© 1995 Oxford University Press Nucleic Acids Research, 1995, Vol. 23, No. 17 **3441–3448**

Transcriptional regulation of the *Sex-lethal* gene by helix-loop-helix proteins

Kazuyuki Hoshijima, Ayako Kohyama+, Ikuko Watakabe§, Kunio Inoue, Hiroshi Sakamoto† and Yoshiro Shimura*

Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606, Japan

Received June 1, 1995; Revised and Accepted August 8, 1995 DDBJ accession no. D50435

ABSTRACT

Somatic sex determination in *Drosophila* depends on the expression of *Sex-lethal (Sxl)*, whose level is determined by the relative number of X chromosomes and sets of autosomes (X:A ratio). The first step in regulation of *Sxl* expression is transcriptional control from its early promoter and several genes encoding transcription factors of the helix-loop-helix (HLH) family such as *daughterless (da), sisterless-b (sis-b), deadpan (dpn)* and *extramacrochaetae (emc)* have been implicated. By the use of transfection assays and *in vitro* binding experiments, here we show that da/sis-b heterodimers bind several sites on the *Sxl* early promoter with different affinities and consequently tune the level of active transcription from this promoter. Interestingly, our data indicate that repression by the dpn product of *da/sis-b* dependent activation results from specific binding of dpn protein to a unique site within the promoter. This contrasts with the mode of emc repression, which inhibits the formation of the da/sis-b heterodimers. These results reveal the molecular mechanisms by which *Sxl* gene transcription is positively or negatively regulated to control somatic sex determination.

INTRODUCTION

Helix-loop-helix (HLH) proteins compose a growing family of transcriptional regulatory factors that have been shown to play important roles in controlling tissue-specific gene expression (for review, see refs 1,2). Helix-loop-helix proteins can form homodimers or heterodimers with other HLH proteins to bind DNA in a sequence-specific manner. The HLH domain is essential for dimerization and the basic region which presents immediately upstream of the HLH domain is required for specific DNA binding (3,4). The consensus binding sites of these basic-HLH proteins follow the sequence motif of CANNTG, which is designated E box (3). In addition to the typical basic-HLH proteins, HLH proteins lacking the basic region or containing a proline residue in this region have been isolated from both vertebrates and invertebrates (2). These aberrant HLH proteins have altered or lost DNA binding properties and some of them have been shown to function negatively (5–9). Thus, members of the HLH protein family can regulate transcription in both positive and negative manners.

In *Drosophila*, HLH proteins have been implicated in somatic sex determination as well as neurogenesis (for review, see ref. 2). In flies, the number of X chromosomes relative to the sets of autosomes (the X:A ratio) is the first signal in the sex determination pathway (10). This X:A ratio regulates the activity state of a binary switch gene, *Sex-lethal (Sxl)*, which plays a pivotal role in sex determination of somatic cells, including dosage compensation (11,12). When the ratio is 1:1, *Sxl* is active and its gene product acts upon downstream genes to produce females. On the other hand, when the X:A ratio is 1:2, *Sxl* remains inactive, resulting in male development. The genes responsible for regulation of *Sxl* expression have been identified by genetic and molecular studies. The *sisterless-a* [*sis-a* (13)], *-b* [*sis-b* (14,15)] and *runt* [*run* (16,17)] genes, which are necessary for *Sxl* activation in females, are all located in the X chromosomes. Conversely, *deadpan (dpn)* which inhibits *Sxl* activation in males is known to reside in an autosome (18). In addition, two maternal genes, *daughterless* [*da* (11,13,14,19)] and *extramacrochaetae* [*emc* (18)] provide further positive and negative regulation. Thus, the X:A signal appears to be achieved by the balance between the activators in the X chromosomes, termed as numerators, and the repressors in the autosomes, termed as denominators. The X:A signal controls *Sxl* expression in conjunction with maternal genes. Several lines of evidence suggested that these genes regulate the activity state of *Sxl* at the transcriptional level. First, the early *Sxl* promoter, which is localized further downstream of the late promoter, is active for a short period in early female embryos which coincides with the period when the X:A ratio determines which of the two sexual pathways to follow (20,21). Throughout other developmental stages, however, the upstream promoter is active in both sexes but its transcripts produce a functional product only in females by sex-specific splicing (22–24). Secondly, all the proteins encoded by the above-mentioned genes have the characteristics of transcriptional regulatory factors; *sis-a* encodes a basic leucine zipper protein (25), *sis-b* (26) and *da* (27,28) encode typical basic-HLH proteins and *dpn* (29) and *emc* (30,31) encode aberrant HLH proteins either containing a proline residue in the basic region or lacking this region, respectively. The dimerization and DNA binding domains of the *run* gene product show similarity with those of putative mammalian and viral

---

* To whom correspondence should be addressed

Present addresses: +Department of Physics, Faculty of Science, University of Tokyo, Tokyo 113, Japan, §Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA and †Department of Biology, Faculty of Science, Kobe University, Kobe 657, Japan transcriptional regulatory factors (32). Third, these genes are all expressed during the period when the early *Sxl* promoter is active (16,18,25,33) and furthermore, mutations in some of these genes have been shown to influence transcription from the early *Sxl* promoter (20,21).

It has been shown that the four HLH proteins involved in sex determination also affect neurogenesis. The *sis-b* gene is identical with the *scute* (*sc*) gene, a member of the *achaete-scute* complex (AS-C) (34) which plays a key role in the development of the central and peripheral nervous systems (35). The *sc* gene cooperates with the *da* gene to activate proneural genes (7). The *da* gene is broadly expressed throughout the developmental stages (33) and is known to also participate in oogenesis (36). The *emc* gene inhibits *da/sis-b* dependent activation of transcription by interfering with da/sis-b heterodimer formation (37). The *dpn* gene was first isolated during the search for essential neurogenic genes, but its function in neurogenesis still remains to be clarified (29). Thus, the molecular mechanisms of these genes with an exception of *dpn*, have been well analyzed in neurogenesis, but not in sex determination.

On the basis of transfection assays in combination with *in vitro* binding experiments, here we present evidence indicating that the HLH proteins, which contribute to the X:A signal, directly regulate the early *Sxl* promoter. We found that the da/sis-b heterodimer directly activates the early *Sxl* promoter by binding to both high and low affinity binding sites. Our data also indicate that the dpn protein represses *da/sis-b* dependent activation by specific binding to a unique site within the promoter. This differs from the mode of emc repression, which inhibits formation of the da/sis-b heterodimers. Our observations reveal insight into the molecular mechanisms by which *Sxl* gene transcription is positively or negatively regulated in control of somatic sex determination.

