Copyright © 1996 by the Genetics Society of America

The Drosophila *ash1* Gene Product, Which Is Localized at Specific Sites on Polytene Chromosomes, Contains a SET Domain and a PHD Finger

Nicholas Tripoulas,[1,3] Dennis LaJeunesse,[2,3] John Gildea and Allen Shearn

Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218

Manuscript received December 27, 1995
Accepted for publication March 14, 1996

ABSTRACT

The determined state of Drosophila imaginal discs depends on stable patterns of homeotic gene expression. The stability of these patterns requires the function of the *ash1* gene, a member of the trithorax group. The primary translation product of the 7.5-kb *ash1* transcript is predicted to be a basic protein of 2144 amino acids. The ASH1 protein contains a SET domain and a PHD finger. Both of these motifs are found in the products of some trithorax group and Polycomb group genes. We have determined the nucleotide sequence alterations in 10 *ash1* mutant alleles and have examined their mutant phenotype. The best candidate for a null allele is $ash1^{22}$. The truncated protein product of this mutant allele is predicted to contain only 47 amino acids. The ASH1 protein is localized on polytene chromosomes of larval salivary glands at >100 sites. The chromosomal localization of ASH1 implies that it functions at the transcriptional level to maintain the expression pattern of homeotic selector genes.

THE determined state of cells within each segment of the newly hatched first instar larva is defined during embryogenesis by the action of segmentation genes along the anterior-posterior axis (reviewed by AKAM 1987). At the molecular level these determined states are equivalent to segment-specific patterns of homeotic selector gene expression. As the imaginal discs within each segment proliferate during larval life they maintain their determined state in the absence of the segmentation gene expression that had initiated the determination process. This maintenance is to some extent mediated by cross regulation (HAFEN *et al.* 1984; HOEY and LEVINE 1988; KRASNOW *et al.* 1989) and auto regulation (BIENZ and TREMML 1988; KUZIORA and MCGINNIS 1988) among homeotic selector genes. There are, however, two other groups of genes that play a key role in maintaining the state of determination of imaginal discs during their proliferative phase. These are the genes of the Polycomb group (reviewed by PARO 1990) and the trithorax group (KENNISON and TAMKUN 1988; SHEARN 1989). Loss of function mutations in genes of the Polycomb group cause ectopic expression of homeotic genes. For example, the *Ultrabithorax* gene is normally expressed in metathoracic imaginal discs but not mesothoracic imaginal discs of third instar larvae (BROWER 1987). Loss of function mutations in genes of the Polycomb group cause ectopic expression of *Ultrabithorax* in the mesothoracic imaginal discs (JONES and GELBART 1990) and lead to transformations of the wing to the haltere. This transformation is similar to transformations caused by gain of function mutations of *Ultrabithorax* such as *Cbx* (LINDSLEY and ZIMM 1992). Loss of function mutations in genes of the trithorax group cause loss of expression of homeotic selector genes (BREEN and HARTE 1991, 1993; TAMKUN *et al.* 1992; SEDKOV *et al.* 1994). For example, mutations in *ash1,* a gene of the trithorax group, cause loss of expression of *Ultrabithorax* in the metathoracic imaginal discs (LAJEUNESSE and SHEARN 1995) and lead to transformation of the haltere to the wing (SHEARN *et al.* 1987; SHEARN 1989). This transformation is similar to the transformation caused by loss of function mutations of *Ultrabithorax* such as *bx* (LINDSLEY and ZIMM 1992).

We have recovered a large number of mutations in the *ash1* gene that is a member of the trithorax group (SHEARN *et al.* 1987; TRIPOULAS *et al.* 1994). Mutations in *ash1* cause transformations of the arista to leg, first leg to second leg, posterior wing to anterior wing, third leg to second leg, haltere to wing, and genitalia to leg and/or antenna (SHEARN *et al.* 1987). The transformation of third leg to second leg and haltere to wing is a consequence of loss of *Ultrabithorax* expression and gain of *Antennapedia* expression (LAJEUNESSE and SHEARN 1995). To begin to understand the molecular mechanism responsible for the role of *ash1* in maintaining the normal expression of homeotic genes, we have isolated the *ash1* gene and identified its product (TRIPOULAS *et al.* 1994). The *ash1* gene product is a 7.5-kb transcript that is present throughout development. It is synthesized in nurse cells during oogenesis and deposited in oocytes. In young embryos the *ash1* transcript is

Corresponding author: Allen Shearn, Department of Biology, The Johns Hopkins University, Baltimore, MD 21218.
E-mail: bio_cals@jhu.edu

[1] *Present address:* Department of Molecular Biology, The Cleveland Clinic Foundation, Cleveland, OH 44195.

[2] *Present address:* LSRC B361, Research Dr., Box 91000, Duke University, Durham, NC 27708.

[3] Both of these authors contributed equally to this work.

distributed uniformly and is presumed to be maternally derived. In larvae the *ash1* transcript accumulates primarily in imaginal discs and is presumed to be zygotically derived (TRIPOULAS *et al.* 1994).

Here we report the recovery of *ash1* cDNAs and present the amino acid sequence predicted by the nucleotide sequence of the entire *ash1* transcript. The predicted ASH1 protein is 2144 amino acids in length. It is proline- and cysteine-rich and very basic; it contains several nuclear localization sequences and three different motifs found in proteins that are localized on chromosomes and/or regulate gene expression. Antibodies raised against an N-terminal part of the ASH1 protein recognize a nuclear protein that is localized at >100 specific bands of salivary gland polytene chromosomes. The mutant phenotype of 10 different alleles was correlated with the nature of the defective protein produced from these alleles as deduced from their altered nucleotide sequence.

MATERIALS AND METHODS

Fly culture and *ash1* stocks: *Drosophila melanogaster* cultures were maintained at 20° or 27° in shell vials on a medium of cornmeal, yeast, molasses and agar with Tegosept added as a mold inhibitor. The *ash1* alleles are marked by *multiple wing hairs* (*mwh*[1]) and *ebony* (*e*[11]) and/or *red Malpighian tubules* (*red*[1]) and maintained over the balancer chromosome *TM1* that is marked with the dominant mutation *Moiré* (*Mé*) or *TM3* that is marked with the dominant mutations *Stubble* (*Sb*[1]) and *Serrate* (*Ser*) now known as *Beaded, Bd*[S]. For full descriptions of balancer chromosomes and deficiencies see LINDSLEY and ZIMM (1992).

Intergenic noncomplementation: Female flies heterozygous for *abx*[1], *brm*[2] or *trx*[e12] were crossed to males heterozygous for *ash1* alleles and the progeny raised at 27°. The doubly heterozygous progeny were identified as those not expressing the dominant markers of the balancer chromosomes. All flies were examined under a dissecting microscope; some were dissected, mounted in Faure's medium (ASHBURNER 1989) and examined under a light microscope. Each third thoracic imaginal disc-derived tissue was scored independently. Homeotic transformations of the third leg were indicated by the presence of a second leg apical bristle, second leg preapical bristle, or sternapleura bristles. Expressivity was classified on a scale of 1–3 indicating the presence of one to three of these features. Haltere transformations were indicated by the presence of anterior wing margin bristles on the anterior capitellum, bristles on the pedicel, mesonotal bristles on the scabellum and/or metanotum, and transformation of the capitellum to wing spread. Expressivity was classified on a scale of 1–4 indicating the presence of one to four of these features.

cDNA library screening: Given the large size of the *ash1* transcript, we considered it likely that most of the transcript would not be represented in oligo-dT-primed cDNA libraries. So, we primarily screened random primed cDNA libraries to obtain cDNAs. Analysis of the temporal pattern of *ash1* transcript accumulation had revealed that it was least abundant in larvae and more abundant in pupae and adults (TRIPOULAS *et al.* 1994). So, we screened one random-primed cDNA library derived from 8-day pupae that was constructed by KENLEY HOOVER in VICTOR CORCES laboratory (HOOVER *et al.* 1993) and one derived from adults that was constructed in lambda ZAP by ELIZABETH MANSFIELD when she was in this laboratory. In larvae, the overall level of transcript is relatively low because it accumulates only in specific tissues such as imaginal discs. So we also screened a library made from imaginal disc poly A RNA by ANDREW COWMAN. The libraries were screened using standard methods with the 5-, 1- and 2.5-kb *Eco*RI genomic fragments of *ash1* as probes (TRIPOULAS *et al.* 1994).

Southern blotting: Genomic DNA from adult flies was isolated using the single fly protocol as described (ASHBURNER 1989) except 200 μl of homogenization buffer and lysis buffer were used to extract DNA from 20 adult flies. The lysis buffer contained 1.25% SDS, and the DNA was precipitated without phenol/chloroform extraction. The DNA from mutant heterozygotes was digested with *Bam*HI, fractionated by agarose electrophoresis and blotted utilizing the protocols for Hybond-N filters (Amersham). The digested DNA was hybridized with both an 8- and a 1.6-kb *Bam*HI fragment, which together encode the *ash1* transcript. The probe DNA was purified by addition of agarase (New England Biolabs) to a diluted agarose solution followed by ethanol precipitation. Radioactive DNA probes were synthesized using the Prime-It kit (Stratagene).

Sequencing the wild-type gene: Double-stranded plasmid DNA was sequenced using standard sequencing reactions with ^{35}S-dATP (DuPont-NEN), dideoxynucleotides and T7 polymerase (Sequenase kit, USB). Primers were synthesized trityl-off and concentrated by ethanol precipitation. Cosmid DNA was sequenced at the Johns Hopkins University Core Facility using the automated sequencing system of Applied Biosciences.

Cloning the breakpoint of an inversion allele: Five larval equivalents of genomic DNA prepared from late third instar wild-type larvae or larvae hemizygous for *ash1*[29] (*ash1*[29]/*Df(3L)]K18*) were digested for 12 hr at 37° with 40 units of *Bam*HI restriction enzyme in 100 μl of TE pH 7.5. The enzyme was heat inactivated for 15 min at 65°, the samples were phenol chloroform extracted and the DNA was ethanol precipitated. The pellet was resuspended in 20 μl $^1/_{10}$ × TE. Two microliters of sample were then ligated in 100 μl volume at 16° for 12 hr using 1× BRL ligation buffer including 1 mM ATP and 4 units T4 DNA ligase. The ligated samples were then phenol-chloroform extracted, ethanol precipitated, and resuspended in 10 μl $^1/_{10}$ × TE. Inverse PCR was performed in a 100 μl volume containing 1× Taq Extension Buffer, 200 μM each dNTPs, 200 nM each inverse PCR primer, 5 units Taq Polymerase, 5 units Taq extender, and 10 μl ligated genomic DNA (0.5 larval equivalents). Hot start was done by adding Taq polymerase and Taq extender at 72°. Cycling was performed in a Perkin Elmer 480 thermal cycler. Cycling condition were as follows: denaturation 1 min 94°, annealing 1 min 63°, extension 4 min 68°, for five cycles, then denaturation 1 min 94°, annealing 1 min 60°, extension 4 min 68°, for 25 more cycles, then 4° soak. Five microliters of inverse PCR product was then run on a 1% agarose gel and stained with ethidium bromide. Primers were chosen so that the product from mutant DNA would contain the new *Bam*HI site. The product from wild-type DNA showed a single band of ~3200 bp, and the product from *ash1*[29] hemizygous DNA showed a single band of ~3500 bp. The product from *ash1*[29] hemizygous DNA was diluted 1:5 with H_2O and 1 μl was cloned using the pCRII TA cloning kit (Invitrogen).

Cycle sequencing of mutant genes: The protocol for cycle sequencing was adapted from the Perkin Elmer AmpliCycle Sequencing Kit. Enough primer for 10 sequencing reactions was 5′ end labeled. The sequencing reaction mixtures contained 24.0 μl H_2O, 1.0 μl of end labeled primer, 4.0 μl 10× cycling mix (from kit) and 1.0 μl template (1 μg/μl of template in water). Six microliters of this mix was dispensed to each termination mix, which was then overlayed with 20 μl

mineral oil. The reaction mix was incubated at 95° for 2 min and then amplified for 30 cycles under these conditions: 95° for 1 min, 52° for 1 min, 72° for 1 min. The mixture was incubated at 4° for <45 min and terminated with the addition of 4 μl stop solution (from kit). Samples were denatured at 95° for 5 min and run on a 6% sequencing gel, 4 μl per lane.

Sequence analysis: DNA sequences were analyzed and conceptually translated using the MacDNAsis program (Hitachi). Protein motif searches were done using MacPattern (FUCHS 1991) and the Prosite data base (BAIROCH 1992). Sequence similarity searches were run on the National Library of Medicine Facility using the BLASTP search algorithm (ALTSCHUL *et al.* 1990). Alignments were performed using the MulAlignment program of the Darwin system, which is available through the Computational Biochemistry Research Group of the Eidgenössische Technische Hochschule in Zurich.

