MOLECULAR AND CELLULAR BIOLOGY, May 1993, p. 2802–2814
0270-7306/93/052802-13$02.00/0

Vol. 13, No. 5

$(CT)_n \cdot (GA)_n$ Repeats and Heat Shock Elements Have Distinct Roles in Chromatin Structure and Transcriptional Activation of the *Drosophila hsp26* Gene

QIN LU,[1] LORI L. WALLRATH,[1] HOWARD GRANOK,[1,2] AND SARAH C. R. ELGIN[1]*

Department of Biology[1] and Program in Molecular Genetics,[2] Division of Biology and Biomedical Sciences, Washington University, St. Louis, Missouri 63130

Received 27 October 1992/Returned for modification 14 December 1992/Accepted 1 February 1993

Previous analysis of the *hsp26* gene of *Drosophila melanogaster* has shown that in addition to the TATA box and the proximal and distal heat shock elements (HSEs) (centered at −59 and −340, relative to the start site of transcription), a segment of $(CT)_n$ repeats at −135 to −85 is required for full heat shock inducibility (R. L. Glaser, G. H. Thomas, E. S. Siegfried, S. C. R. Elgin, and J. T. Lis, J. Mol. Biol. 211:751–761, 1990). This $(CT)_n$ element appears to contribute to formation of the wild-type chromatin structure of *hsp26*, an organized nucleosome array that leaves the HSEs in nucleosome-free, DNase I-hypersensitive (DH) sites (Q. Lu, L. L. Wallrath, B. D. Allan, R. L. Glaser, J. T. Lis, and S. C. R. Elgin, J. Mol. Biol. 225:985–998, 1992). Inspection of the sequences upstream of *hsp26* has revealed an additional $(CT)_n$ element at −347 to −341, adjacent to the distal HSE. We have analyzed the contribution of this distal $(CT)_n$ element (−347 to −341), the proximal $(CT)_n$ element (−135 to −85), and the two HSEs both to the formation of the chromatin structure and to heat shock inducibility. *hsp26* constructs containing site-directed mutations, deletions, substitutions, or rearrangements of these sequence elements have been fused in frame to the *Escherichia coli lacZ* gene and reintroduced into the *D. melanogaster* genome by P-element-mediated germ line transformation. Chromatin structure of the transgenes was analyzed (prior to gene activation) by DNase I or restriction enzyme treatment of isolated nuclei, and heat-inducible expression was monitored by measuring β-galactosidase activity. The results indicate that mutations, deletions, or substitutions of either the distal or the proximal $(CT)_n$ element affect the chromatin structure and heat-inducible expression of the transgenes. These $(CT)_n$ repeats are associated with a nonhistone protein(s) in vivo and are bound by a purified *Drosophila* protein, the GAGA factor, in vitro. In contrast, the HSEs are required for heat-inducible expression but play only a minor role in establishing the chromatin structure of the transgenes. Previous analysis indicates that prior to heat shock, these HSEs appear to be free of protein. Our results suggest that GAGA factor, an abundant protein factor required for normal expression of many *Drosophila* genes, and heat shock factor, a specific transcription factor activated upon heat shock, play distinct roles in gene regulation: the GAGA factor establishes and/or maintains the DH sites prior to heat shock induction, while the activated heat shock factor recognizes and binds HSEs located within the DH sites to trigger transcription.

Eukaryotic gene expression in vivo is regulated at the level of the chromatin template, in which the DNA is packaged with histones and nonhistone chromosomal proteins. To be active, a gene must lie in an open domain, with its regulatory sequences and promoter/transcription start site in an accessible, nonnucleosomal conformation (reviewed in reference 67). Such a nonnucleosomal site can be detected as a DNase I-hypersensitive (DH) site, a region that is hypersensitive to a number of nucleases (16, 26). Over the past few years, extensive analyses of protein-DNA interactions in the promoter regions of many inducible genes have suggested that these genes may be characterized as one of two types, a remodeling gene or a preset gene. In remodeling genes, nucleosomes block some regulatory sequences (such as the TATA box or other protein binding sites) in the inactive state; remodeling (displacement or alteration of the nucleosomes) occurs during gene activation, allowing regulatory factors to gain access to these binding sites. This change in accessibility can be detected as the formation of DH sites. Examples of promoters that undergo remodeling are that of the yeast *PHO5* gene (1, 17, 27, 62) and that of the mouse mammary tumor virus long terminal repeats (3, 11, 46, 49). In contrast, the regulatory sequences of preset genes are free of nucleosomes and are maintained in an accessible conformation while the gene is inactive. Following the activation stimulus, *trans*-acting factors bind to the regulatory sequences and trigger transcription. Thus, in preset genes, there are no requisite changes in the nucleosomal pattern (with concomitant changes in accessibility of the regulatory sequences) before and after gene activation. Such genes presumably must be either packaged into their precise chromatin structure immediately following replication or remodeled shortly thereafter. The *hsp26* gene of *Drosophila melanogaster* provides an excellent example of a preset gene (6, 65).

The promoter region of the *Drosophila hsp26* gene contains well-characterized regulatory elements known to be involved in the control of gene expression. These regulatory elements include the transcription start site, TATA box, heat shock elements (HSEs), and the $(CT)_n \cdot (GA)_n$ repeats (Fig. 1). In the region upstream of the transcription start site (to −1917), there are seven putative HSEs which match, to different extents, the heat shock consensus sequence (4, 39).

* Corresponding author.