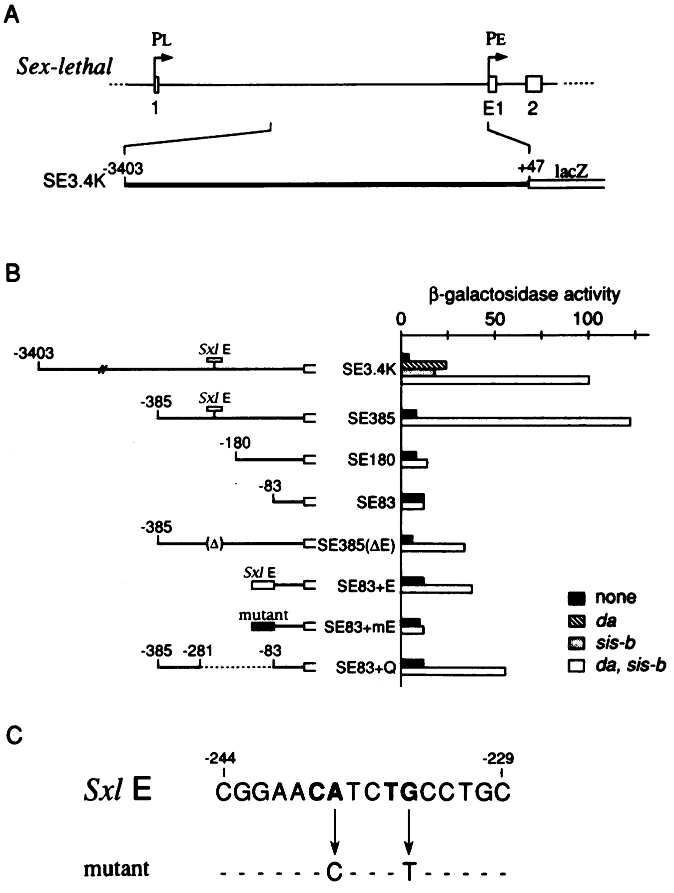

Figure 1. Da and sis-b proteins cooperatively activate the early *Sxl* promoter in transient transfection assays using *Drosophila* Kc cells. (**A**) Schematic representation of the locations of the late and early promoters (PE and PL, respectively) of the *Sxl* gene. The transcription start site of the early promoter and the following embryonic exon (E1) are present ~5 kb downstream of the first exon (20). SE3.4K is a reporter construct containing the region in the early promoter spanning from 3403 bp upstream to 47 bp downstream of the transcription start site, which is fused upstream of the ADH-LacZ reporter gene (see Materials and Methods). (**B**) Analyses of the early promoter region responsible for *da/sis-b* dependent activation. The numbers of individual reporter constructs represent sites upstream of the transcription start site of the early promoter. The open box indicated as *Sxl* E on SE3.4K and SE385 represents the conserved E box located ~240 bp upstream of the transcription start site, whose sequence and location are shown in (C). In SE385(ΔE) and SE298(ΔE), the E-boxes are substituted by the *Sal*I sequence. In SE83 + E and SE83 + mE, the fragments containing the E-box sequence (open box indicated as *Sxl* E) and the mutated E-box sequence (shaded box indicated as mutant) are inserted immediately upstream of the 83 bp promoter region of SE83 respectively. Three micrograms of individual reporter constructs are transfected into *Drosophila* Kc cells with 4 μg of expression vector (filled bar), 2 μg of expression vector and either 2 μg of *da* or *sis-b* expression plasmid (hatched and shaded bars respectively) or 2 μg each of *da* and *sis-b* expression plasmids (open bar), in addition to 2 μg of the luciferase gene expression plasmid as control. The β-galactosidase activities of the reporter constructs are normalized to luciferase activity to control for variations in transfection efficiency. The values of the activities are shown as a percentage of the β-galactosidase activity obtained by transfection of SE3.4K with both *da* and *sis-b* expression plasmids. (**C**) Nucleotide sequence of the E box located ~240 bp upstream of the transcription start site. The conserved CA and TG dinucleotides of the E box, CANNTG, are bold. Mutant represents the substituted nucleotides of the E box.

MATERIALS AND METHODS

Plasmid constructions

A genomic clone containing the early *Sxl* promoter region was isolated from a *Drosophila* genomic library and subcloned into pSP73 (Promega). The reporter constructs were made by introducing various regions of the early *Sxl* promoter into a chimeric ADH-lacZ vector, pCaSpeR-AUG-βgal (38). For construction of SE3.4K, SE385, SE180, SE83 and SE83 + Q, the regions of the early *Sxl* promoter were used as indicated in Figure 1B. All fragments were prepared from the subcloned *Sxl* genomic DNA fragment by restriction enzyme digestion or PCR amplifications. For SE385(ΔE) and SE298(ΔE), the promoter fragments were prepared from subcloned plasmid in which the E-box site, A<u>CA</u>TC<u>TG</u>C, was substituted to a *Sal*I site, <u>G</u>GTCGA<u>C</u>A (the conserved sequences of E-box and *Sal*I site are underlined). For SE83 + E, SE180 + E and SE83 + mE, the fragments containing the wild-type or mutated E box were identical with the oligonucleotides used for the gel retardation assay (see below). These fragments were inserted into the *Bam*HI site located immediately upstream of the promoter fragments in SE83 or SE180, creating SE83 + E, SE180 + E and SE83 + mE. For constructs S1–S11, the regions encompassed in this series are indicated in the legend to Figure 2. These fragments were designed with a *Bgl*II site at the 5′ end and a *Bam*HI site at 3′ end. Three copies of each fragment were inserted in the same orientation into the *Bam*HI site of SE83. For SE83 + E•D, the fragment containing the dpn binding site (D-box), identical with

