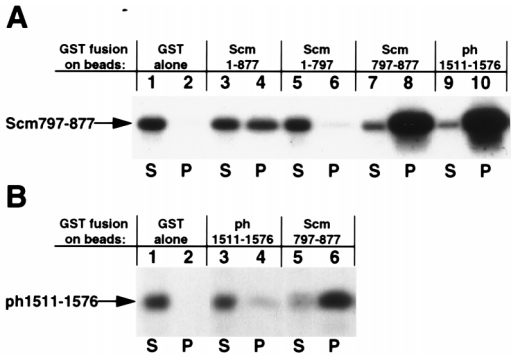

FIG. 3. In vitro binding with minimal domains from the Scm and ph proteins. Autoradiograms of SDS gels are shown. In this and subsequent figures, lanes labelled S contain 4% of the first supernatant removed after the binding reaction, and lanes labelled P contain the bound radiolabelled protein after extensive washing (see Materials and Methods). (A) Radiolabelled Scm797–877 protein was tested for binding to the indicated GST fusion proteins. Scm1–877 is full-length protein, Scm1–797 lacks only the SPM domain, and Scm797–877 and ph1511–1576 are isolated SPM domains. (B) Radiolabelled ph1511–1576 protein was tested for binding to the indicated GST fusion proteins.

down assays (58). Briefly, ^{35}S-radiolabelled proteins were produced by in vitro translation and then tested for binding to GST fusion proteins purified from *E. coli* and immobilized on glutathione-agarose beads (see Materials and Methods).

We first tested the ability of the isolated SPM domain to interact with itself in vitro. The minimal SPM domain from Scm (amino acids 797 to 877) was synthesized in radiolabelled form and tested for binding to GST fusion proteins containing the minimal domain from either Scm or ph (amino acids 1511 to 1576). The pulldown assay data shown in Fig. 3 and subsequent figures consist of results from supernatant (S) samples, which contain a fraction of the unbound material, and pellet (P) samples, which correspond to the bound material. Figure 3A shows that the minimal SPM domain does not bind to the negative control, GST alone (lane 2). However, there is substantial binding to GST fusion proteins containing the minimal domain from either Scm or ph (lanes 8 and 10). Since the GSTScm797–877 and GSTph1511–1576 bead preparations contain equivalent amounts of protein, it appears that the isolated SPM domain from Scm binds about equally well to itself and to the SPM domain from ph. Figure 3A also shows that this isolated SPM domain binds to full-length Scm-GST (lane 4) and that the signal is reduced to background levels if the SPM domain is specifically deleted from the full-length protein (lane 6).

Figure 3B shows the result for self-binding with the isolated SPM domain from ph. We find that there is ph-ph binding activity (lane 4), but that it is dramatically weaker than the Scm-Scm or Scm-ph isolated domain interaction (Fig. 3A). As a further comparison, the same radiolabelled ph1511–1576 protein preparation used in self-binding (Fig. 3B) was tested for cross-binding to GSTScm797–877. This test, shown in Fig. 3B, lane 6, is similar to the test shown in Fig. 3A, lane 10, except the radiolabelled and GST forms of the ph and Scm partners have been reversed. The side-by-side comparison (Fig. 3B, lanes 4 and 6) confirms the dramatic difference in strengths of the ph-ph and Scm-ph domain interactions. This comparison also shows that the relatively weak ph-ph self-interaction is not due to some technical problem with the folding or behavior of the ph fusion proteins used. Rather, we suggest that the qualitative difference in ph-ph and Scm-ph binding affinities reflects intrinsic properties of the different versions of the SPM domain.

We next tested for binding interactions using the Scm and ph proteins in their respective full-length forms. Figure 4A shows that radiolabelled Scm1–877 binds to GSTScm1–877 and to the minimal SPM domain in GSTScm797–877, but not to GST alone. Similarly, radiolabelled ph1–1589 binds to full-length Scm attached to beads (Fig. 4B, lane 4) and to the minimal domain (lane 6). Thus, although the Scm-ph full-length protein interaction was not detected with the *lacZ* reporter in the two-hybrid system (Fig. 2), these tests confirm that Scm-ph binding does occur between the full-length proteins.

The contribution made by the SPM domains to the full-length protein interactions was assessed in the experiment shown in Fig. 4C. In this experiment, deletion derivatives of ph (ph1–1417) and Scm (Scm1–797) that specifically lack the SPM domains were used. In contrast to the binding seen between full-length radiolabelled ph and GST-Scm (lane 2), binding is greatly reduced or eliminated when the SPM domain is specifically removed from either or both proteins tested (lanes 4, 6, and 8). In summary, these results show that the SPM domain is sufficient for binding interactions in vitro and that the domain is required in both partners to mediate the Scm-ph full-length protein interaction.