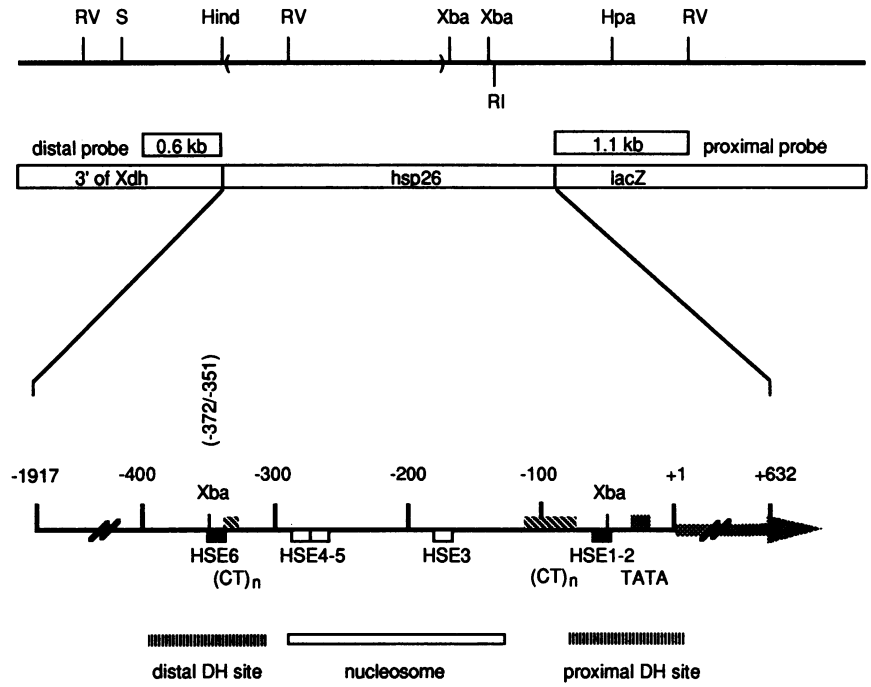

FIG. 1. Map of the *hsp26-lacZ* construct and the probes used for chromatin structure analyses. The structure of CarX is shown. *hsp26* sequences from −1917 to +632 (with the exception of sequences from −371 to −352, which are deleted) are fused in frame to the *E. coli lacZ* gene. Restriction sites shown on the top line are those sites giving marker fragments or those used for mapping the chromatin structure in indirect end-labeling experiments. Sequences located between the parentheses on the top line are not present in transgenes containing cP-351, cPΔCT · GA, and cPri (22). The 1.1-kb *lacZ* fragment, isolated from pMC1403 (7), is used as a probe in experiments to map the DH sites and to quantitate the proximal DH site. The 0.6-kb fragment, containing 3′ sequences of the *Drosophila xdh* gene, is used as a probe for quantitating the distal DH site. The partial restriction map of *hsp26* sequences (−1917 to +632) is enlarged below, with the $(CT)_n$ regions (striped boxes), the TATA box (stippled box), and two required HSEs (HSE 1-2 and HSE 6; filled boxes) diagrammed. HSEs 3 to 5 (shown as open boxes) are sequence elements of lower homology to the HSE consensus and are not involved in heat shock induction (22, 45, 58, 65). Chromatin structural features of the *hsp26* gene are marked below. RI, *Eco*RI; RV, *Eco*RV; Hind, *Hin*dIII; Hpa, *Hpa*I; S, *Sma*I; Xba, *Xba*I.

Functional analyses using P-element-mediated transformation have revealed that sequences from −351 downstream are sufficient to confer wild-type heat shock inducibility; only those HSEs showing a perfect match to two or more contiguous *n*GAA*n* elements are required for the full heat shock response (HSE 1-2 and HSE 6; Fig. 1) (9, 45, 58). High-resolution genomic footprinting analysis indicates that the regions containing HSEs 1-2 and 6-7, which lie in DH sites, bind heat shock factor (HSF) after, but not before, heat shock; HSEs 3, 4, and 5 appear to be associated with a precisely positioned nucleosome and show no indications of HSF binding (65).

Optimal heat shock induction of *hsp26* also requires a stretch of alternating C and T residues located between −135 and −85, adjacent to the proximal HSE. Deletion or substitution of this region results in a four- to fivefold reduction in heat shock-induced expression (22). Chromatin structure analyses indicate that the proximal $(CT)_n$ element is essential for the formation of appropriate DH sites at this locus (41). A *Drosophila* nuclear protein of 66 kDa which binds specifically to segments of alternating C and T residues, including the proximal $(CT)_n$ element in the promoter region of *hsp26*, has been identified (20). Further analysis reported here indicates that this protein is the GAGA factor first isolated by Biggin and Tjian (5).

The mechanism by which the preset structure is established is unknown. Here we investigate which upstream DNA sequence elements are necessary and sufficient to establish the normal chromatin structure of *hsp26* and to facilitate a normal heat shock response. We extend our functional analysis of the *hsp26* gene by using P-element constructs containing deletions and/or site-specific alterations affecting the two $(CT)_n$ elements or the two HSEs within the promoter region of the *hsp26* gene. These constructs have been introduced into the *Drosophila* genome by P-element transformation. Our results suggest that GAGA factor and HSF play distinct roles in gene regulation: the GAGA factor plays a crucial role in establishing and maintaining the chromatin structure prior to heat shock induction, while the HSF is involved in transcription activation upon heat shock. The results also show that the upstream regulatory elements function in a symmetrical manner, supporting a structural model in which the distal HSE/$(CT)_n$ unit is brought close to the proximal HSE/$(CT)_n$ unit by winding the sequences between the two units on a specifically positioned nucleosome. This stable DNA looping could facilitate the interaction of transcription factors bound on these elements with each other and/or with the transcription complex.

MATERIALS AND METHODS

DNA constructs. The numerical assignment of nucleotides and the transcription start site of the *Drosophila hsp26* gene follow the convention of Ingolia and Craig (29). Plasmids cP-351, cPΔCT · GA, and cPri have been described previously (22). The majority of the constructs described in this study were derived from a parental construct, Car26ZΔX.