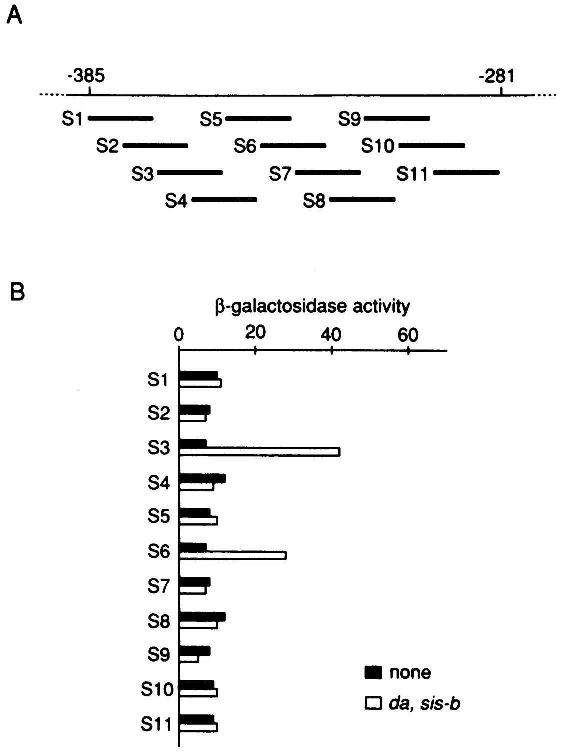

Figure 2. Other elements in addition to the E box responsible for *da/sis-b* dependent activation. (**A**) Schematic representation of the 11 fragments spanning from the region 385 to 281 bp upstream of the transcription start site. Each fragment contains 16 nucleotides with seven nucleotides overlapping with neighboring fragments. The location of each fragment relative to the promoter region (bp) is as follows: S1, –385 to –370: S2, –376 to –361: S3, –367 to –352: S4, –358 to –343: S5, –349 to –334: S6, –340 to –325: S7, –331 to –316: S8, –322 to –307: S9, –313 to –298: S10, –304 to –289: S11, –295 to –280. (**B**) Mapping of the elements able to support *da/sis-b* dependent activation. Three copies of each fragment described in panel A are inserted in the same orientation immediately upstream of the 83 bp promoter region of SE83. The reporter constructs are transfected and their β-galactosidase activities are shown as described in Figure 1B.

the oligonucleotide used in the gel retardation assay (see below), was inserted into the regenerated *Bam*HI site downstream of the E-box fragment in SE83 + E.

The vector used for expression of da, sis-b and dpn products in tissue culture cells was as described previously (39). To construct the *da* expression plasmid, the full coding region with additional 6 bp at the upstream side was PCR amplified from *da* genomic DNA with primers carrying a *Sal*I site at the 5′ end and a *Hin*dIII site at the 3′ end. The resulting product was excised with *Sal*I and *Hin*dIII and inserted into the *Sal*I and *Hin*dIII digested vector. For the *sis-b* expression plasmid, an *Eco*RI fragment from a *sis-b* genomic DNA clone, pC31/B11 (26), was inserted into the *Xba*I digested, Klenow blunted vector. The *dpn* expression plasmid was constructed by PCR-amplification of the protein coding region from *dpn* cDNA in pBluescript. The 5′ end primer contained a *Sal*I site and the 3′ end primer a *Hin*dIII site. The product was cut with *Sal*I and *Hin*dIII and inserted into the *Sal*I–*Hin*dIII digested vector.

The bacterial expression plasmids of da and sis-b were constructed from pGEX-3X (Pharmacia). To construct GST-da, the same *Sal*I–*Hin*dIII fragment used for the *da* expression plasmid in tissue culture cells was Klenow blunted and inserted into *Eco*RI digested and Klenow blunted pGEX-3X. For GST-sis-b, a *Vsp*I-*Eco*RI fragment of pC31/B11 was blunt-ended with Klenow and inserted into *Eco*RI-Klenow blunted pGEX-3X. The bacterial expression plasmids of dpn and emc were constructed from pMAL-cRI (New England Biolabs). For MBP-dpn, the same *Sal*I–*Hin*dIII fragment used for the *dpn* expression plasmid in tissue culture cells was inserted into *Sal*I and *Hin*dIII digested pMAL-cRI. For MBP-emc, the protein coding region with additional 12 bp at the upstream side was PCR amplified from *emc* cDNA cloned in Bluescript with primers carrying a *Sal*I site at the 5′ end and a *Hin*dIII site at the 3′ end. The product was cut with *Sal*I and *Hin*dIII and inserted into the *Sal*I–*Hin*dIII digested pMAL-cRI.

Cotransfection assay

Cultivation of *Drosophila* Kc cells and transient transfection assays were performed as described previously (39). β-galactosidase activity was assayed by the procedure of Edlund and colleagues (41). For all assays, 2 μg of the luciferase gene expression plasmid was cotransfected as control and luciferase activity was assayed as recommended by the manufacturer (Promega Corporation). The β-galactosidase activity was normalized to luciferase activity to control for variations in transfection efficiency.

Protein expression and purification

Recombinant proteins were expressed in BL21(DE3)pLysS with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 1 or 2 h at 30°C. Da and sis-b proteins were individually fused to glutathione *S*-transferase (GST) and purified as recommended by the manufacturer (Pharmacia). Emc and dpn proteins were individually fused to maltose-binding protein (MBP) and purified as recommended by the manufacturer (New England Biolabs). All the proteins were dialyzed overnight against the D′K50 buffer [20 mM K$^+$ HEPES at pH 7.9, 50 mM KCl, 0.2 mM ethylenediaminetetraacetic acid (EDTA), 0.1 mM dithiothreitol (DTT), 0.2 mM 4-amidinophenylmethane-sulfonyl fluoride (APMSF), 10% glycerol] and were quantitated by the protein assay (Biorad Co.).

Gel retardation assay

The probes used in the gel retardation assays were generated by annealing synthesized oligonucleotides and end-labeled with Klenow using [α-^{32}P]dATP. The sequences of the probes are as follows:
Sxl E-box, 5′-GATCTCGGAACATCTGCCTGCG-3′
and 5′-GATCCGCAGGCAGATGTTCCGA-3′;
mutated E-box, 5′-GATCTCGGAACCTCTTCCTGCG-3′ and
5′-GATCCGCAGGAAGAGGTTCCGA-3′;
S2, 5′-GATCTAGCAATACAAAATGCAG-3′ and
5′-GATCCTGCATTTTGTATTGCTA-3′;
S3, 5′-GATCTAAATGCAGCTTGCTTCG-3′ and
5′-GATCCGAAGCAAGCTGCATTTA-3′;
S6, 5′-GATCTCATGCAGCTTGCCACGG-3′ and
5′-GATCCCGTGGCAAGCTGCATGA-3′;
D-box, 5′-GATCTTTAGGTAGCCCACGCGACTGGCACGCGCACCTTGCG-3′ and
5′-GATCCGCAAGGTGCGCGTGCCAGTCGCGTGGGCTACCTAAA-3′. The reaction mixture contained, in 9 μl, the binding buffer (10 mM K$^+$ HEPES at pH 7.9, 150 mM KCl, 50