Generation of an antibody directed against an ASH1 fusion protein: We generated polyclonal rabbit antibodies to a fusion protein that contains six consecutive histidines at the amino terminus fused to 326 amino acids of ASH1 from residues 443–769 (Figure 3). The fusion protein was expressed in bacteria and purified from a bacterial extract by binding it to a nickel NTA-agarose column and eluting with citrate buffer according to the manufacturer's protocol. The eluted protein contained some full length fusion protein as well as smaller fragments (data not shown). Since it is only the six consecutive histidines at the amino terminus that are required for the fusion protein to bind to the column, we infer that the smaller fragments we recovered from bacterial extracts represent amino terminal fragments of the fusion protein. The majority of these smaller fragments are not bacterial proteins since they were only detected in extracts of bacteria that were expressing the fusion protein. The entire mixture of eluted proteins was used as an immunogen. The anti-ASH1 antibodies were purified on affinity columns prepared by cross-linking purified fusion protein to Affi-Gel 10 resin (Bio-Rad) in 1 M urea, 10 mM PO_4 buffer at pH 7 for 4 hr at 4°. Bound antibodies were eluted in 0.5% acetic acid, 0.15 M NaCl and immediately neutralized. Bovine serum albumin (BSA) was added to 1% and the purified antibody was stored at −80°. The specificity of the affinity-purified antibodies was tested on immunoblots of purified fusion protein. The antiserum contained antibodies that recognized the full-length fusion protein as well as fusion protein fragments (data not shown). This suggests that the primary epitope(s) recognized are near the amino terminus of the fusion protein. Otherwise the full-length fusion protein would give a stronger response to the antibodies than the fusion protein fragments.

Immunofluorescent detection of ASH1 and PSC on polytene chromosome: Salivary glands from wild-type or mutant larvae grown at 17° were dissected in 0.1% Triton X-100 in phosphate-buffered saline (PBS). The glands were fixed in 3.7% formaldehyde in 1% Triton X-100 for 6 sec and then in 3.7% formaldehyde plus 50% acetic acid for 2–3 min. The glands were squashed to spread polytene chromosomes and blocked in 3% BSA plus 10% nonfat dry milk in PBS. The chromosomes were incubated with primary antibody overnight, washed in PBS, incubated with fluorescein-labeled goat anti-rabbit immunoglobulin secondary antibody and/or rhodamine-labeled guinea pig anti-mouse immunoglobulin secondary antibody for 1 hr, and mounted in Immunomount to which phenylene diamine (DAPI) was added. The fluorescent images were captured with a cooled CCD camera in black and white and colorized using Oncor image software. For double immunofluorescent images, the fluorescein and rhodamine images were collected separately and then merged.

RESULTS

Recovery and mapping of *ash1* cDNAs: The entire *ash1* gene is contained within 10 kb of cloned DNA and encodes a 7.5-kb transcript (TRIPOULAS *et al.* 1994). We have recovered six cDNAs in addition to the one cDNA we previously reported (TRIPOULAS *et al.* 1994). Together these seven cDNAs account for all but a small portion of the *ash1* transcript. The position of five cDNAs with respect to the genomic restriction map is indicated in Figure 1 as is the position of a 5-kb *Eco*RI genomic fragment.

Nucleotide sequence of *ash1* transcript: We have determined the nucleotide sequence of both strands of 8141 bp of DNA from the *ash1* gene (Genbank accession number U49439). The sequence of one strand was determined from the complete sequence of the seven cDNAs and sequence from both ends of a 5-kb *Eco*RI genomic subclone (Figure 1). The sequence of the complementary strand was determined from genomic DNA included in a cosmid that contains the entire *ash1* gene (TRIPOULAS *et al.* 1994). Comparison of the cDNA and genomic sequences revealed the presence of two small introns of 63 and 59 nucleotides (data not shown). The size, composition and sequence ends of these introns fit the consensus for small Drosophila introns (MOUNT *et al.* 1992). The sequence between a putative cap site for the transcript, ATCAAGC, that is identical in five of seven positions with the consensus cap site, ATCA(G/T)A(C/T) for all Drosophila nonheat shock genes (HULTMARK *et al.* 1986) and a putative poly-adenylation signal, AATAAA, predicts a spliced transcript size before polyadenylation of 7.4 kb that is in good agreement with the observed transcript size of 7.5 kb.

Sequence of the predicted ASH1 protein: There is a single long open reading frame in the nucleotide sequence determined from *ash1* cDNAs. Neither the first nor the second ATG in the open reading frame are in a context that is favorable for translation initiation in Drosophila. However the third ATG is in a context (AAAAATG) that is identical to the consensus for translation initiation [(A/C)AAAATG] (CAVENER 1987). So, assuming that the third ATG is used for initiation of translation, the protein predicted by this sequence would be 2144 amino acids in length (Figure 2). It would be a highly basic protein with a predicted isoelectric point of pH 10 and would be cysteine (2.1%)- and proline (7.2%)-rich. Within the putative ASH1 protein there are seven regions that contain stretches of basic residues that are putative nuclear localization sequences (DINGWALL and LASKEY 1991). These include two nine-amino acid stretches with seven of nine basic residues (lysine or arginine), one seven-amino acid stretch with six of seven basic residues, two six-amino acid stretches with six of six basic residues, and 2 six-amino acid stretches with five of six basic residues (Figure 2). There are seven PEST sequences within the putative ASH1 protein (Figure 2). PEST sequences are regions with an unusually high fraction of proline, glutamic acid, aspartic acid, serine and threonine residues that is characteristic of short-lived proteins (RECHSTEINER 1990). According to the PESTFIND program

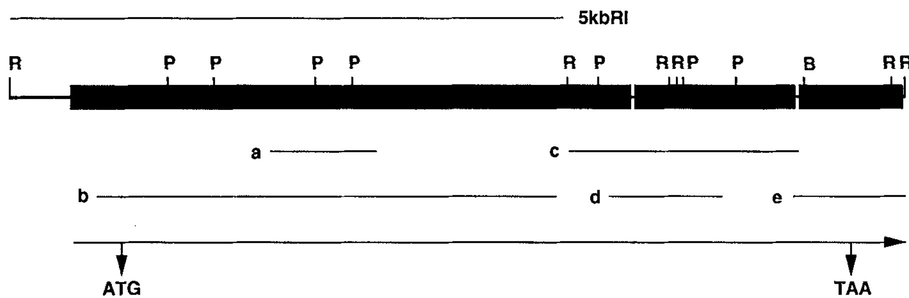

FIGURE 1.—Location of *ash1* cDNAs on restriction map of *ash1* gene. The 7.5-kb *ash1* transcript is detected by three genomic *Eco*RI fragments of 5, 1, and 2.5 kb. Five cDNAs were recovered from cDNA libraries using these fragments as probes. The location of these cDNAs (labeled a–e) was initially determined by restriction mapping and ultimately determined by comparing their sequences to the genomic sequence. The 5′ end of the transcript and a small region of the open reading frame (between the 3′ end of cDNA b and the 5′ end of cDNA c) was determined using the ends of the 5-kb *Eco*RI genomic subclone. The transcribed region is indicated by the broad band on the map. Two interruptions of this broad band indicate the positions of introns. The direction of transcription of the *ash1* mRNA is indicated by the line with the arrow. The position of the start (ATG) and stop (TAA) codons is indicated. E, *Eco*RI; P, *Pst*I; B, *Bam*HI.

(ROGERS *et al.* 1986) the seven significantly high scoring regions are residues 187–226 (score = 7.5), residues 652–664 (score = 16.2), residues 907–922 (score = 11.4), residues 1021–1037 (score = 9.3), residues 1079–1107 (score = 6.5), residues 1125–1142 (score = 9.8), and residues 1727–1738 (score = 7.6).

There are three motifs in the ASH1 protein that are characteristic of some proteins that regulate transcription and/or are bound to chromosomes. At residues 177–186 (Figure 2) there is a $^{9}/_{10}$ residue match with the AT hook motif [(A/T) X (R/K) (R/K) X R G R P (R/K)], that was defined in high mobility group proteins as important for binding to the minor groove of AT-rich regions of double-stranded DNA (REEVES and NISSEN 1990). At residues 977–986 (Figure 2) there is a less well conserved $^{7}/_{10}$ residue match with the AT hook motif. For comparison, there are three AT hook motifs in HRX (DJABALI *et al.* 1992; GU *et al.* 1992; TKACHUK *et al.* 1992), the human homologue of TRX, the product of the Drosophila *trithorax* gene. One of the three (residues 171–180) is as well conserved as the better conserved one in ASH1 ($^{9}/_{10}$ residues). The other two (residues 217–226 and 298–307) are as well conserved as the less conserved one in ASH1 ($^{7}/_{10}$ residues).

Near the middle of the ASH1 protein there is a SET domain (Figure 2). This 150-amino acid sequence is similar to sequences found in three other Drosophila proteins (Figure 3A): SUVAR(3)9, the product of the *Suppressor of variegation (3) 9* gene (TSCHIERSCH *et al.* 1994); E(Z)/PCO, the product of the *Enhancer of zeste/polycombeotic gene* (JONES and GELBART 1993); and TRX, the product of the *trithorax* gene (MAZO *et al.* 1990). This domain is also found in proteins from other organisms. Among the 11 sequences aligned in Figure 3A,

```
   1 MSTQDTESGC SEAKNRAVSK KVKVKRKKLA SSSGISKSDK VSKSKKSQIS AFSSDSEDDL    60
  61 PLKVHQQRAP RVLLSAIIQA GQSGSKPTLD IGISSSDNEL PNLVQAAIKR VESDTEDTTY   120
 121 EGSFRKAAKD KNLPQYQSTL LQDFMEKTQM LGQTVNAKLA EEKVAKAKEE TLVQTAVPRK   180
 181 RRGRPKKVVP TVPAPGNSGP AINESADSGV ISTTSTTQST TPSPKMQNEN AVPTGSLTIA   240
 241 SRRKPKIDMA YLDKRMYATE RVLYPPPRSK RRQNNKKTAC SSSNKEELQL DPLWREIDVN   300
 301 KKFRLRSMSV GAASGTGAST TICSKVLAAK SGYVSDYGSV RHQRSSHNHN SGYKSDASCK   360
 361 SRYSTKSCMS RRSRPKSCGY RSDCKESGKS GLRMRRKRRA SMLLKSSADD TVEDQDILQL   420
 421 AGLSLGQSSE ESNEYISKPS LKSLPTTSAS KKYGEINRYV TTGQYFGRGG SLSATNPDNF   480
 481 ISKMMNQRKE TPAPSKSSCK IKSRRSSAAS MCSSYVSGVS RMRRRHRRKS FSHNKSLNID   540
 541 SKLLTEIEII TSTFNSRCRI QDDRLTGSSG KEKLLADANK LQATLAAPSP AQQLTLNGGG   600
 601 PASTLSKPLK RGLKKRKLSE PLVDFAMLSA SASGTPNGSG SSNGNTKRRH KKSQSNDSSS   660
 661 PDDHKLPLKK RHYLLTPGER PPAEVAFANG KLNAEAWAAA AAAAKSTAST KSQAQFNARS   720
 721 VKSALTPKKR HLLEQPTSVS GAGSSASNSP LRIVVDNNSI SGGKLLDISP SSLCSLKQQR   780
 781 RGGAAKQKVS AAKDLVQLQS PAGSYPPPGV FEPSVELEIQ IPLSKLNESV ITKAEVESPL   840
 841 LSALDIKEDT KKEVGQRVVE TLLHKTGGNL LLKRKRKKIN RTGFPTVRRK KRKVSVEQQT   900
 901 TAVIDEHEPE FDPDDEPLQS LRETRSSNNV NVQAAPNPPL DCERVPQAGE ARETFVARTN   960
 961 QKAPRLSVVA LERLQRPQTP ARGRPRGRKP KNREQAEAAP QPPPKSEPEI RPAKKRGRQP  1020
1021 KQPVLEEPPP TPPPQQKKNK MEPNIRLPDG IDPNTNFSCK IRLKRRKNLE AGTQPKKEKP  1080
1081 VQPVTVEEIP PEIPVSQEEI DAEAEAKRVD SIPTEHDPLP ASESHNPGPQ DYASCSESSE  1140
1141 DKASTTSLRK LSKVKKTYLV AGLLSNHYKQ SLMPPPAKVN KKPGLEEQVG PASLLPPPPY  1200
1201 CEKYLRRTEM DFELPYDIWW AYTNSKLPTR NVVPSWNYRK IRTNVYAESV RPNLAGFDHP  1260
1261 TCNCKNQGEK SCLDNCLNRM VYTECSPSNC PAGEKCRNQK IQRHAVAPGV ERFMTADKGW  1320
1321 GVRTKLPIAK GTYILEYVGE VVTEKEFKQR MASIYLNDTH HYCLHLDGGL VIDGQRMGSD  1380
1381 CRFVNHSCEP NCEMQKWSVN GLSRMVLFAK RAIEEGEELT YDYNFSLFNP SEGQPCRCNT  1440
1441 PQCRGVIGGK SQRVKPLPAV EAKPSGEGLS GRNGRQRKQK AKKHAQRQAG KDISSAVAVA  1500
1501 KLQPLSEKEK KLVRQFNTFL VRNFEKIRRC KAKRASDAAA TASSPALGTT NGDIPGRRPS  1560
1561 TPSSPSLAAQ ISALCSPRNI KTRGLTQAVH DPELEKMAKM AVVLRDICSA METLKMSDLL  1620
1621 TTVSSKKKKP IKTTLSGKLG STAATSKVEF RSIQAQVEQG HYKTPQEFDD HMQQLFVEAK  1680
1681 QQHGDDEGKE KALQSLKDSY EQQKIASYVQ LVEILGDSES LQSFKPKEVL SSEEEPGKIA  1740
1741 VKKSPGAKER DSPIVPLKVT PPPLLPIEAS PDEDVIRCIC GLYKDEGLMI QCSKCMVWQH  1800
1801 TECTKADIDA DNYQCERCEP REVDREIPLE EFTEEGHRYY LSLMRGDLQV RQGDAVYVLR  1860
1861 DIPIKDESGK VLPTKKHTYE TIGAIDYQEC DIFRVEHLWK NELGKRFIFG QHFLRPHETF  1920
1921 HEPSRRFYPN EVVRVSLYEV VPIELVIGPC WLLDRPTFSK GPPMECNDED HCYICELRVD  1980
1981 KTARFFSKAK ANHPACTKSY AFRKFPEKIK ISKSYAPHDV DPSLLKTRKQ KTELDVGAGP  2040
2041 TTMHKVSGRQ EQHQAKMVGR KPCGISAPAD ATAVHVVTPV APNKQMHKKR KSRLENVLVT  2100
2101 MKLKCLDAQT AQEQPIDLSY LLSGRGARQR KTQQSSSSST ANST
```