To make Car26ZΔX, a 5.6-kb *Sal*I fragment, which contains *hsp26* sequences from −1917 to +632 fused in frame to the *Escherichia coli lacZ* gene, was isolated from plasmid pMC1871.26 (23) and cloned into the *Sal*I site of Car20T (designated c70T1 by Xiao and Lis [69]). The DNA of the resulting plasmid (Car26ZT) was digested with *Xba*I and religated. A clone containing a deletion of the two *Xba*I fragments at the *hsp26* promoter region (−372 to −352 and −351 to −52) was recovered and designated Car26ZΔX. Insertion of the 299-bp *Xba*I fragment (−351 to −52) into the *Xba*I site of Car26ZΔX in either orientation gave constructs CarX and CarFX. In CarX, the nucleotide sequence at the −372/−351 junction is 5′-GCGCTCTTT/CTAGAAACTTCG GCTCTCTCTCACTC-3′ (where / indicates the junction). In CarFX, the nucleotide sequence at the −372/−52 junction is 5′-GCGCTCTTT/CTAGAAGAGTCC-3′; the nucleotide sequence at the −352/−51 junction is 5′-GAGAGAGCCGA AGTTT/CTAGAAAA-3′. To make CarX^5, a 304-bp *Xba*I fragment, in which two nucleotides (5′-TA-3′; −87 to −86) have been replaced by a 7-bp *Xho*I linker (5′-CCTCGAG-3′), was isolated from plasmid 88B-X (57). This 304-bp *Xba*I fragment from 88B-X was used to replace the *Xba*I fragment in CarX to give CarX^5, which contains an additional 5 bp in the promoter region.

In plasmid CarXΔ6, HSE 6 was eliminated by deleting sequences from −347 to −341. The 299-bp *Xba*I fragment was first cloned into M13mp18, forming M13mp18X. Oligonucleotide 5′-GGTCGACTCTAGAGGCTCTCTCACTC-3′ was used to make a 6-bp (−346 to −341) deletion in M13mp18X, using oligonucleotide-directed mutagenesis as described by Kunkel et al. (36), resulting in M13mp18XΔ6. The *Xba*I fragment containing the deletion was subsequently isolated from the replicative form of M13mp18XΔ6 and cloned into the *Xba*I site of Car26ZΔX. The nucleotide sequence around the deletion junction of CarXΔ6 is 5′-TTCTAGA/GGCTCTCTCACTC-3′. Similarly, site-directed mutagenesis with oligonucleotide 5′-TAGAAACTTCGGCT CGAGCACTCATACAGGCGC-3′, using M13mp18X as a template, results in M13mp18XmCTd, in which the sequence 5′-TCT-3′ (−335 to −333) is changed to 5′-GAG-3′, therefore changing the distal $(CT)_n$ sequence from 5′-CTCT CTC-3′ (−338 to −332) to 5′-CTCGAGC-3′. The *Xba*I fragment containing this mutation was isolated and cloned into the *Xba*I site of Car26ZΔX, giving CarXmCTd. To make CarXΔCTmCTd, in which the proximal $(CT)_n$ sequence is deleted and the distal $(CT)_n$ sequence is mutated, a 250-bp *Xba*I fragment in which sequences from −135 to −85 have been replaced by 5′-GGTCGAC-3′ was isolated from plasmid 88B-41.1 (57) and cloned into the *Xba*I site of M13mp18, resulting in M13mp18XΔCT. Mutagenesis with oligonucleotide 5′-TAGAAACTTCGGCTCGAGCACTCATACAGGC GC-3′, using M13mp18XΔCT as a template, creates M13mp18XΔCT/mCTd. The *Xba*I fragment that contains the proximal $(CT)_n$ deletion and the distal $(CT)_n$ mutation was isolated and cloned into the *Xba*I site of Car26ZΔX. Note that the resulting construct, CarXΔCT/mCTd, has the same nucleotide changes in the distal $(CT)_n$ region as does CarXmCTd.

Plasmid CarXΔ2, in which sequences containing the proximal HSE (HSE 1-2) are deleted, was made as follows. A 304-bp *Xba*I fragment (−351 to −52), in which an *Xho*I linker (5′-CCTCGA-3′) has replaced nucleotide 5′-TA-3′ (−87 to −86), was isolated from plasmid 88B-X (57) and cloned into the *Xba*I site of M13mp18. A clone which places the −351-containing end of the *Xba*I fragment adjacent to the *Sal*I site in the polylinker of M13mp18 was selected and designated M13mp18X(Xho). A 270-bp *Xho*I-*Sal*I fragment which contains *hsp26* sequences from −351 to −85 was isolated from M13mp18X(Xho) and cloned into the *Sal*I site of M13mp18. A clone which places the −351-containing end of the fragment adjacent to the *Xba*I site in the polylinker of M13mp18 was selected and designated M13mp18XΔ2, in which *hsp26* sequences from −87 to −54 have been replaced by a linker (5′-CCTCGAG-3′). The nucleotide sequence around the deletion junction is 5′-CTCTC/CCTCGAG/TCT AGAAAAGCTCCA-3′ (the linker sequence is underlined). A 280-bp *Xba*I fragment containing this deletion was isolated and cloned into the *Xba*I site of Car26ZΔX, giving CarXΔ2.

To make CarX-70Z, plasmid DNA of pX^{u}S26Z (23) was partially digested with *Xba*I, and the 5′ overhangs of the products were filled in with Klenow fragment in the presence of deoxynucleoside triphosphates (54). A 3.4-kb *Xba*I (filled)-*Sal*I fragment containing the origin of replication and Amp^r gene of pX^{u}S26Z and *hsp26* sequences from −351 to −52 was isolated. A 3.2-kb *Nru*I-*Sal*I fragment containing an *hsp70* promoter (sequences from −50 to +260) fused to the *lacZ* gene was isolated from p70ZT (21). The two isolated fragments were ligated to form pΔZX(−351/−52)70Z, in which *hsp26* sequences from −351 to −52 were fused to *hsp70* sequences at −50. A 3.7-kb *Xho*I-*Sal*I fragment containing *hsp26*, *hsp70*, and *lacZ* sequences was isolated from pΔZX(−351/−52)70Z and inserted into the *Sal*I site of Car20T, forming CarX-70Z. The sequence at the 5′ junction is 5′-ccctcgaggggatcc/TCTAGAAACTTCGG-3′ (lowercase letters represent the linker sequence in Car20T). Plasmid CarXΔCT-70Z was made by the strategy used to make CarX-70Z except that the $(CT)_n$ sequence, located between −135 and −85 of *hsp26*, is replaced by a linker (5′-CCTC-GAG-3′).