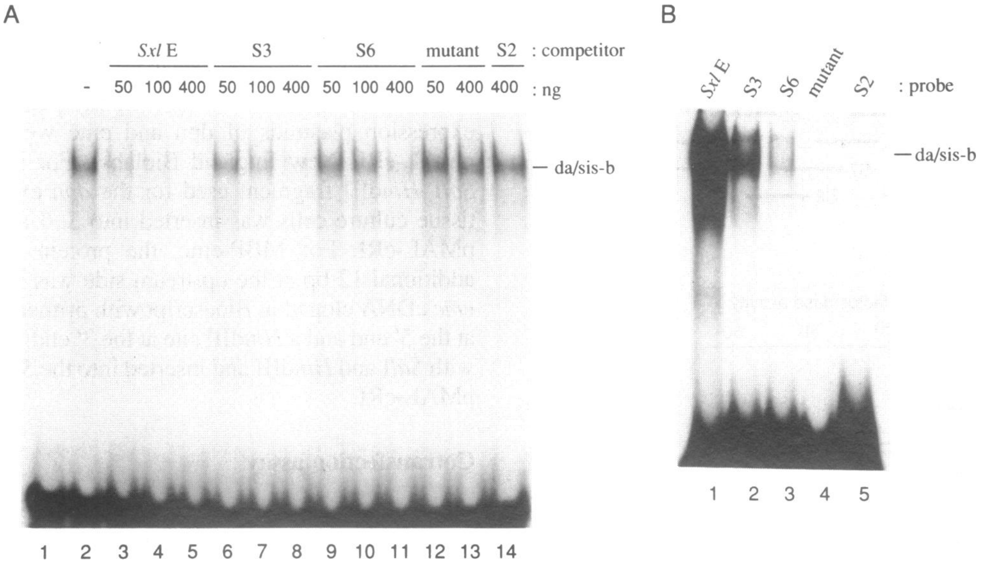

Figure 3. Different binding affinities of the da/sis-b proteins to the responsible elements. (**A**) Competition experiments performed in gel retardation assays. One microgram of da protein and 0.5 μg of sis-b protein are incubated with the ^{32}P-labeled E-box probe in the absence (lane 2) or presence of cold competitor fragments as described above the panel (lanes 3–14). The amount of the competitors is indicated on the top of each lane. Lane 1 contains only the E-box probe. (**B**) Gel retardation assay. One microgram of da protein and 0.5 μg of sis-b protein are incubated with the ^{32}P-labeled probe as indicated at the top of each lane.

mM NaCl, 1 mM EDTA, 1 mM DTT, 10 % glycerol), 400 ng poly dI:dC and 5 μg bovine serum albumin (BSA) in addition to the purified proteins. These mixtures were preincubated for 20 min at 25°C without probe, then 1 μl of ~10 fmol labeled probe was added and incubated for additional 30 min at 25°C. Immediately after incubation, the mixture (9 μl) was electrophoresed on a 4 % polyacrylamide gel as described previously (42). The gel was transferred to Whatman 3MM paper, dried and autoradiographed. In the competition experiments, the competitor oligonucleotides were generated in the same way as the probes except for end-labeling and the amounts indicated in Figure 3B were simultaneously added with the probe to the preincubated reaction mixtures.

DNase I footprinting

Fragments used for DNase I footprinting were obtained by PCR amplification with the primers, 5′-CACGGATCCGCCACCCAAGAAAGTACG-3′ and 5′-CACGAATTCGAAGGAGGCAAGGTGCG-3′, which were annealed to the complementary strand at 180 bp and to the sense strand at 84 bp upstream of the transcription start site of the early *Sxl* promoter respectively. To incorporate radioisotope into only one of the strands, one or the other of these primers was end-labeled with T4 polynucleotides kinase using [γ-^{32}P]ATP. The labeled fragments were incubated with or without dpn protein in reaction mixtures containing 10 mM K$^+$ HEPES at pH 7.9, 150 mM KCl, 50 mM NaCl, 1 mM MgCl$_2$, 1 mM DTT, 10 % glycerol, 400 ng poly dI:dC and 10 μg BSA. After 30 min at 25°C, 5 U DNase I (Takara Co.) was added and incubation continued for additional 2 min. The reactions were terminated by addition of an equal volume of the stop buffer [600 mM Na-acetate at pH 7.0, 40 mM EDTA, 1% SDS] and subsequently incubated for 30 min at 37°C with 80 μg proteinase K (Merck). Nucleic acids were extracted with an equal volume of phenol:chloroform (1:1), ethanol precipitated, dissolved in 80% formamide, 0.01 N NaOH, 1 mM EDTA, with tracking dyes and heated at 90°C for 3 min. The DNA fragments were separated on a 10% polyacrylamide/8 M urea gel. The gel was transferred to Whatman 3MM paper, dried and autoradiographed. The markers partially cleaved at purine residues were prepared according to the standard techniques as described (43).

RESULTS

The *da* and *sis-b* genes cooperatively activate early *Sxl* promoter in *Drosophila* tissue culture cells

Among the four HLH genes involved in regulation of *Sxl* expression, *da* and *sis-b* are required for activation of the early *Sxl* promoter (20,21). To address the action of these gene products on the early *Sxl* promoter, we developed a cotransfection assay system using *Drosophila* tissue culture cells. The transcription start site of the early *Sxl* promoter is located between exons 1 and 2, ~5 kb downstream of that of the late promoter (34), (Fig. 1A, PE and PL, respectively). It has been shown that an ~3 kb region of the early *Sxl* promoter is sufficient to activate transcription in female embryos (20). Based on this knowledge, we first constructed a reporter, SE3.4K, which consists of the 3.4 kb region of the early *Sxl* promoter fused upstream to the lacZ gene (Fig. 1A) and measured β-galactosidase activity of transfected cell lysates. In addition, we constructed *da* and *sis-b* expression plasmids containing the full-length protein coding regions of these genes under the control of the *Drosophila* heat shock protein 70 gene promoter. When we transfected SE3.4K alone, the β-galactosidase activity was very low, corresponding to the basal level (Fig. 1B, SE3.4K). A slight increase in activity was observed when we transfected SE3.4K with either *da* or *sis-b* expression plasmid. However, when SE3.4K was transfected with both *da* and *sis-b*, we observed ~25-fold of activity compared with that of SE3.4K alone (Fig. 1B, SE3.4K). Thus, in the presence of both of the da and sis-b products, the early *Sxl* promoter is efficiently activated in tissue culture cells.