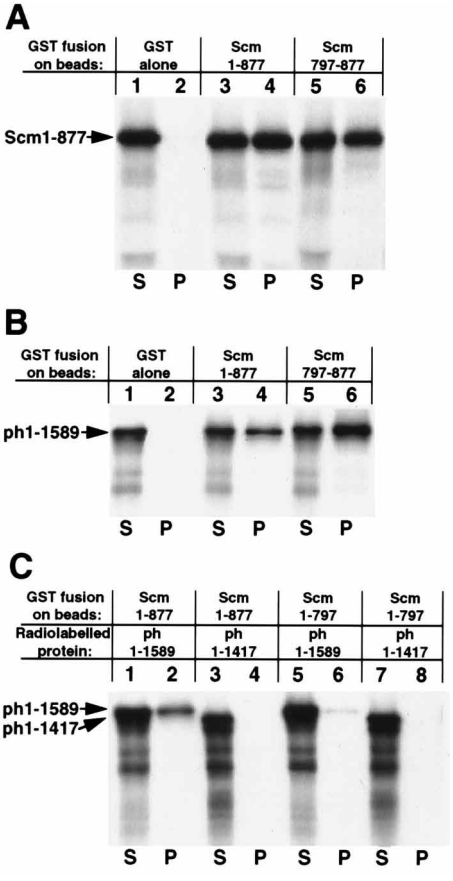

FIG. 4. In vitro binding with full-length Scm and ph proteins and deletion derivatives. (A) Radiolabelled Scm1–877 protein was tested for binding to GST and to the indicated GST-Scm fusion proteins. (B) Radiolabelled ph1–1589 protein was tested for binding to GST and to the indicated GST-Scm fusion proteins. (C) Radiolabelled ph1–1589 and ph1–1417 proteins were tested for binding to the indicated GST-Scm fusion proteins. Comparison of lanes 1 and 3 shows the difference in migration between the ph1–1417 and ph1–1589 proteins (arrows).

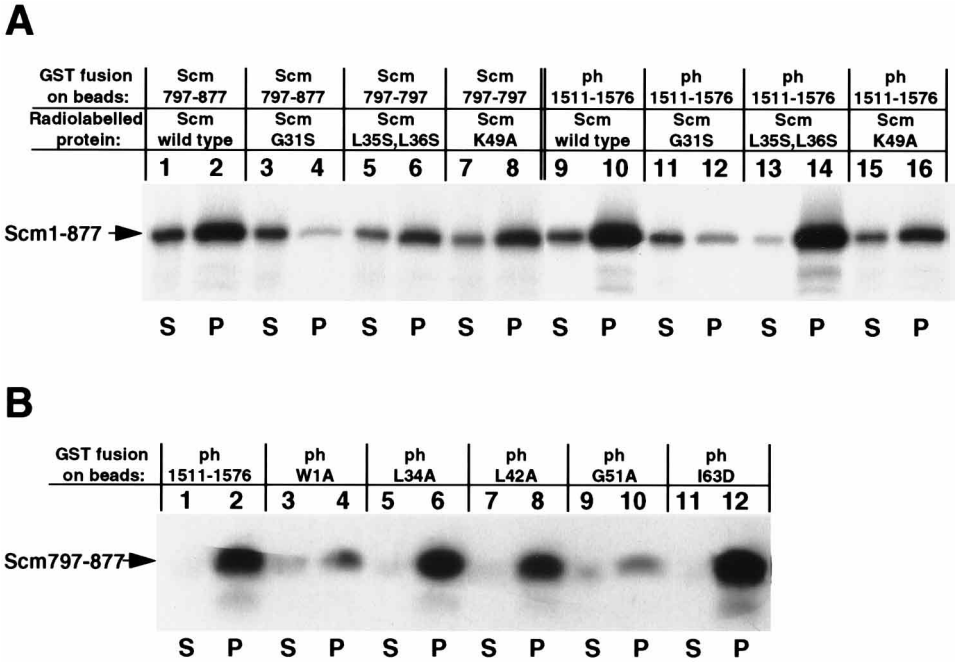

FIG. 5. In vitro binding with Scm and ph mutant proteins. (A) Binding reactions were performed with minimal GST-Scm (lanes 1 to 8) or minimal GST-ph (lanes 9 to 16) fusion proteins. Radiolabelled protein was full-length wild-type Scm or full-length Scm containing the indicated amino acid substitutions. (B) Binding reactions were performed with either wild-type minimal GST-ph fusion protein (lanes 1 and 2) or minimal GST-ph fusion proteins harboring the indicated amino acid substitutions (lanes 3 to 12). Radiolabelled protein was wild-type Scm797–877. The weak signals in the supernatant lanes in panel B reflect the large percentage of radiolabelled protein that binds and is recovered in the pellet sample in the minimal Scm-minimal ph binding reaction. Longer exposure of the gel in panel B (not shown) reveals intact radiolabelled protein in the supernatant lanes as in Fig. 3A, lane 9.

Effects of SPM domain point mutations upon binding in vitro. To begin to investigate the mechanism and specific residues used for SPM domain protein contact, we tested the effects of site-directed mutations in either the Scm or ph domains (see Fig. 6) upon binding in vitro. The mutations were targeted at residues that are highly conserved in alignments of proteins with similar domains. Our point mutations (Fig. 6) fall into two classes: those that target conserved residues in the extended SAM domain family (1, 46) and those that target residues conserved only in the high-homology SPM subgroup.