***Drosophila* germ line transformation and CPRG assays.** Constructs cP-351, cPΔCT · GA, and cPri have been introduced into the *Drosophila* germ line previously (22). All other *hsp26-lacZ* and *hsp26-hsp70-lacZ* constructs were introduced into the *Drosophila* germ line by P-element-mediated transformation (53, 60), using *ry*^{506} as the host stock. Transformants were identified by using the eye color marker; those containing independent single insertions of the P-element transgene were identified by Southern blot analysis. The integrity of the transgenes was confirmed by genomic restriction mapping using the 1.1-kb *lacZ* sequence (Fig. 1) as a probe (data not shown).

Expression of *hsp26-lacZ* transgenes was assessed by determining levels of β-galactosidase activity, using chlorophenol red/β-D-galactopyranoside (CPRG) as a substrate for β-galactosidase (58). Individual males of each line were crossed to the host stock *ry*^{506}; *ry*$^+$ female progeny, which were heterozygous for the P-element insertion, were used for CPRG assays. In total, 30 females between 3 and 5 days old were selected and distributed equally among six Eppendorf tubes (with holes in the top for air). Three of the tubes were transferred to 37°C and incubated for 90 min, while the other three were left at room temperature (23°C) to serve as controls. The flies in each tube were then homogenized in 1 ml of assay buffer (50 mM K_2PO_4 [pH 8.15], 1 mM $MgCl_2$, 0.25 mM phenylmethylsulfonyl fluoride). After a 30-s centrifugation to precipitate the cell debris, 10 μl of the supernatant was added to 1 ml of assay buffer containing 1 mM CPRG, and the mixture was incubated at 37°C for 3 to 5 h. After incubation, the optical density at 574-nm wavelength was determined, using samples from non-heat-shocked flies as controls. The levels of β-galactosidase activity of non-heat-shocked flies for each transgene are 2 to 5% greater

than that of the injection stock ry^{506}. Readings obtained from three parallel assays of the heat-shocked samples were averaged for each construct; results for mutant constructs were normalized to those for the wild-type control transgene (CarX or cP-351). The results from multiple independent lines for each transgene were then averaged to obtain the values shown in Fig. 2, 6, and 7. Standard errors of the mean were calculated and were not more than 15%.

The positive wild-type control transgene for most of the constructs studied is CarX. This construct contains *hsp26* sequences from −1917 to −372 and −351 to +632, which are fused in frame to the *E. coli lacZ* gene. CarX is different from transgene cP26Z, obtained previously (23), only in that sequences from −371 to −352, which contain HSE 7 (65), were deleted in CarX to simplify subsequent cloning. The 21-bp deletion in CarX does not appear to have any significant effect on gene activity; transgenes cP26Z and CarX have the same pattern of chromatin structure and the same levels of heat shock inducibility (40). Transgene CarX is the appropriate wild-type control for the transgenes described in Fig. 2a, as all of them lack sequences from −371 to −352 but otherwise include *hsp26* sequences from −1917 to +632. However, for transgenes cPΔCT · GA and cPri, the pertinent wild-type control is cP-351 (Fig. 2b). The chromatin structure of cP-351 may be influenced by lack of upstream *hsp26* sequences or by sequences from the *rosy* gene immediately upstream of −351 in these transgenes. We do observe a difference in absolute values (accessibility) for *Xba*I cleavage prior to heat shock in both the proximal and the distal DH sites in comparing the results obtained for transgenes CarX and cP-351 (Fig. 2 and Results).

Chromatin structure analyses. At least two independent transformed lines, which showed average β-galactosidase activity as quantitated by CPRG assays, were selected for chromatin structure analyses. Methods used for the isolation of nuclei from larvae, DNase I or *Xba*I treatment of nuclei, DNA purification, and Southern blot analyses using indirect end labeling are described elsewhere (41, 42). The probes used for detecting the DH sites and for quantitating the accessibility of *Xba*I sites within the proximal and distal DH sites are shown in Fig. 1.

To quantitate the accessibility of the *Xba*I site within the proximal DH site, nuclei were treated with an excess of *Xba*I. (Titration with increasing amounts of *Xba*I showed that maximum cleavage was obtained with 200 U of *Xba*I with 20 μg of genomic DNA in nuclei; no further cleavage was observed upon addition of further units of *Xba*I [data not shown].) The genomic DNA was then purified, cleaved to completion with *Eco*RV, fractionated by electrophoresis through a 1% agarose gel, and transferred to a piece of nylon membrane; the membrane was probed with the 1.1-kb *lacZ* fragment (Fig. 1). The intensities of the bands on the autoradiographs were measured with a scanning densitometer (Molecular Dynamics). The percent accessibility of the proximal *Xba*I site was determined by measuring the intensity of the band representing cleavage at that site compared with the total intensity of the bands reflecting cleavage at the proximal *Xba*I site, cleavage at the distal *Xba*I site, and no cleavage by *Xba*I (the parental *Eco*RV band) in the chromatin digestion.

To quantitate the accessibility of the *Xba*I site within the distal DH site, two strategies were used because of the differing lengths of 5′ *hsp26* sequences within the transgenes. For cP-351 and cPΔCT · GA, the DNA from nuclei treated with *Xba*I was purified and restricted to completion with *Eco*RV. For all other transgenes, the DNA was restricted to completion with *Sma*I and *Hpa*I. The DNA was size fractionated on a 1% agarose gel and transferred to a nylon membrane, and the membrane was probed with a 0.6-kb DNA fragment from the 3′ region of the xanthine dehydrogenase (*xdh*) gene; the *xdh* gene is located upstream of the *hsp26* sequences in the constructs used here (Fig. 1). The percent accessibility of the distal *Xba*I site was determined by using the methods described above for the proximal *Xba*I site. To simplify the comparison of relative values of *Xba*I accessibility between pertinent constructs, values for the transgenes shown in Fig. 2a and b have been normalized to those of CarX and cP-351, respectively. The variation for the measurements of *Xba*I accessibility is approximately ±5%.