Other sequences in addition to the E box can mediate the da/sis-b effect

In our transfection system, we found that the da/sis-b products could cooperatively activate the early *Sxl* promoter. To define the sequences responsible for this effect, we introduced deletions from the 5′ end of the early *Sxl* promoter. When we deleted to 385 bp upstream of the transcription start site, the level of transcriptional activity in the presence of da/sis-b remained basically unchanged (Fig. 1B, SE385). However, activation by da/sis-b was no longer seen when we deleted up to 180 or 83 bp upstream of the transcription start site (Fig. 1B, SE180 and SE83). These results indicate that the region lying between 180 and 385 bp includes important elements necessary for activation by the da/sis-b products. Examination of this region reveals an E-box sequence motif located ~240 bp upstream of the transcription start site (Fig. 1B and C). To determine whether this E box responds to *da/sis-b* dependent activation, we made a construct, SE385(ΔE), in which the E box was deleted. When SE385(ΔE) was transfected with the *da* and *sis-b* expression plasmids, the level of transcriptional activity was reduced to one-fourth compared with the original SE385 [Fig. 1B, SE385(ΔE)]. In addition, we constructed SE83 + E which contains the E box immediately upstream of the promoter region of SE83. Although SE83 showed no promoter activity even in the presence of the da/sis-b products (Fig. 1B, SE83), transcriptional induction by da/sis-b could be clearly observed in SE83 + E (Fig. 1B, SE83 + E). Such induction was not observed when a mutated E box (Fig. 1C) was fused to SE83 (Fig. 1B, SE83 + mE). These results indicate that the E box can support *da/sis-b* dependent activation of the early *Sxl* promoter. We showed that this E-box sequence is directly recognized by the da/sis-b proteins by means of gel retardation assays. Using the E-box sequence as a probe, we could detect a slowly migrating band in the presence of both da and sis-b proteins (Fig. 3A, lane 2 and B, lane 1). This band could not be detected when the probe containing the mutated E-box sequence was employed (Fig. 3B, lane 4). These results strongly suggest that the da/sis-b products directly activate the early *Sxl* promoter by binding to the E box.

Deletion of the E box from the SE385 construct reduced the promoter activity in the presence of the da/sis-b products, but did not completely abolish the induced transcriptional activation [Fig. 1B, SE385(ΔE)]. This indicates that other elements within the 385 bp promoter region also modulate the da/sis-b interaction. When we fused the region between 385 and 281 bp upstream of the transcription start site to the promoter region of SE83, which lacks the E-box sequence, partial recovery of da/sis-b induced transcriptional activation is seen (Fig. 1B, SE83 + Q). To further define the critical elements, we divided the region between 385 and 281 bp into 11 fragments (Fig. 2A) and examined which region was responsible for the da/sis-b effect. For this assay, we inserted three copies of each fragment immediately upstream of the promoter region of SE83. Only S3 and S6 were found to have the ability to confer transcriptional activation in the presence of the da/sis-b products (Fig. 2B). However, the degree of activity induced by three copies of either S3 or S6 is comparable to that by only one copy of the E box (compare S3 and S6 in Fig. 2B with SE83 + E in Fig. 1B), suggesting that the potential of these two fragments is inferior to that of the E box. The exact sequences responsible for the da/sis-b effect are yet to be resolved, but analyses of both fragments reveal E-box-like motifs (see Discussion). To verify the ability of the da/sis-b proteins to bind to these two fragments, we performed *in vitro* competition experiments using gel retardation assays and found that when the E box probe was used, the da/sis-b binding complex could be progressively competed out by increasing amounts of the S3 fragment (Fig. 3A, lanes 6–8). However, the E-box fragment itself competed out the complex more efficiently (lanes 3–5). The S6 fragment was also able to compete, but with less efficiency than the S3 fragment (lanes 9–11). Neither the mutated E–box nor the S2 fragment had the ability to compete out the bound da/sis-b–E-box complex (Fig. 3A, lanes 12–14) or to mediate da/sis-b transactivation (Figs 1B and 2B, respectively). Furthermore, gel retardation experiments showed that da/sis-b proteins could bind to both the S3 and S6 fragments, although with lesser efficiencies than to the E box (Fig. 3B, lanes 1–3). Control experiments showed that da/sis-b could not bind to either the mutated E-box or the S2 fragment (lanes 4 and 5). Thus, in addition to the E box, we found that two other elements are directly recognized by the da/sis-b proteins, albeit with weaker affinities. These results indicate that the ability of the da/sis-b product to activate the promoter may be regulated by its different affinities for each of these elements.

Dpn product acts as a repressor of the early *Sxl* promoter by direct binding to a specific site

Genetic and molecular studies have suggested that *dpn* represses the *Sxl* expression in males (18). We verified the possibility that this dpn repression occurs at the early *Sxl* promoter by means of transfection assays. We found that when we transfected SE3.4K (schematically represented in Fig. 1A) with increasing amounts of *dpn* expression plasmid, transcriptional activation was severely reduced even in the presence of the da/sis-b products (Fig. 4A). Thus, dpn efficiently represses *da/sis-b* dependent activation of the early *Sxl* promoter in tissue culture cells.

It has been shown that the emc protein titrates the binding of the da/sis-b heterodimer to the E-box sequence in the promoter of the *acute (ac)* gene, one of the proneural genes (7,37). Figure 5 shows that emc protein can also titrate the binding of da/sis-b proteins to the E box of the early *Sxl* promoter. In gel retardation assays using the E box of the early *Sxl* promoter as a probe, the addition of increasing amounts of emc protein progressively diminishes the band corresponding to the da/sis-b complex (Fig. 5, lanes 3–6). On the contrary, equivalent amounts of dpn protein and MBP could not titrate the da/sis-b complex (lanes 7–10). Furthermore, in gel retardation assays we found that the dpn protein could not bind to the E-box probe (Fig. 6C, lanes 1 and 2). Thus, although dpn could repress *da/sis-b* dependent activation of the early *Sxl* promoter, it was not able to inhibit binding of da/sis-b proteins to the E box by titrating or competing for their binding sites. Consistent with these properties, dpn could not repress transcription from SE83 + E, which contains only the 83 bp promoter region and the E box. When we transfected SE83 + E with *dpn* expression plasmid in the presence of both *da* and *sis-b* expression plasmids, the level of activation was not inhibited, even with large amounts of *dpn* expression plasmid (Fig. 4B, SE83 + E). These findings suggest that the dpn product requires