Three site-directed mutations were generated in the SPM domain of Scm (Fig. 6). The G31S mutation alters a residue that is absolutely conserved in all 23 compiled versions in the extended domain family (1, 46). This mutant was tested in the context of radiolabelled full-length Scm protein for binding to the minimal Scm and ph domains. Figure 5A shows that both Scm-Scm and Scm-ph interactions are greatly reduced in vitro. Consistent with the residual binding activity seen in the G31S lanes, we find that this mutant also mediates a reduced but still detectable interaction in the two-hybrid system (data not shown).

The L35S L36S double mutation and the K49A mutation affect residues conserved in the high-homology subgroup but not in the extended domain family. Figure 5A shows that substantial self- and cross-binding activity is retained with these mutant proteins. The only reduction seen with these two mutants is a modest effect of the L35S L36S double substitution upon the Scm-Scm interaction. It was reproducibly seen that this mutant causes a several-fold loss in Scm-Scm binding but retains Scm-ph cross-binding activity comparable to that of the wild type.

Five site-directed mutations were generated in the SPM domain of ph (Fig. 6). All five mutations alter residues that are highly conserved in the extended domain family. These mutations were inserted into the context of the minimal GSTph1511–1576 fusion protein and then tested for binding to the minimal Scm radiolabelled domain. Figure 5B shows that the W1A and G51A ph mutations cause significant reductions in binding activity to Scm. In contrast, Fig. 5B shows that mutations in the conserved hydrophobic residues, L34A, L42A, and I63D, have little effect upon in vitro Scm-ph interaction.

Colocalization of ph and Scm at chromosomal sites. The two-hybrid and GST pulldown assays show that the Scm and ph proteins can bind each other directly and that their respective SPM domains mediate qualitatively strong interactions. However, these experiments do not address whether the Scm and ph proteins are partners at sites of action in vivo. To assess association in vivo, we compared the Scm and ph distributions on wild-type polytene chromosomes. In addition, we tested for colocalization at an engineered chromosomal site containing an isolated segment of homeotic gene regulatory DNA.

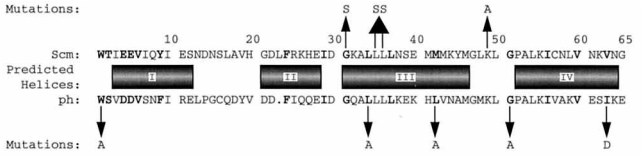

FIG. 6. Mutations in the SPM domains. Alignments of the primary amino acid sequences of the SPM domains from Scm (top) and ph (bottom) are shown. Arrows indicate site-directed mutations analyzed in this work. Scm mutations are shown above the sequence, and ph mutations are indicated below the sequence. Boldface letters indicate amino acid positions that are highly conserved in the extended SAM domain family (1, 46). The positions and extents of potential α-helices in Scm derived from secondary structure prediction with the PHDsec program (48) are shown.

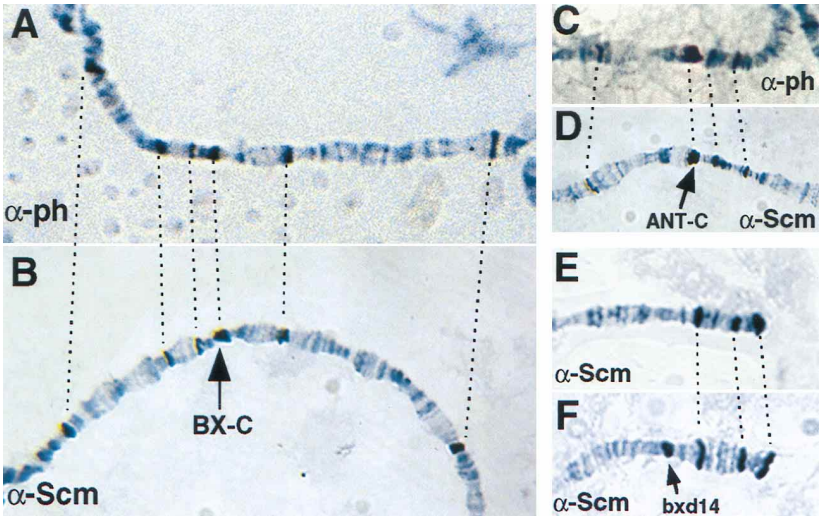

FIG. 7. ph and Scm protein distributions on polytene chromosomes. Chromosomes were immunostained with antibodies specific for the ph or Scm proteins, as indicated. (A) Distribution of ph protein at sites on a central portion of chromosome 3R. (B) Distribution of Scm protein on the same section of chromosome as in panel A. Dashed lines indicate correspondence of immunostaining sites, including the bithorax complex locus (BX-C; arrow). (C) Distribution of ph protein on a proximal portion of chromosome 3R. (D) Distribution of Scm protein on the same section of chromosome as in panel C. Dashed lines indicate correspondence of sites, including the Antennapedia complex (ANT-C; arrow). (E) Scm protein distribution on the distal tip of chromosome 3L from the wild type. Three immunostaining sites are indicated by dashed lines. (F) Scm protein distribution on the same portion of chromosome as in panel E, from a transformant with *bxd* DNA inserted at the indicated site (arrow). The transformed locus creates a novel site of Scm accumulation.