It is noted that restriction enzyme digestion of a particular site in isolated nuclei has never been observed to reach 100% cleavage; a maximum cleavage of 75 to 85% has been obtained for sites that are nucleosome-free by other criteria (15, 30, 31, 70). Cleavage by various restriction enzymes within regions protected by nucleosomes (as assessed by digestion with other enzymes) is 6 to 8%. The lack of complete accessibility or complete protection for a particular site in nuclei may reflect the dynamic nature of the protein-DNA complexes in chromatin.

Purification of cloned GAGA factor. Plasmid pARGAGA was a gift of Walter Soeller and Thomas Kornberg. This plasmid contains a filled-in 1.6-kb *Eco*RI fragment from a GAGA factor cDNA, inserted into the filled-in *Bam*HI site of pAR3038 (59). In this plasmid, the first five codons of the GAGA factor have been replaced by eight codons of the S10 gene from the vector. *E. coli* BL21(DE3)(pLysS) cells transformed with either pARGAGA or pAR3038 were used to inoculate 500 ml of NZCYM medium (54) containing ampicillin (75 μg/ml) and chloramphenicol (34 μg/ml). The cultures were grown at 37°C with vigorous shaking to an optical density at 600 nm of 0.6, isopropylthiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the cultures were shifted to 22°C for 45 min. Rifampin was added to a final concentration of 40 μg/ml, and incubation was continued at 22°C overnight. Cells were harvested at 2,700 × *g* at 4°C for 10 min, resuspended in ice-cold water, and centrifuged as before. The cell pellet was resuspended in 20 ml of 0.2 M KCl–HEGNPD (HEGNPD is 25 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.8], 0.1 mM EDTA, 10% glycerol, 0.1% Nonidet P-40, 0.5 mM phenylmethylsulfonyl fluoride, and 0.5 mM dithiothreitol) containing 1 mM leupeptin. The cells were disrupted three times for 15 s each and two times for 30 s each with a model W-380 sonicator (Heat Systems-Ultrasonics, Inc.) equipped with a microtip. Disrupted cells were placed on ice for 15 min, and bacterial debris was removed by centrifugation at 6,000 × *g* at 4°C for 10 min. The supernatant was further clarified by centrifugation at 100,000 × *g* at 4°C for 1 h. Protein concentration of this supernatant was 5.4 mg/ml, as measured by a Bio-Rad protein assay. This supernatant was used immediately for column chromatography or frozen in liquid nitrogen and stored at −80°C for later use.

For column chromatography, 10 mg of protein was loaded onto a preequilibrated HR5/5 Mono S column (Pharmacia). Bound protein was eluted at a flow rate of 1 ml/min with a 20-ml linear gradient of 0.2 to 0.8 M KCl in HEGNPD, using a Waters 650 protein purification system. GAGA factor-containing fractions were identified by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and gel shift assay. The DNA template used for gel shift assays included *hsp26* sequences from −170 to +48; this DNA

sequence contains a long $(CT)_n \cdot (GA)_n$ stretch to which purified *Drosophila* GAGA factor binds in vitro (20). Extracts from *E. coli* transformed with the vector alone showed no DNA binding activity in gel shift assays (data not shown). GAGA factor eluted predominantly in two 1-ml fractions between 0.34 and 0.38 M KCl and was further purified from these fractions by using DNA affinity chromatography essentially as described by Gilmour et al. (20). Briefly, GAGA factor-containing fractions were pooled and diluted to 0.3 M KCl with HEGNPD, poly(dI-dC) was added to a final concentration of 25 μg/ml, and the pooled fractions were incubated on ice for 5 min. Fractions were then loaded onto a 750-μl DNA affinity column at a flow rate of approximately 0.05 ml/min. The column was washed with 2 volumes of 0.3 M KCl–HEGNPD, and bound protein was eluted with three consecutive 500-μl aliquots of 0.8 M KCl–HEGNPD. The majority of the GAGA factor eluted in the first fraction at 0.67 M KCl. We estimate this fraction to be greater than 65% pure, with a total protein concentration of ca. 60 μg/ml.

DNase I footprint analysis. DNase I footprint analysis of the distal *hsp26* promoter regions of CarX and CarXmCTd was performed by using *hsp26* sequences from −516 to −119. Templates were 5′ end labeled at the *Nde*I site (−516), using [γ-^{32}P]ATP and T4 polynucleotide kinase. After digestion at the *Apa*LI site (−119), end-labeled fragments were gel purified from 6% nondenaturing polyacrylamide gels. Analysis of proximal *hsp26* promoter regions of CarX, cPΔCT · GA, and cPri was performed by using *hsp26* sequences from −278 to +7. Templates were 5′ end labeled at the *Eco*RI site (+7). After digestion at the *Dra*I site (−278), end-labeled fragments were gel purified from 8% nondenaturing polyacrylamide gels. Two microliters of affinity-purified protein was added to a 10.5-μl reaction containing 10 mM HEPES (pH 8.0), 0.1 mM EDTA, 10% glycerol, 5 mM $MgCl_2$, 0.5 mM dithiothreitol, approximately 5 ng of end-labeled fragment (8,000 cpm), and either 50 ng of *Hae*III-digested *E. coli* DNA or 50 ng of $(CT)_n \cdot (GA)_n$ as the competitor. After a 25-min binding reaction at 25°C, 1.2 U of DNase I in DNase I buffer (50 mM NaCl, 20 mM Tris-HCl [pH 8.0], 5 mM $MgCl_2$) was added, and the mixture was incubated at 25°C for 35 s. The digestion was terminated by the addition of 25 μl of stop solution (1% Sarkosyl, 15 mM EDTA, 10 μg of yeast tRNA per ml, 40 μg of proteinase K per ml) followed by incubation at 55°C for 15 min. Samples were phenol extracted, ethanol precipitated, resuspended in sequencing loading dye, and analyzed on 8% sequencing gels. Autoradiography of dried gels was performed at −80°C for 16 to 24 hours with a Du Pont intensifying screen.