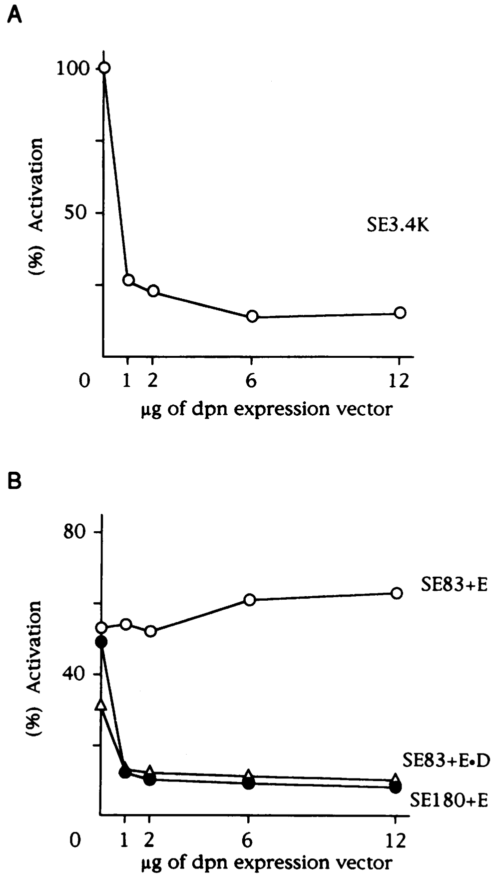

Figure 4. Repression of *da/sis-b* dependent activation of the early *Sxl* promoter by dpn product. (**A**) Three micrograms of SE3.4K (represented in Fig. 1) are cotransfected into *Drosophila* culture cells with 2 µg each of the *da* and *sis-b* expression plasmids and the indicated amount of the *dpn* expression plasmid. In addition, 2 µg of the luciferase gene expression vector is added as control. The β-galactosidase activities of SE3.4K are normalized to luciferase activity to control for variation in transfection efficiency and plotted against the amount of the cotransfected *dpn* expression plasmid. The values of the activities are shown as percentages of the β-galactosidase activity obtained by transfection of SE3.4K with both of the *da* and *sis-b* expression plasmids but without the *dpn* expression plasmid. (**B**) SE83 + E (open circle), SE180 + E (filled circle) or SE83 + E•D (open triangle) is transfected and their β-galactosidase activities are plotted as described in (A).

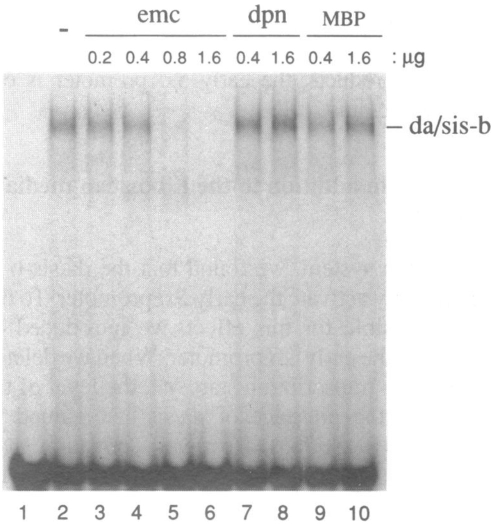

Figure 5. Binding of the da/sis-b complex to the E box is inhibited by emc, but not by dpn. One microgram of da protein and 0.5 µg of sis-b protein are incubated with the ^{32}P-labeled E-box probe in the absence (lane 2) or presence of various amounts of emc protein (lanes 3–6), dpn protein (lanes 7 and 8) or MBP (lanes 9 and 10). The amount of the protein in each lane is indicated above the panel. Lane 1 contains only E-box probe.

some sequences in addition to the E box to repress *da/sis-b* dependent activation. When we used SE180 + E which contains the 180 bp region and the E box of the early *Sxl* promoter as the reporter construct, transcription was efficiently repressed by dpn (Fig. 4B, SE180 + E). This observation indicates that the region between 180 and 84 bp upstream of the start site, present in SE180 + E but not in SE83 + E, contains the sequence which is necessary for dpn-mediated repression. To confirm this possibility, we performed DNase I footprinting on this region. Figure 6 shows that widely spread regions are clearly protected in both sense and complementary strands, depending on dpn protein (Fig. 6B). These protected regions are located between 122 and 93 bp upstream of the transcription start site (Fig. 6A). The binding of dpn protein to this region could also be detected by gel retardation assays. When the fragment containing the promoter region between 125 and 91 bp, designated D box, was used as a probe, a slowly migrating band was observed depending on dpn protein (Fig. 6C, lanes 3 and 4). However, no binding complex was observed with E-box probe (lanes 1 and 2). Thus, dpn protein has the ability to bind to a specific site(s) in the early *Sxl* promoter. To examine whether this dpn binding site is actually involved in repression, we made a construct, SE83 + E•D, in which the dpn binding site was inserted downstream of the E box of SE83 + E. Cotransfection of SE83 + E•D with *dpn*, in the presence of *da* and *sis-b*, results in efficient repression of *da/sis-b* mediated activation (Fig. 4B, SE83 + E•D). These results indicate that dpn represses *da/sis-b* dependent activation directly by site-specific binding to the early *Sxl* promoter.

DISCUSSION

Genetic and molecular studies have identified the genes involved in the X:A signal and suggested that these genes regulate the early *Sxl* promoter. In this paper, by use of a combination of transfection assays and *in vitro* binding experiments, we were able to obtain new insight into the action of the genes encoding HLH proteins on the regulation of the early *Sxl* promoter.

Regulation of the early *Sxl* promoter by positive and negative HLH proteins

Among the genes involved in the X:A signal, *da* and *sis-b* encode typical basic HLH proteins. It has been shown that, during neurogenesis, these proteins form heterodimers (40) and cooperatively activate the *ac* promoter by binding to the E-box sequences (7). Here, we show that the da/sis-b products directly activate the early *Sxl* promoter not only by binding to the E box, but also to other lower affinity sites. On the other hand, *dpn* encodes an aberrant HLH protein with a proline residue in the basic region (29) and has been shown to repress *Sxl* expression in males (18). Consistent with this observation, we showed that the