Polytene chromosome immunostaining experiments have shown that ph protein accumulates at its two most well-characterized target loci, the Antennapedia (ANT-C) and bithorax (BX-C) homeotic gene complexes (13, 19). In addition, ph protein is associated with approximately 100 other sites in the genome. Figure 7A shows immunolocalization of ph protein at the BX-C site as well as at five flanking sites on chromosome 3R. Figure 7B shows the same section of chromosome stained with antibody against Scm protein. There is strong signal at the BX-C locus, and the Scm distribution on flanking sites is identical to the ph distribution. Comparison of Fig. 7C and D shows that the ph and Scm protein distributions in the ANT-C region are also identical. Since the antibodies used in these studies are both rabbit polyclonal antibodies, we did not perform double-staining experiments to determine if all approximately 100 ph and Scm sites are identical. However, comparison of the Scm sites on the five major chromosome arms with the ph sites described previously (13) indicates that there is at least 90% overlap in the distributions of these two proteins on polytene chromosomes.

To compare ph and Scm association with an additional site of action in vivo, we tested for colocalization at a site containing regulatory DNA isolated from a homeotic gene. The germ line transformant, 85-39, contains a 14-kb segment from the *bxd* regulatory region of the BX-C complex (54) inserted near the tip of chromosome 3L at cytological location 62A. Previous work has shown that this transformed DNA segment creates a novel site of ph protein accumulation (13) and that expression programmed by this 14-kb DNA segment is regulated by ph and Scm in vivo (54). Figure 7F shows that Scm protein accumulates at the insertion site of this *bxd* regulatory DNA. Thus, the Scm and ph proteins are both recruited to an engineered chromosomal site containing an in vivo regulatory target. This result, together with the coincidence of the ph and Scm proteins at many wild-type chromosomal sites, provides evidence for association of these proteins in vivo.

DISCUSSION

The SPM domain mediates protein-protein interaction. The shared SPM domains in the Scm and ph proteins (7) comprise a potential common link in the biochemical roles of these two proteins in PcG repression. The present data, from both two-hybrid and in vitro assays, show that the SPM domain is used for protein-protein contact. Our studies with minimal constructs of either Scm or ph, encoding as little as 65 to 70 amino acids, show that the isolated domain is sufficient for protein interaction. Although these binding assays do not allow direct determination of binding constants, the signals observed suggest that the domain mediates a moderate- to high-affinity interaction (16).

The isolated SPM domain from Scm binds both to itself and to the domain from ph (Fig. 3). Thus, this domain is designed both for self-recognition and for heterotypic interaction with a related partner. These properties are reminiscent of the HLH domains in transcription factors (see reference 40 for review), which mediate both self- and cross-interactions among related family members. However, there are functional and predicted structural differences between these two classes of self-binding domains. The HLH proteins typically contain a basic region located adjacent to the HLH region that provides sequence-specific DNA-binding activity. The Scm and ph proteins lack an adjacent basic region, and in vitro tests have so far failed to reveal sequence-specific binding activity for either protein (8a). From a structural standpoint, HLH domains contain two major α-helices separated by a loop (15, 17), whereas secondary structure predictions suggest that the SPM domain and the extended SAM domain family (1, 46) consist of a series of three to five α-helices.

The site-directed mutations shown in Fig. 6 were constructed to identify residues important for the fold or function of the SPM domain. Figure 6 also depicts the locations and extents of four putative α-helices that are derived from secondary structure prediction (48). Although a three-dimen-

sional structure is not yet available, the effects of site-directed mutations can be interpreted in the context of a predicted multihelical bundle with the locations and extents of helices as shown. Severe reductions in binding activity are seen with two glycine substitutions, G31S in Scm and G51A in ph. Given their location bordering predicted helices, these G residues may play structural roles in promoting tight turns that connect these helices. Alternatively, they may lie in regions where the absence of side chains is needed to prevent steric conflicts. The absolute conservation of the G31 residue in SAM domains from functionally diverse proteins (1, 46) is consistent with a role in the basic fold of the domain. We also found that mutation of another absolutely conserved residue, W1, causes reduced binding in vitro (Fig. 5). This hydrophobic residue lies adjacent to the first predicted helix. In contrast, mutations in conserved, hydrophobic residues within the predicted α-helices (for example, the leucine residues in helix III) have little effect upon binding. A possible explanation for this result is that multiple hydrophobic interactions contribute to the fold or are used directly for partner contact and that elimination of single interactions in the context of many has little consequence. We note that none of our single residue mutations completely eliminates the strong Scm-Scm or Scm-ph domain interactions. This may also reflect the use of multiple contacts in the binding mechanism.