FIGURE 2.—Predicted amino acid sequence of primary *ash1* translation product. □, seven putative nuclear localization sequences; ■ with inverse letters, AT hook motifs. PEST sequences are indicated by wavy underline. SET domain is indicated by single underline; PHD finger is indicated by double underline. Brackets mark 326 amino acids included in a fusion protein that was used to immunize a rabbit. Arrowheads indicate positions of introns.

A

```
ASH1       1300  KIQRHAVAFGVERFMTADKGWGVRTKLPIAKGTYILEYVGEVVTEKEFKQRMASIYLND__THHYCLHLDGGLVIDGQ  1375
Y84.5KD     111  QRFQKKQYAPIAIFKTKHKGYGVRAEQDIEANQFIYEYKGEVIEEMEFRDRLIDYDQRHFK_HFYFMMLQNGEFIDAT   188
K09F5.5      84  RFRKRQFCGVETFLTDNGIGHGLRATEEIATGKLILEYRGEAITKAEHNKRV_KRYKKDGIKHSYSFEVGRNYYVDPT   164
HRX        3818  RHLKKTSKEAVGVYRSPIHGRGLFCKRNIDAGEMVIEYAGNVIRSIQTDKREKYYDSKG__IGCYMFRIDDSEVVDAT  3888
Y123.9KD    931  NQLNKRKK_PVMFARSAIHNWGLYALDSIAAKEMIIEYVGERIRQPVAEMREKRYLKNG_IGSSYLFRVDENTVIDAT  1006
C26E6.10    592  INQLKFRKKMIKFARSRIHGWGLYAMESIAPDEMIVEYIGQTIRSLVAEEREKAYERRG_IGSSYLFRIDLHHVIDAT   670
TRX        3604  RTLKETYKDYVGVFRSHIHGRGLYCTKDIEAGEMVIEYAGELIRSTLTDK_RERYYDSRGIG_CYMFKIDDNLVVDAT  3679
SUVAR(3)9   475  LVQHGRQVFLVLFKTANGSGWGVRAATALRKGEFVCEYIEEIITSDEANERGKAYDDNG___RTYLF__DSEYTIDAA   550
G9A         809  RVVQSGIKVRLQLYRTAKMGWGVRALQTIPQGTFICEYVGELISDAEADVREDDSYLFD___LDNKD__GEVYCIDAR   890
ORFKG1T     786  RLVQHGLQVRLQLFKTQNKGWGIRCLDDIAKGSFVCIYAGKILT/MASVDKGESAPVRKNTRQFYDGE_ESCYIIDAK  1213
E(Z)/PCO    607  CVQRGLHKHL_LMAPSDIAGWGIFLKEGAQKNEFISEYCGEIISQDEADRR_GKVYD_KYMCSFLFN_LNNDFVVDAT   680
                                    *                  *                              •

ASH1       1376  RMGSDCRFVTHSCEPNCEIQKWSVNG___LSRMVLFAKRAIEEGEELTYDYNFSLFN_PSEGQPCRCNTPQCRGVIG  1448
Y84.5KD     189  IKGSLARFCNHSCSPNAYVNKWVVKD___KLRMGIFAQRKILKGEEITFDYNVDRYG__AQAQKCYCEEPNCIGFLG   260
K09F5.5     165  RKGNSARFINHSCNPNALVKVWTVPD__RPMKSLGIFASKVIKPGEEITFDYGTSFRN____DQPCQCGEAACRGWIG   236
HRX        3889  MHGNRARFINHSCEPNCYSRVINIDG___QKHIVIFAMRKIYRGEELTYDYKFPIED_ASNKLPCNCGAKKCRKFLN  3961
Y123.9KD   1007  KKGGIARFINHCCDPNCTAKIIKVGG___RRRIVIYALRDIAASEELTYDYKFEREKDDEERLPCLCGAPNCKGFLN  1080
C26E6.10    671  KRGNFARFINHSCQPNCYAKVLTIEG___EKRIVIYSRTIIKKGEELTYDYKFPIED___DKIDCLCGAKTCRGYLN   739
TRX        3680  MRGNAARFINHCCEPNCYSKVVDILG___HKHIIIFAVRRIVQGEELTYDYKFPF___EDEKIPCSCGSKRCRKYLN  3750
SUVAR(3)9   551  NYGNISHFINHSCDPNLAVFPCWIEHLNVALPHLVFFTLRPIKAGEELSFDYIRADNEDVPYENLSTAVRVECRCGRD   628
G9A         891  YYGNISRFINHLCDPNIIPVRVFMLHQDLRFPRIAFFSSRDIRTGEELGFDYGDRFWDIKSKYFTCQCGSEKCKHSAE   968
ORFKG1T    1214  LEGNLGRYLNHSCSPNLFVQNVFVDTHDLRFPWVAFFASKRIRAGTELTWDYNYEVGSVEGKELLCCCGAIECRGRLL  1291
E(Z)/PCO    681  RKGNKIRFANHSINPNCYAKVMMVTGDH____RIGIFAKRAIQPGEELFFDYRYRYGPTEQLKFVGIEREMEIV      754
                   *         *    **                           *     *     **
```

B

```
TRX-1      CFLCCGSTGLDP.........LIFCAC..CCEPYHQYCVQDEYNLKHGSFEDTTLMGSLLETTVNASTGPSSSLNQLTQRLNWLCPRC
TRX-2      CYTCNMSSGS...........KVKCQK..CQKNYHSTCLGTSKRLLGADRP.............................LICVNC
TRX-3      CQRCYDDNDFDLK........MMECGD..CGQWVHSKCEGLSDEQYNLLSTLPESIE.......................FICKKC
TRX-4      CTVCGNRGA............TVGCNVRSCGEHYHYPCARSIDCAFLTDKS.............................MYCPAC
PCL-1      CVACKRSDIEDV.........VEICER..CGRGYHRGCTVEIVTGSGI................................WSCKRC
PCL-2      C.YCGKPGKFDHN........MMQCCK..CRNWFHTQCMQNFKKKLLRGDMFFV..........................FCCTVC
ASH1       C.ICGLYKDEGL.........MIQCSDK.CMVWQHTECTKADIDADN.................................YCCERC

CONSENSUS  C aC                 aa C     C    YH  Ca                                          W C  C
                                               F                                              Y
                                                                                              F
```

FIGURE 3.—Alignments of SET domain and PHD finger in ASH1 with similar motifs in other proteins. (A) Alignment of SET domain in ASH1 and other proteins. Amino acids in the 11 aligned sequences that are identical to the amino acid at that position in ASH1 are indicated by inverse letter. The order of alignment is from most to least similar to ASH1. Gaps are indicated by underscore; insertion in ORFKG1T is indicated by /. Y84.5KD is hypothetical yeast protein in CPS1-FPP1 intergenic region. K09F5.5 is hypothetical *C. elegans* protein encoded by cosmid CELK09F5. HRX is human homologue of TRX that has a total length of 3968. Y123.9KD is hypothetical yeast protein in ORC6-MSH1 intergenic region. C26E6.10 is hypothetical *C. elegans* protein encoded by cosmid CELC26E6. TRX is Drosophila *trithorax* protein that has a total length of 3759. SUVAR(3)9 is Drosophila *Suppressor of variegation (3) 9* protein that has a total length of 628. G9A is human G9a protein that has a total length of 3679. ORFKG1T is hypothetical human protein. E(Z)/PCO is protein product of Drosophila *Enhancer of zeste* also known as *polycombeotic* that has a total length of 754. *, positions of 11 amino acids that are identical in all eleven sequences. (B) Alignment of PHD finger in ASH1 with those in TRX and PCL. Inverse letters indicate match with consensus defined by R. D. NICHOLLS and P. J. HARTE (unpublished results). a in the consensus represents any aliphatic amino acid.

there are 11 residues that are invariant in all of the sequences. In human HRX, yeast 123.9KD (Swiss-Prot accession number P38827) and Drosophila TRX, SU-VAR(3)9, and E(Z)/PCO the SET domain is found in the carboxy terminus of the protein, but in ASH1, yeast 84.5KD (Swiss-Prot accession number P46995), *Caenorhabditis elegans* KO9F5.5 (Genbank accession number U37430) and C26E6.10 (Genbank accession number U13875), and human G9a (MILNER and CAMPBELL 1993) and ORFKG1T (Genbank accession number D31891) it is found in the middle of the protein.

At the carboxy terminus of the ASH1 protein (Figure 2) there is a double zinc finger-like motif called the PHD finger (SCHINDLER *et al.* 1993). This domain occurs four times in TRX (R. D. NICHOLLS and P. J. HARTE, unpublished results), twice in PCL, the product of the Drosophila Polycomb group gene *Polycomblike*. (LONIE *et al.* 1994), and in many other putative transcription factors (AASLAND *et al.* 1995; R. D. NICHOLLS and P. J. HARTE, unpublished results). An alignment of this domain in ASH1 with those in TRX and PCL is presented in Figure 3B. This domain in ASH1 matches the consensus at 12/14 positions. For comparison, the four copies of this domain in TRX match the consensus at 14/14, 11/14, 11/14 and 12/14 positions respectively, while the two copies of this domain in PCL match the consensus at 12/14 and 13/14 positions, respectively.

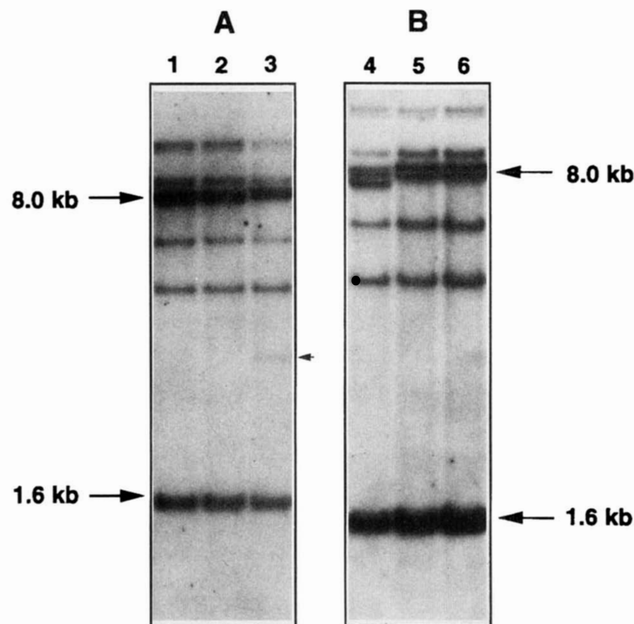

FIGURE 4.—DNA rearrangements in *ash1* mutant alleles. Genomic DNA from mutant heterozygotes was digested with *Bam*HI, fractionated by agarose gel electophoresis, blotted onto nylon membranes and hybridized with labeled 8.0- and 1.6-kb *Bam*HI fragments isolated from cosmid 6. Arrows points to 8.0- and 1.6-kb bands. Other bands represents hybridization with contaminating cosmid fragments. (A) Lane 1, control+/*TM3*; lane 2, $ash1^{28}$/*TM3*; lane 3, $ash1^{29}$/*TM3*. Arrowhead points to 4-kb band. (B) Lane 4, $ash1^{17}$/*TM3*; lane 5, $ash1^{18}$/*TM3*; lane 6, control+/*TM3*.