RESULTS

A distal $(CT)_n$ sequence contributes to the levels of heat shock inducibility of *hsp26*. We have previously reported that a sequence element containing $(CT)_n \cdot (GA)_n$ repeats (−135 to −85) in the promoter region of the *D. melanogaster hsp26* gene is required for optimal heat shock-induced expression; deletion or substitution of this proximal $(CT)_n$ element with a random sequence results in a four- to fivefold reduction in heat shock-induced expression (22). Computer analysis of the *Drosophila hsp26* promoter sequences to −351 (those required for optimal heat shock induction) indicates that there is a second $(CT)_n$ element, located from −338 to −332, near the distal HSE (Fig. 1). To assess the role of this 7-bp distal $(CT)_n$ element on *hsp26* expression, we constructed CarXmCTd and CarXΔCT/mCTd. In CarXmCTd, the distal $(CT)_n$ sequence 5′-CTCTCTC-3′ is changed to 5′-CTCGAGC-3′, whereas the proximal $(CT)_n$ sequence (−135 to −85) is intact; in CarXΔCT/mCTd, the distal $(CT)_n$ is mutated as described above and the proximal $(CT)_n$ is deleted. As shown in Fig. 2a, mutation of the distal $(CT)_n$ sequence alone reduces the heat shock-induced expression of the transgene to 64% of that of the wild-type control (CarX; see description of DNA constructs in Materials and Methods); the double mutation (CarXΔCT/mCTd) reduces the heat shock-induced expression to 8% of that of the wild-type control. These results demonstrate that the distal $(CT)_n$ sequence element also contributes to the levels of heat shock inducibility of this promoter.

The distal $(CT)_n$ sequence contributes significantly to the formation of the DH sites. Our previous studies have shown that the accessibility of a particular DNA region in nuclei can be quantitated by measuring the accessibility of a restriction site located within that region to cleavage by the enzyme; thus, changes in the DH sites can be quantitated by treatment of isolated nuclei with saturating amounts of a restriction enzyme which has a recognition site within the original DH site (31, 41, 42). To quantitate the hypersensitivity of both the proximal and the distal DH sites of the transgenes containing specific changes in either of the $(CT)_n$ sequences, we examined the accessibility of the *Xba*I site within each of the DH sites, using the strategies described in Materials and Methods and the probes shown in Fig. 1. We have previously determined that deletion of the proximal $(CT)_n$ sequence (−135 to −85) (in cPΔCT · GA) dramatically affects accessibility within the proximal DH site, reducing the *Xba*I cleavage to 15% of that of the wild-type control (cP-351) (41). As shown in Fig. 3a, deletion of the proximal $(CT)_n$ sequence (cPΔCT · GA) also reduces the accessibility of the distal DH site to 49% of that of the wild-type transgene cP-351. Previous analysis indicated that the change in spacing due to the deletion of the proximal $(CT)_n$ sequence in cPΔCT · GA has no significant effect per se, as similar results (levels of inducible gene expression, patterns of DH sites, and accessibility of *Xba*I to its site located within the proximal DH site) were obtained when the proximal $(CT)_n$ sequence was replaced by a random salmon sperm DNA sequence (cPri) (41).

The functional analysis described above shows that the 7-bp $(CT)_n$ sequence (−348 to −342) located next to HSE 6 also contributes to heat shock inducibility (compare results for CarX and CarXmCTd in Fig. 2a). To examine the effect of the mutation of the distal $(CT)_n$ sequence on chromatin structure, we carried out indirect end-labeling experiments using the strategies described in Materials and Methods (see Fig. 1 for probes used) to examine the pattern of DH sites of the transgene CarXmCTd. As shown in Fig. 4, CarXmCTd showed a DH site pattern similar to that of the wild-type control (CarX); however, the hypersensitivity was significantly reduced. Mutation at the distal $(CT)_n$ sequence (CarXmCTd) reduces the accessibility to cleavage by *Xba*I both at the distal DH site (56% of that of CarX) (Fig. 3b) and at the proximal DH site (46% of that of CarX) (Fig. 3c).

We also analyzed the chromatin structure of transgene CarXΔCT/mCTd, in which the distal $(CT)_n$ is mutated and the proximal $(CT)_n$ is deleted. Analysis with DNase I suggested that the DH sites were dramatically reduced (Fig. 4). Indeed, accessibility to both DH sites as measured by *Xba*I cleavage is severely reduced, to 9% of that of CarX for the proximal DH site (Fig. 3c) and to 18% of that of CarX for the distal DH site (Fig. 3b).

The $(CT)_n$ sequences are bound by the GAGA factor. The promoter-proximal $(CT)_n$ sequence has been shown to be a binding site for a CT-binding protein (20). The apparent