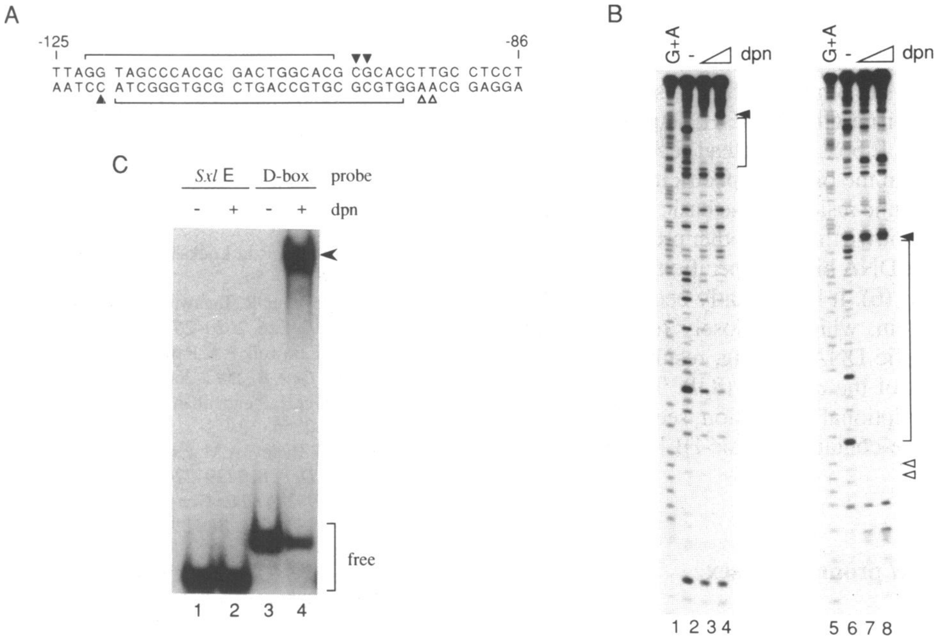

Figure 6. Binding of dpn protein to a specific site on the early *Sxl* promoter. (**A**) Schematic representation of the region protected from DNase I digestion by the binding of dpn protein. The bars and open triangles indicate the protected regions and sites. The filled triangles indicate the hypersensitive site to DNase I digestion. (**B**) DNase I footprinting experiment. Fragments containing the promoter region between 180 and 84 bp upstream of the transcription start site are end-labeled with [γ-^{32}P]ATP either on the sense strand (lanes 2–4) or the complementary strand (lanes 6–8). The fragments are incubated in the absence (lanes 2 and 6) or presence of 4 μg (lanes 3 and 7) or 8 μg (lanes 4 and 8) of dpn protein and subsequently digested with DNase I. G + A represents the marker partially cleaved at purine residues of fragments labeled on the sense strand (lane 1) or the complementary strand (lane 5). The bars and open triangles indicate the protected regions and sites from DNase I digestion, respectively. The filled triangles indicate the hypersensitive site of DNase I digestion. (**C**) Gel retardation experiment. The ^{32}P-labeled E box (*Sxl* E, lanes 1 and 2) or the fragment containing the promoter region between 125 and 91 bp (D-box, lanes 3 and 4) is incubated with (lanes 2 and 4) or without (lanes 1 and 3) 0.5 μg of dpn protein. The arrow head represents the dpn–D-box complex.

dpn product represses the *da/sis-b* dependent activation of the early *Sxl* promoter. Furthermore, this mechanism of repression differs from that of the emc protein, which has been shown to titrate the da/sis-b heterodimer (7,37). We showed that the dpn product acts as a direct repressor that binds to a specific site. Thus, these positive- and negative-acting HLH proteins regulate the activity of the early *Sxl* promoter.

Activation dependent on da/sis-b products

Recently, Estes and colleagues showed that the 1.4 kb promoter region is required for full activation of the early *Sxl* promoter in female embryos (21). In female embryos, not only the da/sis-b products but also other activators including sis-a and run products are necessary for activation of the early *Sxl* promoter. However, it was suggested that an element located within the 400 bp promoter region plays an essential role for female-specific activation. Consistent with this notion, activation mediated by the 400 bp promoter region is severely impaired by *sis-b* mutation. Our transfection assays showed the ability of the da/sis-b products to activate transcription from the 385 bp region of the early *Sxl* promoter as efficiently as the 3.4 kb promoter region. Thus, these observations point to the importance of elements located within the 385 bp promoter region for the *da/sis-b* mediated transcriptional activation. Analysis of the nucleotide sequence revealed only one E box located within this region. We showed that this E box is in fact important for *da/sis-b* dependent activation of the promoter in tissue culture cells, and that the da/sis-b protein complex does bind efficiently to the E box *in vitro*. This supports the idea that the da/sis-b heterodimers act as transcriptional activators by direct binding to the E box. This was also suggested recently in the case of the *ac* gene promoter (7).

An interesting observation is that despite the absence of the exact E-box sequence, two other subregions can contribute to the *da/sis-b* dependent activation to some extent. These two regions encode E-box related sequences (G<u>CA</u>GCT<u>TG</u>C), which differ from the E-box motif (CANNTG) by an insertion in the internal spacer. Since the CANNNTG motif can stand as a minor binding site for other HLH proteins (44), the da/sis-b heterodimers may be able to interact with this sequence albeit with low efficiency. Alternatively, other factors may be involved in more efficient recognition of this sequence *in vivo* as suggested by Thayer and Weintraub, who have shown that the DNA-binding affinity of MyoD/E47 heterodimers is stimulated by a cellular factor (45). At present, the molecular mechanisms underlying this observation are not fully elucidated. However, our finding may provide some clues to better understanding of the dosage effect of X chromosomes in controlling the *Sxl* promoter activity.

Dpn product as a repressor

Our results indicated that like emc, the dpn protein can repress *da/sis-b* dependent activation of the promoter, although the molecular mechanism of this repression is very different. The emc repression occurs by inhibition of the da/sis-b heterodimer, most likely by competitive heterodimeric interaction with either da or sis-b protein (37). The absence of the basic DNA binding region in emc protein may render the resulting heterodimer transcriptionally non-functional. Dpn protein, on the other hand, has been shown in our assays to bind to DNA in a sequence-specific manner which differs from other basic-HLH proteins. Dpn protein does not bind to the E box, but instead recognizes other sequences. This difference in the binding specificity is probably due to the inclusion of a proline residue in the basic region of dpn protein. Indeed it has been shown that with other proline-containing basic-HLH proteins, the DNA binding specificity differs from typical basic-HLH proteins (6). It has recently been shown that the *Drosophila* hairy protein, which is closely related to dpn protein, also exhibits specific DNA binding, resulting in direct repression of transcription of the *ac* gene (8,9). Thus, specific DNA binding and transcriptional repression seem to be the general properties of proline-containing basic-HLH proteins.

Regulation of the early *Sxl* promoter in sex determination

During sex determination in flies, the X:A ratio controls the activity of the early *Sxl* promoter. This promoter is active in female embryos but remains inactive in males. How are these different states achieved between sexes? Although the HLH proteins involved in sex determination have not been well quantitated in embryos, the maternal factors, da and emc, and the zygotic autosome-linked dpn appear to be equally present in both sexes. In contrast, the *sis-b* gene is located in two X chromosomes in females, but only one in males. Thus, the amount of sis-b product would be different between females and males. This difference in the amount of sis-b protein may result in different levels of active da/sis-b heterodimeric complexes. The repression by emc and dpn may further modulate *da/sis-b* mediated activation of the early *Sxl* promoter, especially in male embryos which have lower concentrations of sis-b product. As a consequence, sufficient levels of active da/sis-b complex necessary for promoter activation may be present only in females. This does not rule out the possibility of other gene products (i.e. *sis-a* and *run*) to interact with these HLH proteins for more precise regulation of sex-specific determination. Further analyses using cell lines in combination with transgenic flies will undoubtedly reveal more precise mechanisms of early *Sxl* promoter regulation during sex determination.