Similarity of ASH1 to other proteins: A search of all available protein databases using the BLAST algorithm (ALTSCHUL *et al.* 1990) revealed that the proteins most similar to ASH1 are Drosophila TRX (MAZO *et al.* 1990) and its human homologue called HRX (DJABALI *et al.* 1992; TKACHUK *et al.* 1992) or ALL (GU *et al.* 1992). There are three regions of similarity: the SET domain (residues 1300–1448 of ASH1, 30% identity), the PHD domain (residues 1778–1818 of ASH1), and a proline-rich region of 65 amino acids from 918 to 982 of ASH1 that is 24% identical to a sequence in HRX.

Analysis of mutant genes: We have recovered 31 *ash1* mutant alleles (SHEARN *et al.* 1987; TRIPOULAS *et al.* 1994). DNA rearrangements in three of the alleles were used to identify the gene within a chromosome walk. Two of these three alleles had *P*-element insertions ($ash1^{25}$ and $ash1^{26}$) and one ($ash1^{27}$) was a translocation induced by γ-rays (TRIPOULAS *et al.* 1994). We have used genomic Southern blotting to examine the other 28 alleles for DNA rearrangements. Only two of the 28 alleles revealed DNA rearrangements. One of these alleles, $ash1^{17}$, was recovered in an F_2 screen for EMS-induced alleles (TRIPOULAS *et al.* 1994). Mutant heterozygotes have both a 7.5- and an 8-kb *Bam*HI fragment each of which is half the intensity of the control 8-kb *Bam*HI fragment (Figure 4B). We inferred that the 7.5-kb fragment represents the 8-kb fragment with an internal deletion. To determine which part of the 8-kb *Bam*HI fragment might be deleted by the mutation, the entire 8-kb *Bam*HI fragment from mutant heterozygotes was amplified by PCR in eight fragments each of which was ~1 kb in length. From seven of the eight reactions only a single fragment was recovered. However, from one of the eight reactions both a 1 kb and a smaller fragment was recovered. This smaller fragment was cloned, and its nucleotide sequence was determined. The sequence revealed a deletion of 451 bases starting at nucleotide 2483. This deletion would create a truncated protein that includes the amino terminal 490 amino acids of ASH1 fused to a novel sequence of 27 amino acids (Table 1). $ash1^{17}$ appears to be an antimorphic allele. Mutant hemizygotes with the genotype $ash1^{17}$/*Df(3L)JK18* die before the third larval instar, which is earlier than the stage of lethality of the putative amorphic allele, $ash1^{22}$ (Table 1). As double heterozygotes with *trithorax* or *brahma* null alleles or with abx^{1}, $ash1^{17}$ causes a higher index of metathoracic transformations (Table 2) than observed for other *ash1* alleles (SHEARN 1989; TRIPOULAS *et al.* 1994; LAJEUNESSE and SHEARN 1995) or for $ash1^{22}$ (Table 2).

The other allele that revealed a DNA rearrangement is $ash1^{29}$, which was recovered in an F_1 screen for γ ray-induced alleles (TRIPOULAS *et al.* 1994). In mutant heterozygotes the 8-kb *Bam*HI fragment is represented by two fragments (Figure 4A). One is slightly larger than 8 kb and the other is 4 kb. This rearrangement is caused by an inversion with one breakpoint in the *ash1* gene. Additional Southern blots revealed that the breakpoint was between the *Eco*RI site at nucleotides 6075–6080 and the *Bam*HI site at nucleotides 7195–7200 (data not shown). To precisely determine the breakpoint in $ash1^{29}$, a portion of the *Bam*HI fragment that is slightly larger than 8 kb was amplified by inverse PCR, and the nucleotide sequence starting at the *Eco*RI site and proceeding toward the new *Bam*HI site was determined. The breakpoint was found to be at nucleotide 6453 (Table 1). A database search of the nucleotide sequence beyond the breakpoint revealed a 71 out of 74 nucleotide match with a sequence from the 5′ untranslated region of the *retinal degeneration C* gene that is located in salivary gland polytene chromosome band 77B1 (STEELE *et al.* 1992) just proximal to *ash1* at 76B6-11. Assuming that the proximal breakpoint of the $ash1^{29}$ inversion is in the *retinal degeneration C* gene, the calculated size of the altered 8-kb *Bam*HI fragment we detected (Figure 4A) would be 8.5 kb in length. The 4-kb fragment we also detected would be expected to be less intense than the larger fragment, as is observed, because only 0.6 kb of it would be derived from *ash1* and could hybridize to the probe. The predicted $ash1^{29}$ gene product would be a truncated protein that includes the amino terminal 1794 amino acids of ASH1 fused to a novel sequence of 21 amino acids. It would include all of the seven nuclear localization signals and the SET domain but only half of the PHD finger (Figure 10). Mutant hemizygotes with the genotype $ash1^{29}$/

TABLE 1

Alterations of DNA sequences in *ash1* mutant alleles and their consequences

Allele name	Mutagen	DNA sequence alteration	Predicted protein alteration	Stage of lethality[a,b]	Phenotype[a]
$ash1^{22}$	EMS	C 1155 → T	Gln47 → termination	3L/EP	Null
$ash1^{17}$	EMS	Deletion 2483-2934	Termination at 490	1L/3L	Antimorph
$ash1^{4}$	EMS	A 3369 → T	Lys875 → termination	2L/3L	Antimorph
$ash1^{1}$	ICR 170	Insertion of C at 4108	Termination at 1141	3L/EP	Strong hypomorph
$ash1^{11}$	γrays	C 4759 → A	Ser1248 → termination	3L/EP	Strong hypomorph
$ash1^{29}$	γrays	Inversion at 6453	Termination at 1794	PP	Hypomorph
$ash1^{14}$	EMS	T 7127 → A	Termination at 2001	LP/PhA	Weak hypomorph
$ash1^{16}$	EMS	G 7184 → A	Termination at 2001	LP/PhA	Weak hypomorph
$ash1^{21}$	EMS	G 4863 → A	Glu1284 → Lys	PP/LP	Hypomorph
$ash1^{10}$	EMS	A 5167 → T	Asn1385 → Ile	PP/LP	Hypomorph

[a] Consequences of mutant alleles was examined in mutant hemizygotes, *i.e.*, *ash1/Df(3L)JK18*.
[b] 1L, first larval instar; 3L, third larval instar; PP, prepupae; LP, late pupae; PhA, pharate adult.

Df(3L)JK18 die after puparium formation, which is later than the stage of lethality of the putative amorphic allele, $ash1^{22}$ (Table 1). As double heterozygotes with *trithorax* or *brahma* null alleles or with abx^{1}, $ash1^{29}$ causes a lower index of metathoracic transformations than observed for $ash1^{22}$ (Table 2). These data suggest that $ash1^{29}$ is a hypomorphic allele, so its truncated protein product must have partial activity.

We determined the nucleotide sequence of the 7.5-kb coding region for eight of the 26 mutant alleles that did not have alterations detectable by Southern blotting. These alleles were chosen for further analysis because they represent a range of phenotypes. The best candidate for an amorphic allele is $ash1^{22}$ (Table 1). There is a C → T transition at nucleotide 1155 in $ash1^{22}$. This causes the 47th codon to change from CAG, encoding glutamine, to TAG, a termination codon. The protein product of this mutant allele is predicted to only contain the N terminal 46 amino acids. Since $ash1^{22}$ is very likely to be a null allele, we have interpreted the phenotype of the other alleles using $ash1^{22}$ as the reference point. The stage of lethality of $ash1^{22}$ mutant hemizygotes ranges from the end of the third larval instar to puparium formation (Table 1).

The $ash1^{4}$ allele has a transversion of A → T at nucleotide 3369. This is predicted to cause the 875th codon to change from AAA, encoding lysine, to TAA, a termination codon. $ash1^{4}$ also appears to be an antimorphic allele. Mutant hemizygotes with the genotype $ash1^{4}$/*Df(3L)JK18* die before the third larval instar, which is earlier than the stage of lethality of the putative amorphic allele, $ash1^{22}$ (Table 1). As double heterozygotes with *trithorax* or *brahma* amorphic alleles or with abx^{1}, $ash1^{4}$ causes a higher index of metathoracic transforma-

TABLE 2

Dominant enhancement of homeotic transformations in *abx*, *trx* and *brm* heterozygotes caused by different *ash1* mutations

ash1 allele name	$ash1^{x}$ +/+ abx^{1}			$ash1^{x}$ +/+ trx^{el2}			$ash1^{x}$ +/+ brm^{2}		
	No. of flies	L3 → L2[a]	H → W[b]	No. of flies	L3 → L2[a]	H → W[b]	No. of flies	L3 → L2[a]	H → W[b]
+		0	0		0	0		0	0
$ash1^{22}$	86	37	86	116	102	56	122	158	69
$ash1^{17}$	120	85	68	56	143	93	96	281	134
$ash1^{4}$	122	86	194	76	168	225	98	286	303
$ash1^{1}$	170	19	34	114	40	26	154	106	11
$ash1^{11}$	292	51	50	164	119	14	120	240	135
$ash1^{29}$	180	63	38	114	62	11	158	141	16
$ash1^{14}$				26	57	12	192	58	0
$ash1^{16}$	168	78	36	124	107	25	186	95	4
$ash1^{21}$	142	4	44	120	0	1	180	6	0
$ash1^{10}$	128	35	17	110	8	1	122	19	3

[a] Index of leg 3 to leg 2 transformations is the product of the penetrance (as a percentage) and the expressivity (classified on a scale of 1–3 as described in MATERIALS AND METHODS. The range is from 0 (none of the flies have any transformations) to 300 (all of the flies have extreme transformations of leg 3 to leg 2). Examples of these transformations are presented in Figure 5.

[b] Index of haltere to wing transformations is the product of the penetrance (as a percentage) and the expressivity (classified on a scale of 1–4 as described in MATERIALS AND METHODS). The range is from 0 (none of the flies have any transformations) to 400 (all of the flies have extreme transformations of haltere to wing). Examples of these transformations are presented in Figure 6.

tions than observed for other *ash1* alleles including *ash1*[22] (Table 2).

The *ash1*[1] allele was induced by the frame shift mutagen ICR-170 (SHEARN *et al.* 1971); it has an insertion of a C after nucleotide 4108. This is predicted to cause a frame shift at amino acid residue 1032 leading to a novel sequence of 109 residues and then termination at residue 1141. *ash1*[1] appears to be a strong hypomorphic allele. Mutant hemizygotes with the genotype *ash1*[1]/*Df(3L)JK18* die at the end of the third larval instar or just after puparium formation, which is the same stage of lethality as the putative amorphic allele, *ash1*[22] (Table 1). However, as double heterozygotes with *trithorax* or *brahma* null alleles or with *abx*[1], *ash1*[1] causes a lower index of metathoracic transformations than observed for *ash1*[22] (Table 2).

The *ash1*[11] allele has a transversion of C → A at nucleotide 4759. This is predicted to cause the 1248th codon to change from TCA, encoding serine, to TAA, a termination codon. *ash1*[11] appears to be nearly an amorphic allele. Mutant hemizygotes with the genotype *ash1*[11]/*Df(3L)JK18* die at the end of the third larval instar or just after puparium formation, which is the same stage of lethality as the putative amorphic allele, *ash1*[22] (Table 1). Moreover, as double heterozygotes with *trithorax* or *brahma* null alleles or with *abx*[1], *ash1*[11] causes a similar index of metathoracic transformations as observed for *ash1*[22] (Table 2).

There are two temperature-sensitive alleles of *ash1*, *ash1*[14] and *ash1*[16]. The viability of one, *ash1*[14], is heat-sensitive (TRIPOULAS *et al.* 1994). This mutant allele has a transversion of T → A at nucleotide 7127, which alters the consensus 5′ splice site of the second intron from GT to GA. The other allele, *ash1*[16], is lethal at all temperatures, but its imaginal disc phenotype is heat-sensitive (LAJEUNESSE 1995). This mutant allele has a transition of G → A at nucleotide 7184, which alters the consensus 3′ splice site of the second intron from TAG to TAA. Both of these temperature-sensitive mutations would be expected to interfere with splicing of the second intron. If no splicing occurs, then because of stop codons in the intron, the protein product of either mutant allele would terminate after 2001 amino acids of ASH1; these mutant ASH1 proteins would be missing the C terminal 143 amino acids.