Biological role of related domains in Ets oncoproteins. Members of the Ets family of transcription factors, which include the human proteins TEL, ETS-1, and ERG-2 (see reference 59 for review) and the fly proteins YAN and PNT (32, 34), contain the SAM homology domain in their respective N-terminal regions. The homology domain in these proteins has also been variously termed the B domain (8), the pointed domain (32), or the HLH domain (23, 51). The most detailed information about this domain in Ets proteins has been derived from studies of the TEL oncoprotein.

The TEL gene was identified by its association with chromosome translocation breakpoints that lead to leukemia in humans (22, 23). A common property of these translocations is that they create novel fusion proteins that contain the N-terminal portion of TEL, including the SAM domain. Fusions of TEL to either of two tyrosine kinases, the platelet-derived growth factor receptor β or the ABL tyrosine kinase, have been described previously (21, 23). Studies of the TEL–platelet-derived growth factor receptor β and TEL-ABL fusions have shown that a key step in transformation is activation of the respective kinase activities by oligomerization of the fusion proteins (21, 29). These workers have shown that the self-binding is mediated specifically by the SAM homology domain from TEL. A third type of leukemia involves TEL fusion to the AML1 transcription factor (22). In this case, TEL-AML fusion converts the transcriptional activator into a repressor (26). This change requires the SAM domain and correlates with oligomerization of the TEL-AML1 fusion protein (38). Thus, the common link among the TEL fusion proteins is the SAM domain, which by triggering oligomerization alters the biochemical activities of the proteins and contributes to the disease state.

Differential binding specificities of the Scm and ph domains. Binding assays with isolated SPM domains show that the Scm-Scm and Scm-ph interactions are much stronger than the ph-ph interaction (Fig. 3). Although we have not precisely measured the difference in affinities, the fractions of radiolabelled proteins bound in these experiments suggest that the difference is at least 10-fold. Since these experiments use minimal domain fusions, and since the same minimal ph1511–1576 protein binds avidly to the domain from Scm, but weakly to itself, we conclude that the binding specificity is an intrinsic property of the domain. We suggest that specific amino acid residues that differ between the Scm and ph domains impart the interaction specificities. Further mutational studies will be needed to identify these specificity residues.

Studies of the SAM domains from TEL and related Ets-proteins also provide information about interaction specificities. Although the TEL domain mediates self-interaction, the comparable homology domains from ETS-1, ERG-2, and GABPα lack self-binding activity in vitro (29). In addition, the ETS-1 protein has been reported to be monomeric in solution (29). Our Scm-ph binding studies establish that a single version of the domain can participate in both homotypic and heterotypic interactions. Taken together, these results show that specificity is built into different versions of the SAM domain so that only certain combinations can form oligomers.

Several families of transcription factors that share a common self-binding domain, such as the basic HLH leucine zipper (bHLH-LZ) domain (3, 6 [see reference 2 for review]) or the POZ (also known as BTB) domain (5), have been described. In these cases, differential binding between specific family members can have important consequences for regulation in vivo. For example, within the bHLH-LZ subfamily that includes Myc, Max, Mad, and Mxi-1, the Myc/Max heterodimer functions as a transcriptional activator (3), whereas Mad/Max and Mxi-1/Max dimers contribute to repression (4, 50). Thus, the distribution of Max into different heterodimeric forms serves as a control point for the biological effects of this group of proteins. It will be important to identify forms of Scm-ph oligomers that occur in vivo and to determine how SPM domain binding specificity influences formation or function of these oligomers. We note that the number of possible oligomeric forms is increased by the presence of distinct ph proteins encoded by the two gene copies in *Drosophila* (12).

ph and Scm function and PcG multimeric complexes. Cytological and biochemical studies suggest that the fly ph protein works together in multimeric protein complexes with other PcG proteins (19, 36, 47). Recently, biochemical studies have also shown association of a mouse ph protein homolog with other PcG protein homologs in embryonic extracts (1). These ph data, together with our ph-Scm binding studies, suggest that direct contact between the ph and Scm proteins contributes to the formation and/or function of PcG protein complexes. In addition, the colocalization of Scm and ph on polytene chromosomes (Fig. 7) provides evidence for association of these proteins at sites of action in vivo.

The ability of the SPM domain to mediate both homo- and hetero-oligomerization of these proteins could supply a variety of contacts used in assembly or regulation of PcG complexes. The specific removal of this domain in the cases of *Scm* mutant alleles *XF24* (7) and *D1* (7a) is consistent with an important role in vivo. The evolutionary conservation of the ph SPM domain in mice (41) and two-hybrid studies that implicate this domain in self-association of mouse ph protein (1) also support an important functional role for the domain. Recently, a mouse Scm homolog that contains an SPM domain with 63% identity to its fly counterpart has been identified (46a). It will be interesting to test whether the mouse ph and Scm homologs also show direct interactions.

Several of the other highly conserved domains in PcG proteins have also been implicated in protein-protein contacts that contribute to complex formation (see reference 52 for review). For example, the chromodomain from Pc protein is required for its own distribution in chromatin (39) and for the distribution of ph protein (18). In addition, the Pc chromodomain is sufficient to target other PcG proteins to ectopic chromatin

locations (45). The identity of direct chromodomain binding partners remains to be determined.