ACKNOWLEDGEMENTS

We are grateful to T. Uemura for providing a *Drosophila* genomic library, C. S. Thummel for pCaSpeR-AUG-βgal, C. V. Cabrera for pC31/B11, E. Bier for dpn cDNA plasmid, R. Ueda and J. Modolell for emc cDNA plasmid. We thank S. Kobayashi, R. Yu and N. Kataoka for discussion and critical reading of the manuscript. The genomic sequence of the early *Sxl* promoter region described in this paper will appear in the GSDB, DDBJ, EMBL and NCBI nucleotide sequence databases with the following accession no. D50435. This work was supported by Grants-in-Aid from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1 Murre, C. and Baltimore, D. (1991) In McKnight, S.L. and Yamamoto, K.R. (eds) *Transcriptional Regulation*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. II, pp. 861–879

2 Jan, Y.N. and Jan, L.Y. (1993) *Cell*, **75**, 827–830.

3 Murre, C.,McCaw, P.S. and Baltimore, D. (1989) *Cell*, **56**, 777–783.

4 Davis, R.L.,Cheng, P.,Lassar, A.B. and Weintraub, H. (1990) *Cell*, **60**, 733–746.

5 Benezra, R.,Davis, R.L.,Lockshon, D.,Turner, D.L. and Weintraub, H. (1990) *Cell*, **61**, 49–59.

6 Sasai, Y.,Kageyama, R.,Tagawa, Y.,Shigemoto, R. and Nakanishi, S. (1992) *Genes Dev.*, **6**, 2620–2634.

7 Van Doren, M.,Powell, P.A.,Pasternak, D.,Singson, A. and Posakony, J.W. (1992) *Genes Dev.*, **6**, 2592–2605.

8 Ohsako, S.,Hyer, J.,Panganiban, G.,Oliver, I. and Caudy, M. (1994) *Genes Dev.*, **8**, 2743–2755.

9 Van Doren, M.,Bailey, A.M.,Esnayra, J.,Ede, K. and Posakony, J.W. (1994) *Genes Dev.*, **8**, 2729–2742.

10 Cline, T.W. (1993) *Trends Genet.*, **9**, 385–390.

11 Cline, T.W. (1983) *Dev. Biol.*, **95**, 260–274.

12 Sanchez, L. and Nöthiger, R. (1983) *EMBO J.*, **2**, 485–491.

13 Cline, T.W. (1986) *Genetics*, **113**, 641–663.

14 Cline, T.W. (1988) *Genetics*, **119**, 829–862.

15 Torres, M. and Sanchez, L. (1989) *EMBO J.*, **8**, 3079–3086.

16 Duffy, J.B. and Gergen, J.P. (1991) *Genes Dev.*, **5**, 2176–2187.

17 Torres, M. and Sánchez, L. (1992) *Genet. Res.*, **59**, 189–198.

18 Younger-Shepherd, S.,Vaessin, H.,Bier, E.,Jan, L.Y. and Jan, Y.N. (1992) *Cell*, **70**, 911–922.

19 Bopp, D.,Bell, L.R.,Cline, T.W. and Schedl, P. (1991) *Genes Dev.*, **5**, 403–415.

20 Keyes, L.N.,Cline, T.W. and Schedl, P. (1992) *Cell*, **68**, 933–943.

21 Estes, P.A.,Keyes, L.N. and Schedl, P. (1995) *Mol. Cell. Biol.*, **15**, 904–917.

22 Bell, L.R.,Horabin, J.I.,Schedl, P. and Cline, T.W. (1991) *Cell*, **65**, 229–240.

23 Sakamoto, H.,Inoue, K.,Higuchi, I.,Ono, Y. and Shimura, Y. (1992) *Nucleic Acids Res.*, **20**, 5533–5540.

24 Horabin, J.I. and Schedl, P. (1993) *Mol. Cell. Biol.*, **13**, 1408–1414.

25 Erickson, J.W. and Cline, T.W. (1993) *Genes Dev.*, **7**, 1688–1702.

26 Villares, R. and Cabrera, C.V. (1987) *Cell*, **50**, 415–424.

27 Caudy, M.,Vässin, H.,Brand, M.,Tuma, R.,Jan, J.Y. and Jan, Y.N. (1988) *Cell*, **55**, 1061–1067.

28 Cronmiller, C.,Schedl, P. and Cline, T.W. (1988) *Genes Dev.*, **2**, 1666–1676.

29 Bier, E.,Vaessin, H.,Younger-Shepherd, S.,Jan, L.Y. and Jan, Y.N. (1992) *Genes Dev.*, **6**, 2137–2151.

30 Ellis, H.M.,Spann, D.R. and Posakony, J.W. (1990) *Cell*, **61**, 27–38.

31 Garrell, J. and Modolell, J. (1990) *Cell*, **61**, 39–48.

32 Kagoshima, H.,Shigesada, K.,Satake, M.,Ito, Y.,Miyoshi, H.,Ohki, M.,Pepling, M. and Gergen, P. (1993) *Trends Genet.*, **9**, 338–340.

33 Cronmiller, C. and Cummings, C. (1993) *Mech. Dev.*, **42**, 159–169.

34 Erickson, J.W. and Cline, T.W. (1991) *Science*, **251**, 1071–1074.

35 Cabrera, C.V.,Martinez-Arias, A. and Bate, M. (1987) *Cell*, **50**, 425–433.

36 Cline, T.W. (1989) *Cell*, **59**, 231–234.

37 Van Doren, M.,Ellis, H.M. and Posakony, J.W. (1991) *Development*, **113**, 245–255.

38 Thummel, C.S.,Boulet, A.M. and Lipshitz, H.D. (1988) *Gene*, **74**, 445–456.

39 Inoue, K.,Hoshijima, K.,Sakamoto, H. and Shimura, Y. (1990) *Nature*, **344**, 461–463.

40 Cabrera, C.V. and Alonso, M.C. (1991) *EMBO J.*, **10**, 2965–2973.

41 Edlund, T.,Walker, M.D.,Barr, P.J. and Rutter, W.J. (1985) *Science*, **230**, 912–916.

42 Matsuo, I.,Katamura, M.,Okazaki, K. and Yasuda, K. (1991) *Development*, **113**, 539–550.

43 Sambrook, J.,Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.*Cold Spring Harbar Laboratory Press, Cold Spring Harbor, NY.

44 Sun, X. and Baltimore, D. (1991) *Cell*, **64**, 459–470.

45 Thayer, M.J. and Weintraub, H. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6483–6487.