The *ash1*[21] allele has a transition of G → A at nucleotide 4863. This is expected to cause a substitution of lysine (AAG) for glutamic acid (GAG) at residue 1284, which is near the SET domain. The *ash1*[10] allele has a transversion of A → T at nucleotide 5167. This is expected to cause a substitution of isoleucine (ATC) for asparagine (AAC) at residue 1385. This asparagine at residue 1385 is one of only 11 residues in the SET domain that is identical in all of the sequences aligned in Figure 3A. Both of these mutations are considered hypomorphs (Table 1). Mutant hemizygotes with the genotype *ash1*[21]/*Df(3L)JK18* or *ash1*[10]/*Df(3L)JK18* die during the pupal period, which is later than the stage of lethality of the putative amorphic allele, *ash1*[22] (Table 1). Moreover, as double heterozygotes with *trithorax* or *brahma* null alleles or with *abx*[1], either *ash1*[21] or *ash1*[10] causes a lower index of metathoracic transformations than observed for *ash1*[22] (Table 2).

Immunohistochemical localization of the ASH1 protein: Since the ASH1 putative protein is presumed to be a nuclear protein by virtue of its multiple nuclear localization sequences and since *ash1* mutations affect gene expression (LAJEUNESSE and SHEARN 1995), we examined polytene chromosomes from larval salivary glands for localization of ASH1 protein at specific sites. We found 108 bands that reacted with the ASH1 antibody with varying intensity and a diffuse speckled pattern along the length of the chromosome (Figure 7B). As a negative control for the specificity of the antibody binding to ASH1 on polytene chromosomes, we examined polytene chromosomes dissected from *ash1*[22] hemizygous mutant salivary glands for accumulation of ASH1 protein. The polytene chromosomes from these amorphic mutant larvae (Figure 7C) are smaller and more fragile than those from wild-type larvae (Figure 7A) as can be discerned by DAPI staining. Polytene chromosomes from *E(z)* mutant larvae have similar abnormalities (RASTELLI *et al.* 1993). We detected diffuse speckled staining along the chromosomes but no accumulation at specific bands (Figure 7D). The *ash1*[22] allele is predicted to make a protein that is truncated before the fragment of ASH1 that was used as an immunogen. So, the antibody could not recognize this truncated protein even if it accumulated. Therefore, we interpret the diffuse speckled pattern as background, *i.e.*, the antibody cross-reacts with some other antigen that is distributed along the chromosomes in a diffuse speckled pattern. This negative control shows that the antibody directed against a fragment of ASH1 does not cross-react with any other protein that accumulates at specific sites on polytene chromosomes. As a positive control for our ability to detect proteins bound to *ash1*[22] mutant chromosomes, we labeled this preparation with a monoclonal antibody that detects PSC, the product of the *Posterior sex combs* gene (MARTIN and ADLER 1993). We found that *ash1*[22] does not significantly affect the number of sites of PSC localization (data not shown) although the chromosome morphology is not good enough to identify these sites.

The *ash1*[17] and *ash1*[4] mutations as hemizygotes causes lethality before the end of the third larval instar, so it is not possible to recover mutant larvae with polytene chromosomes that are suitable for immunohistochemical localization. However, we have examined polytene chromosomes from other mutant alleles for ASH1 and PSC binding (Table 1). None of the seven alleles examined affect PSC binding, our positive control. This indicates that none of these alleles disrupt polytene chromosomes to such an extent as to preclude binding or detection of proteins. Two of the alleles, *ash1*[1] and *ash1*[11], eliminate chromosome localization of ASH1.

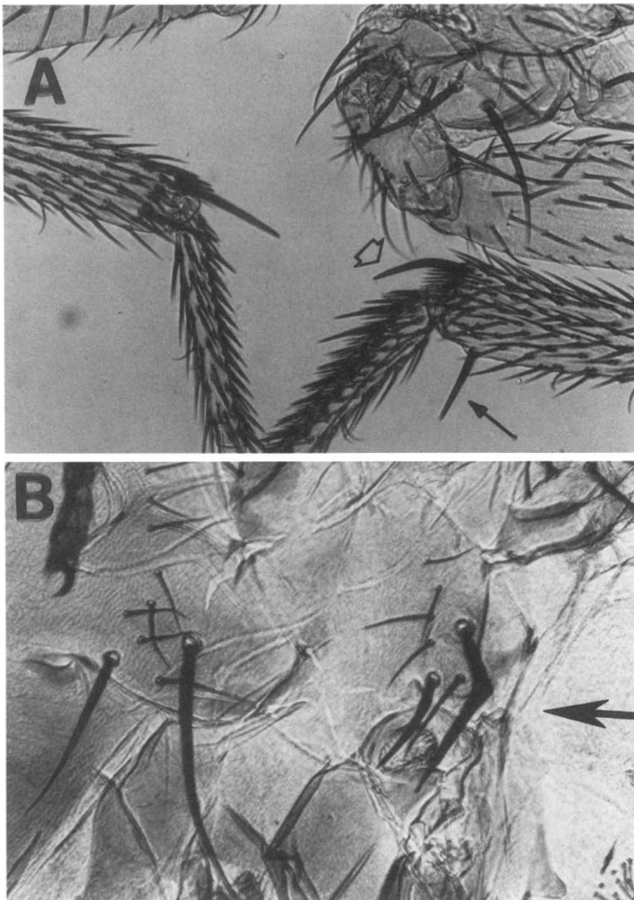

FIGURE 5.—Leg transformations observed in *ash1/abx*[1] transheterozygotes. (A) Ectopic apical bristle at anterior base of tibia on metathoracic leg (open arrowhead); ectopic preapical bristle near base of tibia on metathoracic leg (arrow). (B) Ectopic sternapleural bristles on ventral metathorax (arrow).

The *ash1*[29] mutation, an inversion that removes the 3′ end of the transcript and creates a fusion protein that includes the amino terminal 85% of ASH1, eliminates all but two or three sites of ASH1 binding. The substitution of lysine for glutamic acid at residue 1284 in *ash1*[21] or the substitution of isoleucine for asparagine at residue 1385 in *ash1*[10] reduces the intensity of accumulation of ASH1 at specific sites. The significance of this negative evidence cannot yet be determined because we do not have immunoblotting data documenting the accumulation of ASH1 protein in these mutants. By contrast, the temperature-sensitive alleles, *ash1*[14] and *ash1*[16], have wild-type accumulation of ASH1 at specific sites regardless of the temperature at which they are grown. This positive evidence indicates that these mutant proteins must have near normal stability so they accumulate at near normal levels. Since both *ash1*[14] and *ash1*[16] are predicted to cause truncation of the carboxy terminal 143 amino acids, the normal localization of these mutant proteins suggests that the carboxy terminus is not essential for chromosome localization.

An example of a preparation used to classify the location and relative intensity of the chromosomal sites of ASH1 localization is presented in Figure 8B. Some of the immunofluorescent bands correspond to ones that are intensely labeled by DAPI (Figure 8A) and some do not. We also labeled this preparation with a monoclonal antibody that detects PSC at 70 sites (Figure 9). For comparison, polyclonal antibodies have detected PSC localization at 45 sites (MARTIN and ADLER 1993) and 83 sites (RASTELLI *et al.* 1993). Some of the locations that bind PSC are near ones that bind ASH1. The location and relative intensity of the bands detected on polytene chromosomes from wild-type larvae by the anti-ASH1 antibody is presented in Table 3. Some of those bands are at locations reported for the binding of TRX, the product of the *trithorax* gene, and some are at locations reported for the binding of the products of Polycomb group genes such as PSC (Table 3). To examine whether these similar positions represented colocalization, we performed double labeling using the monoclonal antibody that detects PSC. More than 30 of the ASH1 binding sites are at or near sites reported for PSC binding and we detect many examples of ASH1 binding sites adjacent to PSC binding sites (Figure 9). Nevertheless, there are only two or three sites where we detect colocalization of ASH1 and PSC binding (Figure 9).

DISCUSSION

Analysis of the ASH1 protein: Comparison of the nucleotide sequence of five cDNAs to the nucleotide sequence of genomic *ash1* DNA revealed the presence of only two small introns. The absence of large introns provides an explanation of how a transcript as large as *ash1*, 7.5 kb, can be encoded by a 10-kb gene (TRIPOULAS *et al.* 1994). The nucleotide sequence of the *ash1* transcript predicts a protein that would be localized in nuclei because of seven potential nuclear localization signals (DINGWALL and LASKEY 1991) and would be rapidly degraded because of seven potential PEST sequences (RECHSTEINER 1990). There are three motifs in the ASH1 protein that are characteristic of some proteins that regulate transcription and/or are bound to chromosomes. There are one or two AT hook motifs (REEVES and NISSEN 1990), a SET domain (TSCHIERISCH *et al.* 1994) and a PHD finger (SCHINDLER *et al.* 1993).

The AT hook motif is important for the binding of some proteins to DNA (REEVES and NISSEN 1990). However, we have not yet determined whether ASH1 binds to DNA. Although the biochemical function of neither the SET domain nor the PHD finger have been elucidated, both of them are found in many proteins in different organisms (Figure 3; AASLAND *et al.* 1995; STASSEN *et al.* 1995; R. D. NICHOLLS and P. J. HARTE, unpublished results). This conservation of sequence suggests that each of these domains has a conserved function that could involve either protein:protein or protein:nucleic acid interactions. We are particularly interested in the fact that these domains have been found in Dro-

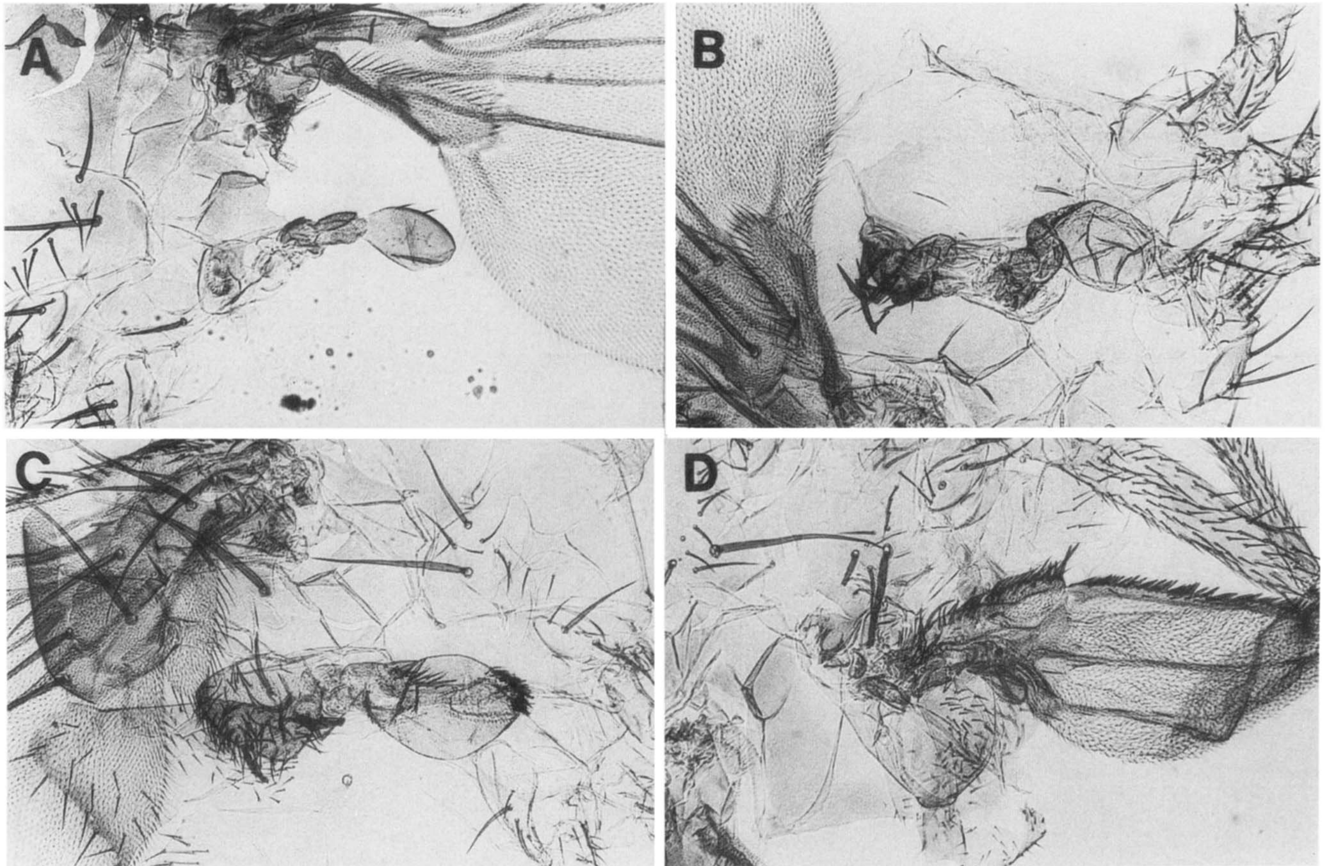

FIGURE 6.—Haltere transformations observed in *ash1/abx*^{1} transheterozygotes. (A) Ectopic bristles on capitellum. (B) Ectopic bristles on metanotum and scabellum. (C) Ectopic bristles on capitellum, metanotum and scabellum. (D) Ectopic bristles on metanotum and scabellum, transformation of capitellum to wing spread.

sophila proteins encoded by genes both of the trithorax group and the Polycomb group, since the products of these two groups of gene act antagonistically to each other. The SET domain is found in SUVAR(3)9, the product of the *Suppressor of variegation(3)9* gene that is a suppressor of position-effect variegation, E(Z)/PCO, the product of the *Enhancer of zeste/polycombeotic* gene, which has been classified as a Polycomb group gene (JONES and GELBART 1990; PHILLIPS and SHEARN 1990) but may also be classified as a trithorax group gene (LAJEUNESSE and SHEARN 1996), and TRX, the product of the trithorax group gene *trithorax*. The SET domain is at the C-terminus of SUVAR(3)9, E(Z)/PCO and TRX but is in the middle of the predicted ASH1 protein, so its location is not conserved. While there are four PHD finger domains in TRX and two PHD finger domains in PCL, the product of the Polycomb group gene *polycomblike;* there is a single PHD finger in ASH1. So, the number of PHD fingers is not conserved. Neither the SET domain nor the PHD finger is characteristic specifically of the products of the trithorax group or of the Polycomb group. However, the presence of these same domains in proteins encoded by two groups of genes that are responsible, respectively, for maintaining either the activation or repression of homeotic selector genes suggests that the biochemical functions of the products of these two groups of genes must be similar in some respect.