Although there is substantial evidence for associations of PcG proteins, much about the biochemical nature of multimeric moieties remains unclear. Since discrete PcG protein complexes have yet to be purified from embryos, the subunit constituents and their stoichiometries are not known. Further biochemical studies will be needed to assess if and how Scm-ph oligomers collaborate with other PcG proteins in complexes. Additional conserved domains in both the Scm and ph proteins (1, 7, 41) provide potential sites for interactions with other PcG proteins. Indeed, two-hybrid studies implicate a mouse ph domain, distinct from the SPM domain, in contact with the Bmi-1 PcG protein (1), and similar interactions occur between their fly homologs, ph and Psc (33). The remarkable conservation of PcG proteins shows that many of the components that control *Hox* gene expression have been maintained during evolution. Further studies of the conserved building blocks, such as the SPM domain, will determine if molecular mechanisms of *Hox* transcriptional repression are similar in organisms as diverse as flies and mice.

ACKNOWLEDGMENTS

The first three authors contributed equally to this work.

We thank Ellen Miller for generating several of the constructs and yeast strains used in this work. We thank Erica Golemis and Roger Brent for providing LexA antibodies. We thank Fil Randazzo for communicating results prior to publication and Vivian Bardwell and Mike O'Connor for comments on the manuscript.

This work was supported by a grant from the Medical Research Council of Canada to H.W.B. and by NSF grant IBN-9304936 and, subsequently, NIH grant GM49850 to J.S. D.B. was supported in part by NIH training grant HD07480.

REFERENCES

1. **Alkema, M. J., M. Bronk, E. Verhoeven, A. Otte, L. J. van't Veer, A. Berns, and M. van Lohuizen.** 1997. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian Polycomb complex. Genes Dev. **11:**226–240.
2. **Amati, B., and H. Land.** 1994. Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. Curr. Opin. Genet. Dev. **4:**102–108.
3. **Amati, B., S. Dalton, M. W. Brooks, T. D. Littlewood, G. I. Evan, and H. Land.** 1992. Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. Nature **359:**423–426.
4. **Ayer, D. E., Q. A. Lawrence, and R. N. Eisenman.** 1995. Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. Cell **80:**767–776.
5. **Bardwell, V. J., and R. Treisman.** 1994. The POZ domain: a conserved protein-protein interaction motif. Genes Dev. **8:**1664–1677.
6. **Blackwood, E. M., and R. N. Eisenman.** 1991. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science **251:**1211–1217.
7. **Bornemann, D., E. Miller, and J. Simon.** 1996. The *Drosophila Polycomb* group gene *Sex comb on midleg (Scm)* encodes a zinc finger protein with similarity to polyhomeotic protein. Development **122:**1621–1630.

7a. **Bornemann, D., E. Miller, and J. Simon.** Unpublished data.

7b. **Bornemann, D., and J. Simon.** Unpublished data.

8. **Boulukos, K., P. Pognonec, B. Rabault, A. Begue, and J. Ghysdael.** 1989. Definition of an Ets1 protein domain required for nuclear localization in cells and DNA-binding activity in vitro. Mol. Cell. Biol. **9:**5718–5721.

8a. **Brock, H. W., D. Bornemann, and J. Simon.** Unpublished results.