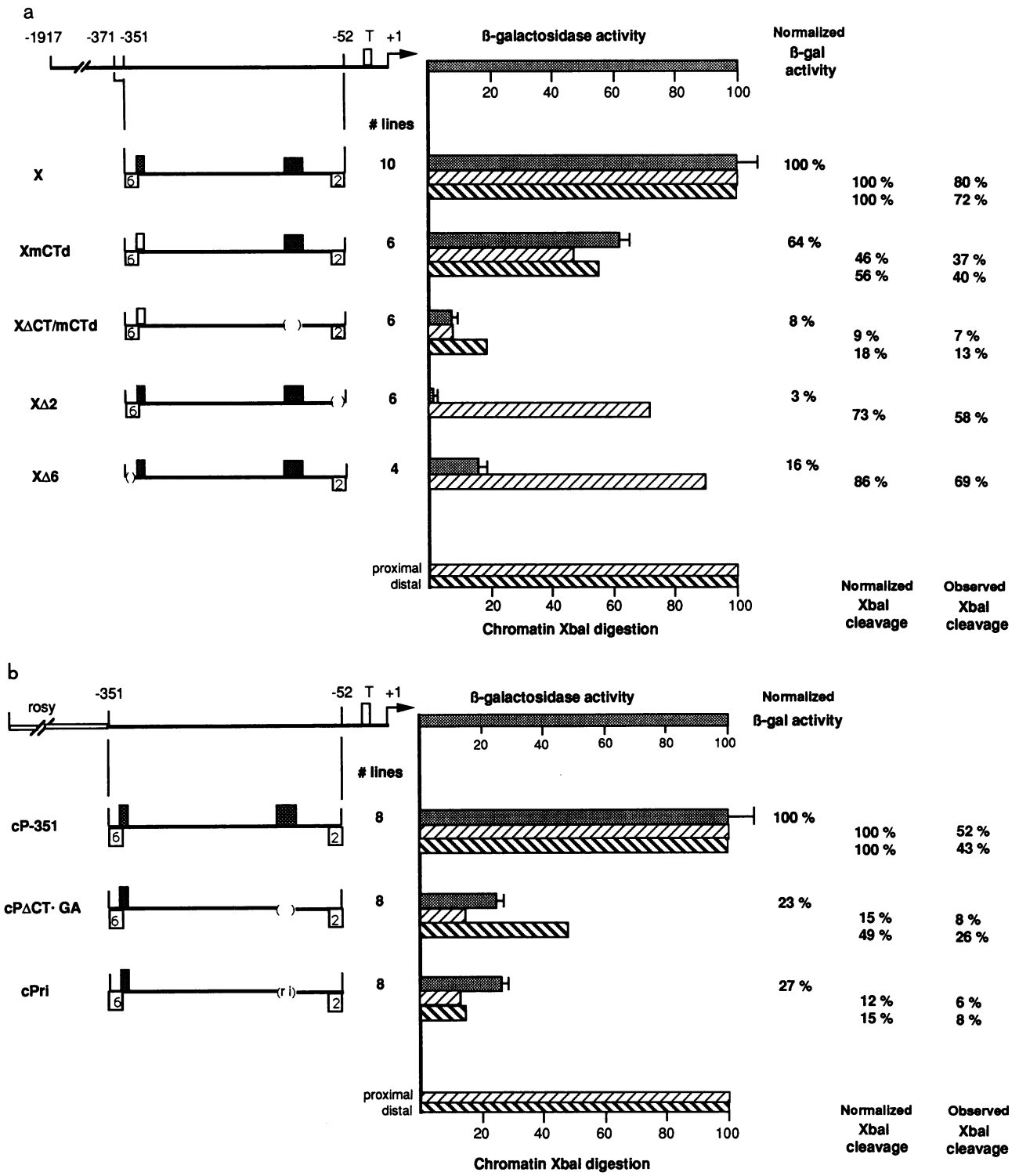

FIG. 2. Roles of HSEs and $(CT)_n$ sequences in optimal heat shock expression and in formation of the wild-type chromatin structure. (a) Constructs with alterations or deletions in HSE or $(CT)_n$ sequences, schematically presented on the left. Functional HSEs (HSE 1-2 and HSE 6; numbered boxes below the line) and sequences containing $(CT)_n \cdot (GA)_n$ repeats (grey [wild type] and open [mutant] boxes above the line) are positioned as indicated. T (top line) indicates the TATA box. Parentheses indicate the deletion of the corresponding elements. The number of independent transformed lines used to determine heat shock-inducible β-galactosidase activity is shown. The percentage values represent average levels of heat-induced activity and accessibility of the *XbaI* sites for each transgene, shown normalized to the values obtained for CarX. Within the bar graph, the first bar for each transgene shows relative levels of heat shock-induced expression, the second bar shows relative values of accessibility of the proximal DH site (from results shown in Fig. 3c), and the third bar, where given, shows relative values of accessibility of the distal DH site (from results shown in Fig. 3b). (b) Heat shock-induced β-galactosidase activity and chromatin structure of cPΔCT · GA and cPri normalized to the values obtained for cP-351. Symbols are as described above. Levels of β-galactosidase activity and accessibility of the proximal DH site were determined previously (41); levels of accessibility of the distal DH site were determined in this study (Fig. 3a).