We had already reported that the *ash1* and *trithorax* genes were functionally related (SHEARN 1989) and that *trithorax* mutations were recovered in an F_1 screen for *ash1* mutations (TRIPOULAS *et al.* 1994). However, this functional relationship did not necessarily predict that the products of these two genes would be structurally related. So, we were quite interested to discover that the Drosophila protein most similar in sequence to ASH1 is TRX (MAZO *et al.* 1990). The similarity includes the presence of a SET domain and a PHD finger. The similarity between ASH1 and TRX also includes the presence of numerous PEST sequences, which suggests that both of these proteins are rapidly turned over. We speculate that of the ASH1 protein that is synthesized in cells only the fraction that is bound to chromosomes accumulates and the rest is degraded. This may also be true of other trithorax group proteins and could help to explain their dosage sensitivity.

Analysis of mutant ASH1 proteins: Based on phenotypic analyses, we believed that *ash1*^{17} represented an amorphic allele. As a hemizygote it causes an earlier stage of lethality than other *ash1* alleles (Table 1) and it causes a higher frequency of homeotic transformations as double heterozygotes with *trithorax* and *brahma*

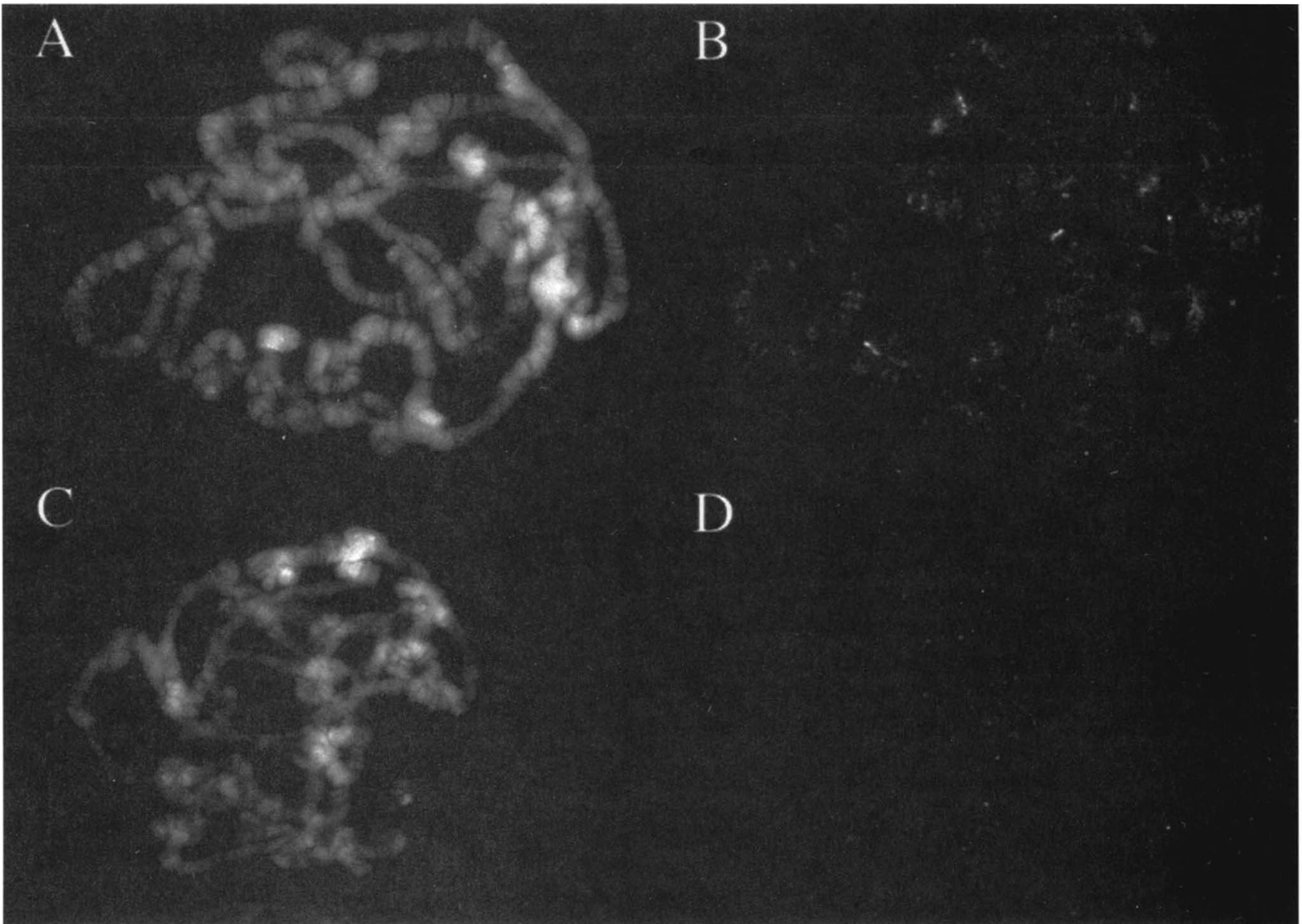

FIGURE 7.—Immunofluorescent localization of ASH1 protein on polytene chromosomes from *ash1* mutants. (A and C) DAPI-stained chromosomes. (B and D) Same chromosomes as in A and C, respectively, reacted with ASH1 antisera; the bound antibodies were detected with fluorescein-labeled goat anti-rabbit secondary antibody. (A and B) Chromosomes from wild-type larva. (C and D) Chromosomes from $ash1^{22}$ hemizygous mutant larva.

alleles than other *ash1* alleles (Table 2). Nucleotide sequence analysis indicated that the truncated protein encoded by the $ash1^{17}$ mutant gene would contain 492 amino acids derived from the amino terminus of ASH1 rather than the 2144 amino acids of full-length ASH1 (Figure 10). This truncated protein would be missing five of the seven nuclear localization signals, the SET domain and the PHD finger (Figure 10). So, we were surprised to discover that an allele that causes a less extreme phenotype than $ash1^{17}$ is more likely to represent an amorphic allele. That allele, $ash1^{22}$, encodes a protein that would only contain the amino terminal 47 amino acids (Figure 10). If $ash1^{22}$ is indeed an amorphic allele, then the viability of $ash1^{22}$ hemizygotes until the end of the third larval instar implies that zygotic expression of *ash1* is not required for larval viability. We have already shown that there is a maternal requirement for *ash1* expression (TRIPOULAS *et al.* 1994). So, we interpret the viability of $ash1^{22}$ hemizygotes through the third larval instar as dependent on maternally deposited *ash1* product. The imaginal discs of *ash1* mutant larvae have a variegated pattern of reduced accumulation of homeotic selector gene products (LAJEUNESSE and SHEARN 1995). The variegation of this pattern may reflect the stochastic loss of maternally derived ASH1 in imaginal discs.

Six of the mutant alleles we have sequenced represent a series of carboxy terminal truncations of the ASH1 protein (Figure 10). The most extreme truncation, caused by $ash1^{22}$, retains only 47 amino acids from the amino terminus. As discussed above, we interpret this allele as an amorph. The $ash1^{4}$ allele, like $ash1^{17}$, causes a more extreme phenotype than the putative amorphic allele, $ash1^{22}$. Also like $ash1^{17}$, $ash1^{4}$ causes truncation of most of the ASH1 protein (Figure 10). It is predicted to contain 875 amino acids derived from the amino terminus. We interpret the larval lethality of $ash1^{17}$ and $ash1^{4}$ hemizygotes as a consequence of the truncated products of these mutant genes interfering with the function of the wild-type maternally derived ASH1 product in young larvae. So, we have classified these alleles as antimorphic. The $ash1^{17}$ gene product would contain the well-conserved AT hook motif but would be missing six of the seven PEST sequences; the ash^{4} gene product would contain the same AT hook motif but would be missing five of the seven PEST sequences.

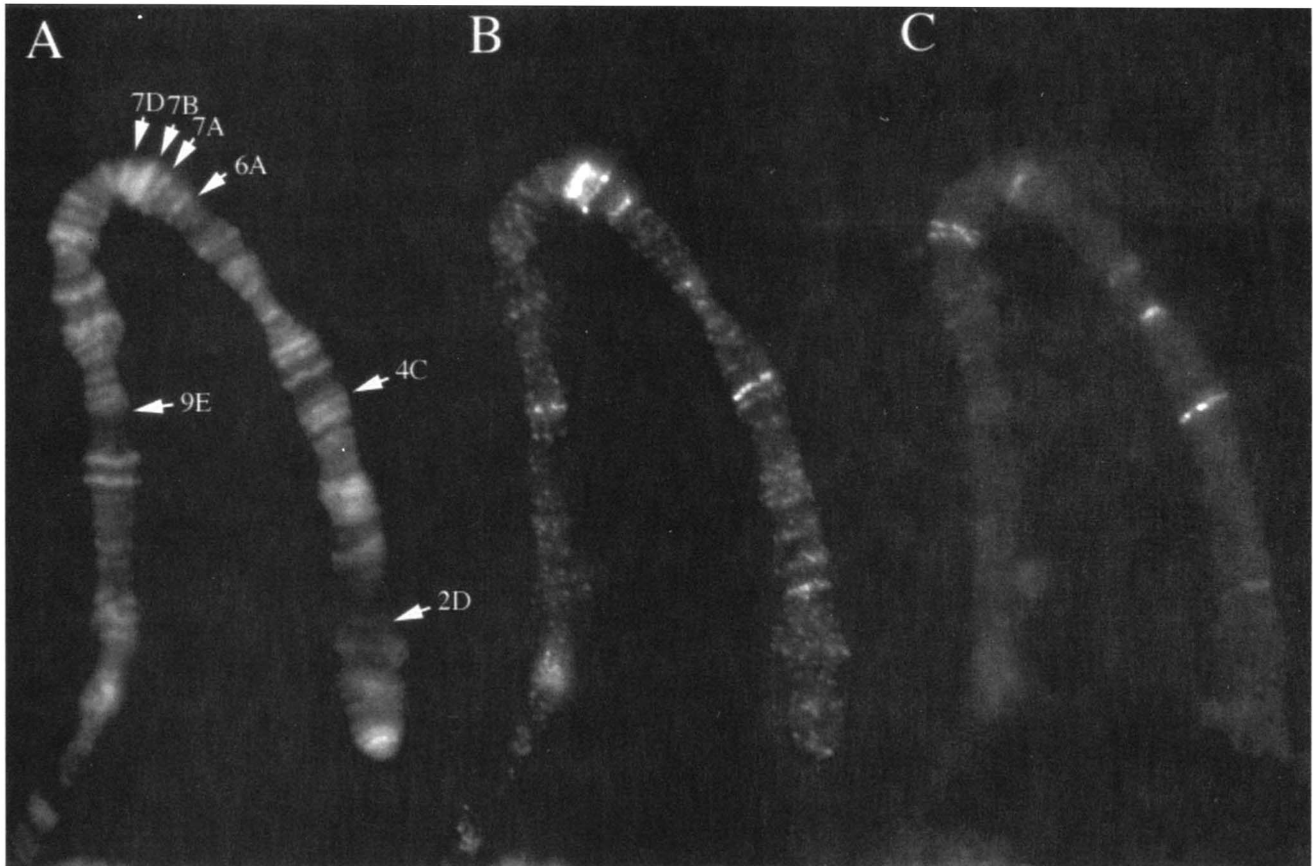

FIGURE 8.—Immunofluorescent localization of ASH1 and PSC proteins on distal half of *X* chromosome. Chromosomes from a wild-type larva were reacted simultaneously with affinity-purified polyclonal ASH1 antibodies and monoclonal PSC antisera. (A) DAPI-stained chromosome. (B) Same chromosome as in A reacted with ASH1 antisera; the bound antibodies were detected with fluorescein-labeled goat anti-rabbit secondary antibody. (C) Same chromosome as in A reacted with monoclonal PSC antisera; the bound antibodies were detected with rhodamine-labeled guinea pig anti-mouse secondary antibody.