9. **Brunk, B. P., E. C. Martin, and P. N. Adler.** 1991. *Drosophila* genes *Posterior sex combs* and *Suppressor two of zeste* encode proteins with homology to the murine *bmi-1* oncogene. Nature **353:**351–353.
10. **Carrington, E. A., and R. S. Jones.** 1996. The *Drosophila Enhancer of zeste* gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution. Development **122:**4073–4083.
11. **Celniker, S. E., D. J. Keelan, and E. B. Lewis.** 1989. The molecular genetics of the bithorax complex of *Drosophila*: characterization of the products of the *Abdominal-B* domain. Genes Dev. **3:**1424–1436.
12. **Deatrick, J., M. Daly, N. B. Randsholt, and H. W. Brock.** 1991. The complex genetic locus *polyhomeotic* in *Drosophila melanogaster* potentially encodes two homologous zinc-finger proteins. Gene **105:**185–195.
13. **DeCamillis, M., N. Cheng, D. Pierre, and H. W. Brock.** 1992. The *polyhomeotic* gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome-binding sites with *Polycomb*. Genes Dev. **6:**223–232.
14. **Dunn, B., and C. R. Wobbe.** 1994. Preparation of protein extracts from yeast, p. 13.13.4–13.13.5. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. G. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley and Sons, Inc., New York, N.Y.
15. **Ellenberger, T., D. Fass, M. Arnaud, and S. C. Harrison.** 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev. **8:**970–980.
16. **Estojak, J., R. Brent, and E. A. Golemis.** 1995. Correlation of two-hybrid affinity data with in vitro measurements. Mol. Cell. Biol. **15:**5820–5829.
17. **Ferre-D'Amare, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley.** 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature **363:**38–45.
18. **Franke, A., S. Messmer, and R. Paro.** 1995. Mapping functional domains of the Polycomb protein of *Drosophila melanogaster*. Chromosome Res. **3:**351–360.
19. **Franke, A., M. DeCamillis, D. Zink, N. Cheng, H. W. Brock, and R. Paro.** 1992. *Polycomb* and *polyhomeotic* are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. EMBO J. **11:**2941–2950.
20. **Golemis, E. A., J. Gyuris, and R. Brent.** 1994. Interaction trap/two-hybrid system to identify interacting proteins, p. 13.14.1–13.14.17. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. G. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley and Sons, Inc., New York, N.Y.
21. **Golub, T. R., A. Goga, G. F. Barker, D. E. H. Afar, J. McLaughlin, S. K. Bohlander, J. D. Rowley, O. N. Witte, and D. G. Gilliland.** 1996. Oligomerization of the ABL tyrosine kinase by the Ets protein TEL in human leukemia. Mol. Cell. Biol. **16:**4107–4116.
22. **Golub, T. R., G. F. Barker, S. K. Bohlander, S. W. Hiebert, D. C. Ward, P. Bray-Ward, E. Morgan, S. C. Raimondi, J. D. Rowley, and D. G. Gilliland.** 1995. Fusion of the *tel* gene on 12p13 to the *AML1* gene on 21q22 in acute lymphoblastic leukemia. Proc. Natl. Acad. Sci. USA **92:**4917–4921.
23. **Golub, T. R., G. F. Barker, M. Lovett, and D. G. Gilliland.** 1994. Fusion of PDGF receptor β to a novel *ets*-like gene, *tel*, in chronic myelomonocytic leukemia with t[5;12] chromosomal translocation. Cell **77:**307–316.
24. **Gutjahr, T., E. Frei, C. Spicer, S. Baumgartner, R. A. H. White, and M. Noll.** 1995. The Polycomb-group gene, extra sex combs, encodes a nuclear member of the WD-40 repeat family. EMBO J. **14:**4296–4306.
25. **Gyuris, J., E. Golemis, H. Chertkov, and R. Brent.** 1993. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. Cell **75:**791–803.
26. **Hiebert, S. W., W. Sun, J. N. Davis, T. Golub, S. Shurtleff, A. Buijs, J. R. Downing, G. Grosveld, M. F. Roussel, D. G. Gilliland, N. Lenny, and S. Meyers.** 1996. The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. Mol. Cell. Biol. **16:**1349–1355.
27. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153:**163–168.
28. **Jones, R. S., and W. M. Gelbart.** 1993. The *Drosophila* Polycomb-group gene *Enhancer of zeste* contains a region with sequence similarity to *trithorax*. Mol. Cell. Biol. **13:**6357–6366.
29. **Jousset, C., C. Carron, A. Boureux, C. Tran Quang, C. Oury, I. Dusanter-Fourt, M. Charon, J. Levin, O. Bernard, and J. Ghysdael.** 1997. A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerization interface essential to the mitogenic properties of the TEL-PDGFRβ oncoprotein. EMBO J. **16:**69–82.
30. **Karch, F., W. Bender, and B. Weiffenbach.** 1990. *abdA* expression in *Drosophila* embryos. Genes Dev. **4:**1573–1587.
31. **Kaufman, T. C., R. A. Lewis, and B. T. Wakimoto.** 1980. Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homeotic gene complex in polytene chromosome interval 84A-B. Genetics **94:**115–133.
32. **Klambt, C.** 1993. The *Drosophila* gene *pointed* encodes two ETS-like proteins which are involved in the development of the midline glial cells. Development **117:**163–176.
33. **Kyba, M., and H. W. Brock.** The *Drosophila Polycomb* group protein Psc contacts ph and Pc through specific conserved domains. Submitted for publication.
34. **Lai, Z.-C., and G. M. Rubin.** 1992. Negative control of photoreceptor development in *Drosophila* by the product of the *yan* gene, an ETS domain protein. Cell **70:**609–620.
35. **Lewis, E. B.** 1978. A gene complex controlling segmentation in *Drosophila*. Nature **276:**565–570.
36. **Lonie, A., R. D'Andrea, R. Paro, and R. Saint.** 1994. Molecular characterization of the *Polycomblike* gene of *Drosophila melanogaster*, a trans-acting negative regulator of homeotic gene expression. Development **120:**2629–2636.
37. **McKeon, J., and H. W. Brock.** 1991. Interactions of the *Polycomb* group of genes with homeotic loci of *Drosophila*. Roux's Arch. Dev. Biol. **199:**387–396.
38. **McLean, T. W., S. Ringold, D. Neuberg, K. Stegmaier, R. Tantravahi, J. Ritz, H. P. Koeffler, S. Takeuichi, J. W. G. Janssen, T. Seriu, C. R. Bartram, S. E. Sallan, D. G. Gilliland, and T. R. Golub.** 1996. TEL/AML1 dimerizes and is

associated with a favorable outcome in childhood acute lymphoblastic leukemia. Blood **88:**4252–4258.
39. **Messmer, S., A. Franke, and R. Paro.** 1992. Analysis of the functional role of the *Polycomb* chromodomain in *Drosophila melanogaster*. Genes Dev. **6:** 1241–1254.
40. **Murre, C., G. Bain, M. A. van Dijk, I. Engel, B. A. Furnari, M. E. Massari, J. R. Matthews, M. W. Quong, R. R. Rivera, and M. H. Stuiver.** 1994. Structure and function of helix-loop-helix proteins. Biochim. Biophys. Acta **1218:**129–135.
41. **Nomura, M., Y. Takihara, and K. Shimada.** 1994. 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.
42. **Okajima, T., T. Tanabe, and T. Yasuda.** 1993. Nonurea sodium dodecyl sulfate-polyacrylamide gel electrophoresis with high molarity buffers for the separation of proteins and peptides. Anal. Biochem. **211:**293–300.
43. **Paro, R., and D. S. Hogness.** 1991. The *Polycomb* protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. Proc. Natl. Acad. Sci. USA **88:**263–267.
44. **Pirrotta, V.** 1995. Chromatin complexes regulating gene expression in *Drosophila*. Curr. Opin. Genet. Dev. **5:**466–472.
45. **Platero, J. S., E. J. Sharp, P. N. Adler, and J. C. Eissenberg.** 1996. In vivo assay for protein-protein interactions using *Drosophila* chromosomes. Chromosoma **104:**393–404.
46. **Ponting, C. P.** 1995. SAM: a novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. Protein Sci. **4:**1928–1930.
46a. **Randazzo, F. (Chiron Corp.).** Personal communication.
47. **Rastelli, L., C. S. Chan, and V. Pirrotta.** 1993. Related chromosome binding sites for *zeste*, suppressors of *zeste* and *Polycomb* group proteins in *Drosophila* and their dependence on *Enhancer of zeste* function. EMBO J. **12:**1513–1522.
48. **Rost, B., and C. Sander.** 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. **232:**584–599.
49. **Sathe, S. S., and P. J. Harte.** 1995. The *Drosophila extra sex combs* protein contains WD motifs essential for its function as a repressor of homeotic genes. Mech. Dev. **52:**77–87.
50. **Schreiber-Agus, N., L. Chin, K. Chen, R. Torres, G. Rao, P. Guida, A. I. Skoultchi, and R. A. DePinho.** 1995. An amino-terminal domain of Mxi1 mediates anti-myc oncogenic activity and interacts with a homolog of the yeast transcriptional repressor SIN3. Cell **80:**777–786.
51. **Seth, A., and T. S. Papas.** 1990. The *c-ets-1* proto-oncogene has oncogenic activity and is positively autoregulated. Oncogene **5:**1761–1767.
52. **Simon, J.** 1995. Locking in stable states of gene expression: transcriptional control during *Drosophila* development. Curr. Opin. Cell Biol. **7:**376–385.
53. **Simon, J., D. Bornemann, K. Lunde, and C. Schwartz.** 1995. The *extra sex combs* product contains WD40 repeats and its time of action implies a role distinct from other *Polycomb* group products. Mech. Dev. **53:**197–208.
54. **Simon, J., A. Chiang, W. Bender, M. J. Shimell, and M. B. O'Connor.** 1993. Elements of the *Drosophila* bithorax complex that mediate repression by *Polycomb* group products. Dev. Biol. **158:**131–144.
55. **Simon, J., A. Chiang, and W. Bender.** 1992. Ten different *Polycomb* group genes are required for spatial control of the *abdA* and *AbdB* homeotic products. Development **114:**493–505.
56. **Soto, M. C., T.-B. Chou, and W. Bender.** 1995. Comparison of germline mosaics of genes in the Polycomb group of *Drosophila melanogaster*. Genetics **140:**231–243.
57. **Struhl, G., and M. E. Akam.** 1985. Altered distributions of *Ultrabithorax* transcripts in *extra sex combs* mutant embryos of *Drosophila*. EMBO J. **4:**3259–3264.
58. **Tzamarias, D., and K. Struhl.** 1994. Functional dissection of the yeast Cyc8-Tup1 transcriptional co-repressor complex. Nature **369:**758–761.
59. **Wasylyk, B., S. J. Hahn, and A. Giovane.** 1993. The Ets family of transcription factors. Eur. J. Biochem. **211:**7–18.
60. **White, R. A. H., and M. Wilcox.** 1985. Distribution of *Ultrabithorax* proteins in *Drosophila*. EMBO J. **4:**2035–2043.
61. **Wismar, J., T. Loffler, N. Habtermichael, O. Vef, M. Geissen, R. Zirwes, W. Altmeyer, H. Sass, and E. Gateff.** 1995. 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.
62. **Zink, B., and R. Paro.** 1989. *In vivo* binding pattern of a *trans*-regulator of homeotic genes in *Drosophila melanogaster*. Nature **337:**468–471.