Perhaps, these mutant proteins are more stable than the wild-type protein because of the reduced number of PEST sequences. If so, they could accumulate in mutant larvae. Their antimorphic phenotype could be derived from competition by these more stable mutant protein fragments with less stable wild-type protein derived from maternal transcripts for interaction with some other protein or nucleic acid.

The truncations caused by *ash1*^{1} and *ash1*^{11} are slightly less extreme than that caused by *ash1*^{4}. The product of *ash1*^{1} is predicted to contain 1141 amino acids from the amino terminus and the product of *ash1*^{11} is predicted to contain 1248 amino acids from the amino terminus. Both products would be missing the SET domain and the PHD finger. The phenotypes caused by *ash1*^{1} and *ash1*^{11} are also less extreme than those caused by *ash1*^{17} and *ash1*^{4} even though the products of *ash1*^{1} and *ash1*^{11} contain all of the sequences contained in *ash1*^{17} and *ash1*^{4}. Our interpretation of the observation that neither *ash1*^{1} nor *ash1*^{11} have antimorphic properties is that the product of neither of these alleles is able to interfere with the function of the wild-type maternally deposited *ash1* product. One possible explanation is that the products of both *ash1*^{1} and *ash1*^{11} contain four additional PEST sequences that might cause them to have shorter half-lives than *ash1*^{17} and *ash1*^{4} (Figure 10). Another possibility is that the additional sequences of *ash1*^{1} and *ash1*^{11} cause them to fold in such a way that they are not able to compete with wild-type ASH1. The truncation caused by *ash1*^{29} removes the carboxy terminal 350 amino acids including part of the PHD finger (Figure 10). The *ash1*^{29} allele causes a less extreme phenotype than the putative amorphic allele, *ash1*^{22} (Tables 1, 2), so we have classified it as a hypomorph. The fact that *ash1*^{29} is hypomorphic implies that the product of *ash1*^{29} retains some degree of function despite missing part of the PHD finger.

Temperature-sensitive mutations are generally interpreted as the consequence of missense mutations that lead to abnormal protein conformations at restrictive temperature. However, we found that both of the mutations that cause a temperature-sensitive phenotype, *ash1*^{14} and *ash1*^{16}, have altered nucleotides at the consensus 5′ and 3′ splice sites of the second intron, respectively, rather than missense mutations as might be expected. There are, at least, two different interpretations of these data. One interpretation is that at permissive temperatures some splicing occurs regardless of these mutations leading to some accumulation of functional

TABLE 3

Localization of ASH1, TRX, SU(Z)2, PSC, PH, PC, and Z on salivary gland polytene chromosomes[a]

Chromosome localization	Intensity	TRX	SU	PSC	PH	PC	Z
X							
1B	+						
1F	+		√				
2A	+						
2B	+						
2D	++	√		√	√	√	
3A	+		√				
3B	+						
3C	++		√				
3D	+						
4A	+						
4B	+						
4C	++	√	√	√	√	√	√
4D	+						
4F	+						
5A	+		√	√	√	√	√
5B	++						
6A	+++						
6B	++						
7A	++++	√					
7B	++++	√	√	√	√	√	√
8C/D	++++	√					
9E	++		√	√			√
10C	+						
10F	++						
11C	++	√					
12B	+						
12D	++++		√		√	√	√
12F	++++						
13A	++++						
13E	+						
14A/B	+		√	√	√	√	
15A	+						
15F	+						
*III*L							
61A	+			√	√	√	
61C	++++	√	√		√	√	
61E	+						
61F	+		√	√	√	√	
62C/D	++						
62F	+++		√	√	√	√	
63B	+						
64B	++++			√			
67B	+						
68B/C	+++		√				√
69C	++++	√	√	√	√	√	
69F/70A	+	√					
70C	+++						√
70D	++	√	√	√	√	√	
70F	++						
71C	+++						
71D	+						
71E	+						
72B	+						
72F	++++		√	√			√
74D/E	++						
75A	+						
76A	+						
77B	+						
78A	+		√				

Chromosome localization	Intensity	TRX	SU	PSC	PH	PC	Z
*II*L							
21B	+		√	√	√	√	
21C	+		√				
26A/B	+++	√					
30E	++++						
31F	++						
32A	++++						
32D	+++						
34A	+						
35A/B	+	√	√	√	√	√	
36A	++		√		√	√	
36C	+						
*II*R							
47A/B	+		√	√			√
49C	+						
49D	+						
49E/F	+	√	√	√	√	√	√
50B	+						
50C	++		√	√			√
51E	+						
51F	+						
53C	+	√					√
56C	++	√	√	√	√	√	√
57D	++++						
57E	+						
58D/E	++++	√	√		√	√	√
58F	+			√			
60E	+		√	√	√	√	√
*III*R							
83F	++++						
84A/B	+	√	√	√	√	√	√
84E/F	++		√	√	√	√	
85D/E	+++	√			√	√	
85F	+++						
87D-F	+		√	√	√	√	
88D	+	√					
88E	+						
89A/B	++++	√					
89C	+		√	√	√	√	√
89E/F	+		√	√	√	√	√
90B	+						
90A	+						
90C/D	++						
91E/F	+	√					
92A/B	+++	√		√			
92C	+						
94A/B	++++						
94C	+						
98A	++++						
98D	++++	√			√	√	
100E/F	+	√					

[a] TRX-binding sites from KUZIN *et al.* (1994) and CHINWALLA *et al.* (1995). SU(Z)2-PSC- and Z-binding sites from RASTELLI *et al.* (1993); PH-binding sites from DECAMILLIS *et al.* (1992), and PC-binding sites from ZINK and PARO (1989).

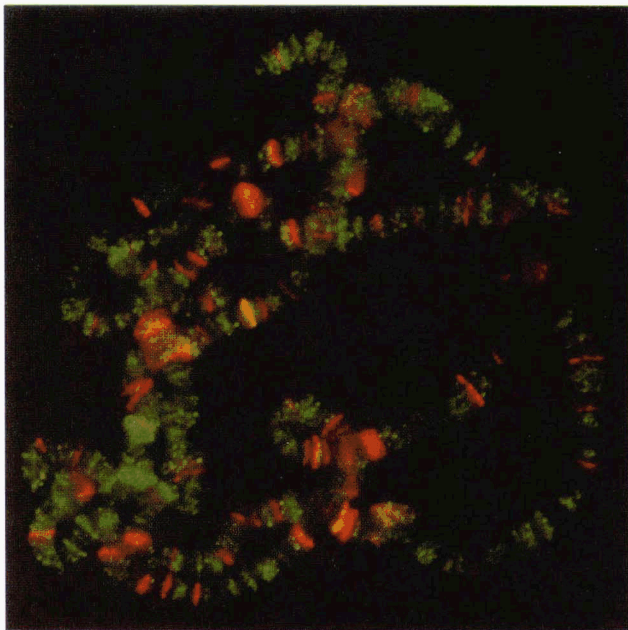

FIGURE 9.—Immunofluorescent colocalization of ASH1 and PSC proteins on polytene chromosomes. Polytene chromosomes from a wild-type larva were reacted simultaneously with affinity-purified polyclonal ASH1 antibodies and monoclonal PSC antisera. The bound antibodies were detected with fluorescein labeled goat anti-rabbit secondary antibody and rhodamine-labeled guinea pig anti-mouse secondary antibody. The image represents the merging of two separate images. Green bands indicate sites of ASH1 localization, red bands indicate sites of PSC localization and yellow bands indicate sites of ASH1 and PSC colocalization.

ASH1 product, but less or no splicing occurs at restrictive temperatures. An alternative interpretation is that little or no splicing occurs at either permissive or restrictive temperatures, but at permissive temperatures the mutant proteins retain some function. The fact that $ash1^{16}$ causes a slightly more extreme phenotype than does $ash1^{14}$ implies that disruption of the 3′ splice site of the second intron allows less splicing than alteration of the 5′ splice site.

The SET domain appears to be essential for ASH1 function since the amino acid substitution in $ash1^{10}$ alters one of 11 invariant residues of the domain, the asparagine that is residue 1391 of ASH1 (Table 1; Figure 10). However, an alternative interpretation is that this amino acid substitution prevents accumulation of mutant protein. Immunoblotting of mutant extracts would allow us to distinguish between these interpretations.

Chromosomal localization of ASH1: An antibody made against a fragment of ASH1 protein detects ~100 sites of ASH1 localization on polytene chromosomes from wild-type larvae. The absence of detectable localization to specific sites on polytene chromosomes from larvae hemizygous for a putative amorphic allele indicates that the localization of antibody to specific sites represents accumulation of ASH1 protein. We have argued that the viability of larvae hemizygous for this allele, $ash1^{22}$, depends upon ASH1 protein derived from maternal transcript. So the absence of detectable localization at specific sites indicates that the amount of maternally derived ASH1 on mutant polytene chromosomes is below the level of detectability. It is possible, but we think unlikely, that the diffuse speckled pattern of staining we observe on wild-type and mutant polytene chromosomes represents maternally derived ASH1. Mutations of *ash1* cause homeotic imaginal disc transformations like those caused by loss of function mutations in homeotic selector genes (SHEARN 1989). We have shown that these mutations cause such phenotypes in thoracic imaginal discs by reducing accumulation of the products of homeotic selector genes (LAJEUNESSE and SHEARN 1995). Genetic evidence suggests that *ash1* mutations cause reduced accumulation of the *Ultrabithorax* gene product by disrupting transcription of the *Ultrabithorax* gene (LAJEUNESSE and SHEARN 1995). The evidence presented here that ASH1 is localized at specific sites on polytene chromosomes supports the genetic evidence that ASH1 regulates homeotic gene expression at the level of transcription.

Of the 108 sites of ASH1 localization on polytene chromosomes, 23 of them are similar to sites reported for TRX localization (KUZIN *et al.* 1994; CHINWALLA *et al.* 1995). This is noteworthy because ASH1 and TRX are both activators of homeotic gene expression and it has been proposed that they are components of a multimeric protein complex (SHEARN 1989). Colocalization of these two proteins would certainly be consistent with such a proposal. Recent data that in salivary glands from larvae homozygous for the temperature-sensitive allele, $ash1^{14}$, TRX binding to polytene chromosomes is reduced at nonpermissive temperatures is also consistent with such a proposal (KUZIN *et al.* 1994). There are many more sites of ASH1 and TRX binding than there are homeotic selector genes. This implies that ASH1, TRX and probably the products of other trithorax group genes are responsible for maintaining the active expression of many different genes not only the homeotic selector genes.

As indicated in Table 3, some of the sites of ASH1 l localization on polytene chromosomes are also similar to sites of PC, PH, SU(Z)2, PSC and Z localization (ZINK and PARO 1989; DECAMILLIS *et al.* 1992; RASTELLI *et al.* 1993). PC, PH, and PSC are the products of the *Polycomb, Polyhomeotic* and *Posterior Sex Combs* genes, respectively, which are members of the polycomb group of genes that are repressors of homeotic gene expression. CHINWALLA *et al.* (1995) have reported that 32 of the 63 TRX sites colocalize with PC, and 27 of the 63 are at sites reported for PSC localization. Of the 108 sites of ASH1 localization on polytene chromosomes, 27 of them are similar to sites reported for PSC localization (RASTELLI *et al.* 1993). However, we have found only two or three sites of colocalization of ASH1 and PSC (Figure 9). So, while many of the sites of ASH1 localization are near sites of PSC localization, these sites are

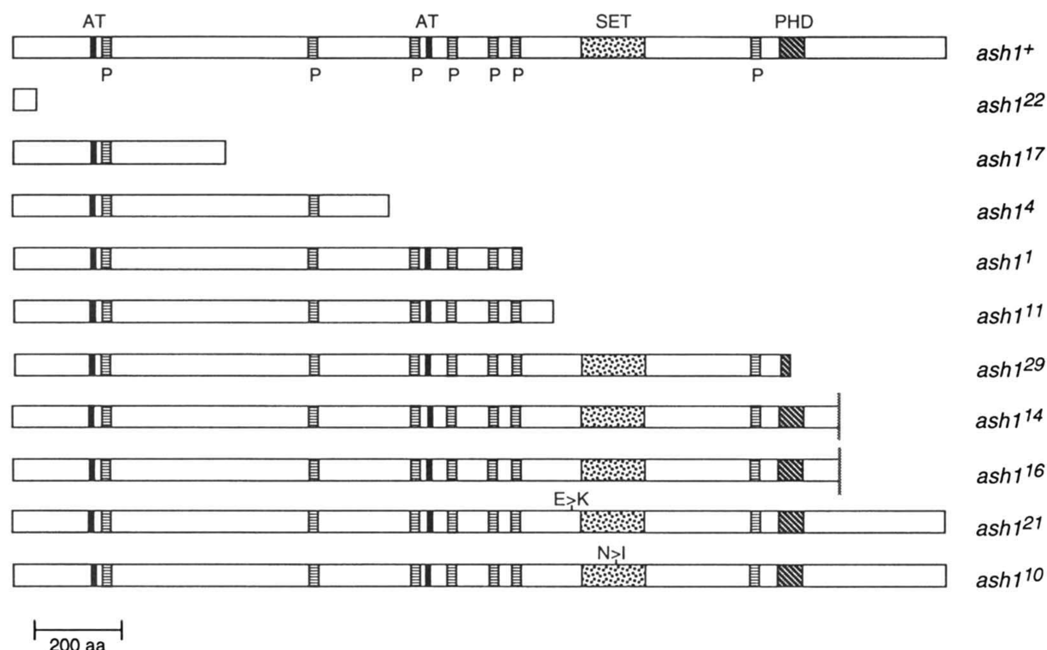

FIGURE 10.—Diagram of wild-type and mutant ASH1 proteins including locations of motifs: AT, AT hook motifs indicated by ■; P, PEST sequences indicated by ▤; SET, SET domain indicated by ▨; PHD, PHD finger indicated by ▧; wavy line at C terminus of $ash1^{14}$ and $ash1^{16}$ indicates site of truncation if no splicing occurs.

not identical. We are not certain of how near adjacent sites of ASH1 and PSC localization could be without appearing to be colocalized. The data of WEEKS *et al.* (1993) on localization of *Pol*IIA and heat shock factor at *hsp10-lacZ* transgene sites suggests that sites separated by as little as 11 kb may not appear to be colocalized. Since most homeotic selector genes have large regulatory regions, it is possible that ASH1 and PSC are both localized in more than two or three genes but at distant sites within those genes. Nevertheless, the absence of extensive colocalization at a given stage of development is consistent with any model of gene regulation that involves either the binding of activators of gene expression excluding the binding of repressors of gene expression or the binding of repressors excluding the binding of activators. Loss of ASH1 function leads to the loss of expression of some homeotic selector genes, and loss of function of Polycomb group genes leads to ectopic expression of the same genes. If the sites of ASH1 binding on polytene chromosomes represent genes that are being expressed and the sites of Polycomb group protein binding represent genes that are not being expressed, then it might be expected that the ASH1 sites would be occupied by Polycomb group proteins in an ASH1 mutant that was a protein null. However, we detected no alteration of PSC binding in *ash1* mutants, including larvae hemizygous for $ash1^{22}$, a putative amorphic allele. Either this expectation is based on a false interpretation of the antagonism between the trithorax and Polycomb group or polytene chromosome binding is not an appropriate assay system to examine this antagonism.

We thank ROXANN ASHWORTH for nucleic acid sequencing, ELIZABETH MANSFIELD for help with purifying the ASH1 fusion protein, GLORIA LIM for purifying antibody, EVELYN HERSPERGER for advice on chromosome squashes, PETER HARTE and ALEXANDER MAZO for discussions about work done in their labs on the TRX protein and for sharing data on chromosome binding of TRX before publication, PETER HARTE for providing *trithorax* alleles, JIM KENNISON for providing *brahma* alleles and *Df(3L)JK18* (a deletion that includes *ash1*), PAUL ADLER for providing us with PSC monoclonal antibody, THOMAS BLUMENTHAL for information about the nomenclature of the *C. elegans* proteins aligned with ASH1 in Figure 3, MARTIN RECHSTEINER and GREG PRATT for performing the PESTFIND analysis, and JAMES T. VANRENSSELAER for photography. This work was supported by a grant from the National Institutes of Health (GM-53058) to A.S.

LITERATURE CITED

AASLAND, R., T. J. GIBSON and A. F. STEWART, 1995 The PHD finger: implications for chromatin mediated transcriptional regulation. Trends Biochem. Sci. **20:** 56–59.

AKAM, M., 1987 The molecular basis for metameric pattern in the Drosophila embryo. Development **101:** 1–22.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 404–410.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

BAIROCH, A., 1992 PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **20:** (Supplement) 2013–2018.

BIENZ, M., and G. TREMML, 1988 Domain of *Ultrabithorax* expression in Drosophila visceral mesoderm from autoregulation and exclusion. Nature **333:** 576–578.

BREEN, T. R., and P. J. HARTE, 1991 Molecular characterization of the *trithorax* gene, a positive regulator of homeotic gene expression in Drosophila. Mech. Dev. **35:** 113–127.

BREEN, T. R., and P. J. HARTE, 1993 *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.

BROWER, D. L., 1987 *Ultrabithorax* gene expression in Drosophila imaginal discs and larval nervous system. Development **101:** 83–92.

CAVENER, D. R., 1987 Comparison of the consensus sequence flank-

ing translation start sites in Drosophila and vertebrates. Nucleic Acids Res. **15:** 1353–1361.

CHINWALLA, V., E. P. JANE and P. J. HARTE, 1995 *In vivo* chromosome sites of *trithorax*, a regulator of the homeotic genes in Drosophila. EMBO J. **14:** 2056–2065.

DECAMILLIS, M., C. NIANSHENG, 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.

DINGWALL, C., and R. A. LASKEY, 1991 Nuclear targeting sequences—a consensus? Trends Biochem Sci. **16:** 478–481.

DJABALI, M., L. SELLERI, P. PARRY, M. BOWER, B. D. YOUNG *et al.*, 1992 A *trithorax* like gene is interrupted by 11q23 translocations in acute leukemias. Nature Genet. **2:** 112–118.

FUCHS, R., 1991 MacPattern: protein pattern searching on the Apple Macintosh. Comput. Applic. Biosci. **5:** 227–232.

GU, Y., T. NAKAMURA, H. ADLER, R. PRASAD, O. CANAANI *et al.*, 1992 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.

HAFEN, E., M. LEVINE and W. J. GEHRING, 1984 Regulation of *Antennapedia* transcript distribution by the bithorax complex in Drosophila. Nature **307:** 287–289.

HOEY, T., and M. LEVINE, 1988 Divergent homeo box proteins recognize similar DNA sequences in Drosophila. Nature **332:** 858–861.

HOOVER, K. K., A. J. CHIEN and V. G. CORCES, 1993 Effects of transposable elements on the expression of the *forked* gene of *Drosophila melanogaster*. Genetics **135:** 507–526.

HULTMARK, D., R. KLEMENZ, and W. GEHRING, 1986 Translational and transcriptional control elements in the untranslated leader of the heat shock gene *hsp22*. Cell **44:** 429–438.

JONES, R. S., and W. M. GELBART, 1990 Genetic analysis of the *Enhancer of zeste* locus and its role in gene regulation in *Drosophila melanogaster*. Genetics **126:** 185–199.

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.

KENNISON, J. A., and J. W. TAMKUN, 1988 Dosage-dependent modifiers of *Polycomb* and *Antennapedia* mutations in Drosophila. Proc. Natl. Acad. Sci. USA **85:** 8136–8140.

KRASNOW, M. A., E. E. SAFFMAN, K. KORNFELD and D. S. HOGNESS, 1989 Transcriptional activation and repression by Ultrabithorax proteins in cultured Drosophila cells. Cell **57:** 1031–1043.

KUZIN, B., S. TILLIB, Y. SEDKOV, L. MIZROKHI and A. MAZO, 1994 The Drosophila *trithorax* gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene *fork head*. Genes Dev. **8:** 24787–2490.

KUZIORA, M. A., and W. MCGINNIS, 1988 Autoregulation of a Drosophila homeotic selector gene. Cell **55:** 477–485.

LAJEUNESSE, D., 1995 Trans-regulation of the Drosophila homeotic selector genes by three trithorax group genes. PhD thesis. The Johns Hopkins University.

LAJEUNESSE, D., and A. SHEARN, 1995 Trans-regulation of thoracic homeotic selector genes of the Antennapedia and bithorax complexes by the trithorax group genes: *absent, small, and homeotic discs 1* and *2*. Mech. Dev. **53:** 123–139.

LAJEUNESSE, D., and A. SHEARN, 1996 E(z): a Polycomb group gene or a trithorax group gene? Development (in press).

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

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.

MARTIN, E. C., and P. N. ADLER, 1993 The Polycomb group gene *Posterior Sex Combs* encodes a chromosomal protein. Development **117:** 641–655.

MAZO, A. M., D.-H. HUANG, B. A. MOZER and I. B. DAWID, 1990 The *trithorax* gene, a trans-acting regulator of the bithorax complex in Drosophila, encodes a protein with zinc-binding domains. Proc. Natl. Acad. Sci. USA **87:** 2112–2116.

MILNER, C. M., and R. D. CAMPBELL, 1993 The *G9A* gene in the human major histocompatibility complex encodes a novel protein containing ankyrin-like repeats. Biochem. J. **290:** 811–818.

MOUNT, S. M., C. BURKS, G. HERTZ, G. D. STORMO, O. WHITE *et al.*, 1992 Splicing signals in Drosophila: intron size, information content, and consensus sequences. Nucleic Acids Res. **20:** 4255–4262.

PARO, R., 1990 Imprinting a determined state into the chromatin of Drosophila. Trends Genet. **6:** 416–421.

PHILLIPS, M. D., and A. SHEARN, 1990 Mutations in *polycombeotic*, a Drosophila polycomb group gene, causes a wide range of maternal and zygotic phenotypes. Genetics **125:** 91–101.

RASTELLI, L., C. S. CHAN and V. PIROTTA, 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.

RECHSTEINER, M., 1990 PEST sequences are signals for rapid intranuclear proteolysis. Sem. Cell Biol. **1:** 433–440.

REEVES, R., and M. S. NISSEN, 1990 The AT DNA-binding domain of mammalian high mobility group I proteins. A novel peptide motif for recognizing DNA structure. J. Biol. Chem. **265:** 8573–8582.

ROGERS, S., R. WELLS and M. RECHSTEINER, 1986 Amino acid sequences common to rapidly degraded protein: the PEST hypothesis. Science **234:** 364–368.

SCHINDLER, U., H. BECKMANN and A. CASHMORE, 1993 HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. Plant J. **4:** 137–150.

SEDKOV, Y., S. TILLIB, L. MIZROKHI and A. MAZO, 1994 The bithorax complex is regulated by *trithorax* earlier during Drosophila embryogenesis than is the Antennapedia complex, correlating with a bithorax-like expression pattern of distinct early trithorax transcripts. Development **120:** 1907–1917.

SHEARN, A., 1989 The *ash1*, *ash2* and *trithorax* genes of *Drosophila melanogaster* are functionally related. Genetics **121:** 517–525.

SHEARN, A., T. RICE, A. GAREN and W. GEHRING, 1971 Imaginal disc abnormalities in lethal mutants of Drosophila. Proc. Natl. Acad. Sci. USA **68:** 2695–2698.

SHEARN, A., E. HERSPERGER and G. HERSPERGER, 1987 Genetic studies of mutations at two loci of *Drosophila melanogaster* which cause a wide variety of homeotic transformations. Roux's Arch. Dev. Biol. **196:** 231–242.

STASSEN, M. J., D. BAILEY, S. FOX, V. CHINWALLA and P. J. HARTE, 1995 The Drosophila *trithorax* proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. Mech. Dev. **52:** 1–15.

STEELE, F. R., T. WASHBURN, R. RIEGER and J. E. O'TOUSA, 1992 Drosophila *retinal degeneration C* (*rdgC*) encodes a novel serine/threonine protein phosphatase. Cell **69:** 669–676.

TAMKUN, J. W., R. DEURING, M. P. SCOTT, M. KISSINGER, A. M. PATTATUCCI *et al.*, 1992 *brahma:* A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell **68:** 561–572.

TKACHUK, D. C., S. KOHLER and M. L. CLEARY, 1992 Involvement of a homolog of Drosophila *trithorax* by 11q23 chromosomal translocations in acute leukemias. Cell **71:** 691–700.

TRIPOULAS, N. A., E. HERSPERGER, D. LAJEUNESSE and A. SHEARN, 1994 Molecular genetic analysis of the *Drosophila melanogaster* gene *absent, small or homeotic discs1* (*ash1*). Genetics **137:** 1027–1038.

TSCHIERSCH, B., A. HOFMANN, V. KRAUSS, R. DORN, G. KORGE *et al.*, 1994 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.

WEEKS, J. R., S. E. HARDIN, J. SHEN, J. M. LEE and A. GREENLEAF, 1993 Locus-specific variation in phosphorylation state of RNA polymerase II *in vivo:* correlations with gene activity and transcript processing. Genes Dev. **7:** 23298–2344.

ZINK, B., and R. PARO, 1989 *In vivo* binding of a trans-regulator of homeotic genes in *Drosophila melanogaster*. Nature **337:** 468–471.

Communicating editor: R. E. DENELL