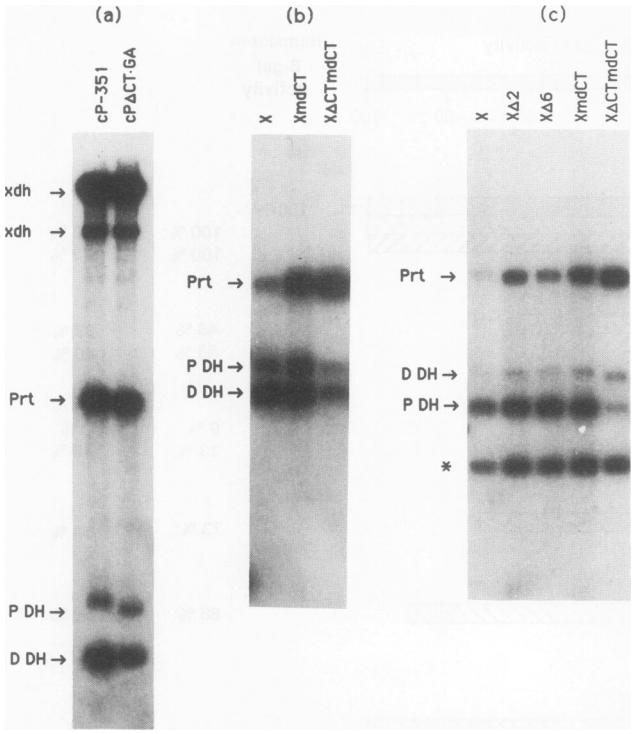

FIG. 3. *Xba*I accessibility within the proximal and distal DH sites. (a) *Xba*I accessibility within the distal DH site. Shown is an autoradiograph of an indirect end-labeling analysis of DNA isolated from *Xba*I-treated nuclei from transformants probed with the 0.6-kb *xdh* fragment (Fig. 1). The particular transformants used are identified above the lanes. The bands generated by cleavage at the *Xba*I site within the proximal DH (P DH) and distal DH (D DH) sites are indicated. The bands labeled *xdh* are derived from the endogenous *xdh* gene present in the host shock ACR ($Adh^{fn6}cn; ry^{502}$). The band labeled Prt is derived from *Sma*I and *Eco*RV digestion of the DNA that was not cleaved in nuclei at either *Xba*I site. (b) *Xba*I accessibility within the distal DH site. Shown is an autoradiograph of an indirect end-labeling analysis of DNA isolated from *Xba*I-treated nuclei, probed with the 0.6-kb *xdh* fragment (Fig. 1). The band labeled Prt is derived from *Sma*I and *Hpa*I digestion of DNA not cleaved in nuclei at either *Xba*I site. No bands from the endogenous *xdh* gene are present in panel b because the region containing the 0.6-kb *xdh* fragment, which is used as the probe for hybridization, has been deleted in the host stock ry^{506}. (c) *Xba*I accessibility within the proximal DH site. Shown is an autoradiograph of an indirect end-labeling analysis of DNA isolated from *Xba*I-treated nuclei, probed with the 1.1-kb *lacZ* fragment (Fig. 1). The parental band, which is created by *Eco*RV digestion of the DNA not cleaved in nuclei at either *Xba*I site, is indicated by Prt. The band indicated by an asterisk is unrelated to this group of constructs. All other designations for panels b and c are as described for panel a.

molecular size of this protein as estimated by SDS-PAGE (66 kDa) and its DNA binding specificity suggested that the CT-binding protein might be similar or identical to the GAGA factor isolated by Biggin and Tjian (5, 20). Using the method of Vinson et al. (68), we screened a *Drosophila* ovary cDNA expression library (61) to identify recombinant proteins which bound to $(CT)_n$ sequences. Approximately 10^6 clones were screened by using the ligated double-stranded oligonucleotide 5′-AGAGAGAGAGAGAGAAAAG AGAGAG-3′/3′-CTCTCTCTCTCTTTTCTCTCTCTCT-5′. Four positive clones were identified which bound this probe much more strongly than did a probe composed of 5′-CCA TTGGAGCTTAAGAGGGAGAATT-3′/3′-ACCTCGAATT CTCCCTCTTAAGGTA-5′. We compared these clones to a recently isolated GAGA factor cDNA (59); all were shown by restriction mapping, Southern hybridization, and sequence comparison to be partial GAGA factor cDNAs (data not shown).

We have purified the *E. coli*-expressed GAGA factor and have used this recombinant protein for DNase I footprinting analysis of the proximal and distal regions of the *hsp26* promoter. As shown in Fig. 5a, the recombinant GAGA factor protects the promoter proximal $(CT)_n$ stretch (−135 to −85) in CarX from digestions by DNase I (lane 2); the binding of recombinant GAGA factor is abolished by the addition of specific competitor DNA to the binding reaction (lane 3). The pattern of protection over this long $(CT)_n$ stretch is identical to the footprint produced by the native *Drosophila* CT-binding protein with use of the same DNA fragment as a binding substrate (19, 20, 24). Recombinant GAGA factor also binds to the promoter distal $(CT)_n$ stretch (−338 to −332), protecting 14 residues, including 7 bp of alternating C and T, from DNase I digestion (Fig. 5c, lane 2, solid bar). A less perfect $(CT)_n$ stretch extending from −346 to −384 is also weakly protected by the recombinant protein (Fig. 5c, lane 2, stippled bar) (note that sequences from −371 to −352 are not present in CarX and CarXmCTd). From previous results (20), the results from screening for additional $(CT)_n$-binding proteins, and the results of the footprinting assays, we conclude that the CT-binding protein is identical to the GAGA factor.

The binding of recombinant GAGA factor to *hsp26* sequences is eliminated by alterations or substitutions of the $(CT)_n$ stretches examined above. In construct cPΔCT · GA, the promoter-proximal $(CT)_n$ stretch has been deleted. As expected, recombinant GAGA factor fails to bind the proximal region, including the novel sequence at the deletion junction (data not shown). In cPri, the promoter-proximal $(CT)_n$ stretch (−135 to −85) has been replaced by a 47-bp sequence from salmon sperm DNA (22) which contains no $(CT)_n \cdot (GA)_n$ repeats (Fig. 5e). As shown in Fig. 5b, binding of recombinant GAGA factor is not detected over this region by DNase I footprinting analysis (compare Fig. 5b, lane 2, bracketed region, and Fig. 5a, lane 2, solid bar). CarXmCTd contains a 3-bp substitution which disrupts the promoter-distal $(CT)_n$ sequence; recombinant GAGA factor fails to bind to this abbreviated $(CT)_n$ stretch, although the imperfect $(CT)_n$ repeats immediately upstream are still weakly protected (compare Fig. 5c, lane 2, and Fig. 5d, lane 2, stippled bars). The results from DNase I footprinting analysis are summarized in Fig. 5e. These results indicate that the changes in the $(CT)_n$ regions, which affect chromatin structure and heat shock-induced expression, eliminate GAGA factor binding at the mutated sites.

Deletion of HSEs severely reduces heat shock inducibility but has only a minor effect on the formation of the DH sites. In transgenes containing a deletion of either of the two HSEs CarXΔ6 and CarXΔ2, heat shock expression was reduced to 16 and 3%, respectively, of that of the wild-type control CarX (Fig. 2a). This finding is consistent with previous observations that two HSEs are required for optimal heat shock-induced expression of this gene (9, 22, 45, 58). Multiple HSEs are also required for optimal heat shock-induced expression of other *Drosophila* heat shock genes (2, 14, 35, 50, 66).

To examine the contribution of HSEs to the formation of chromatin structure, we examined transgenes with either of the two HSEs deleted. As shown in Fig. 4, in both cases wild-type patterns of DH sites were observed. Determination of the accessibility of the *Xba*I site within the proximal DH

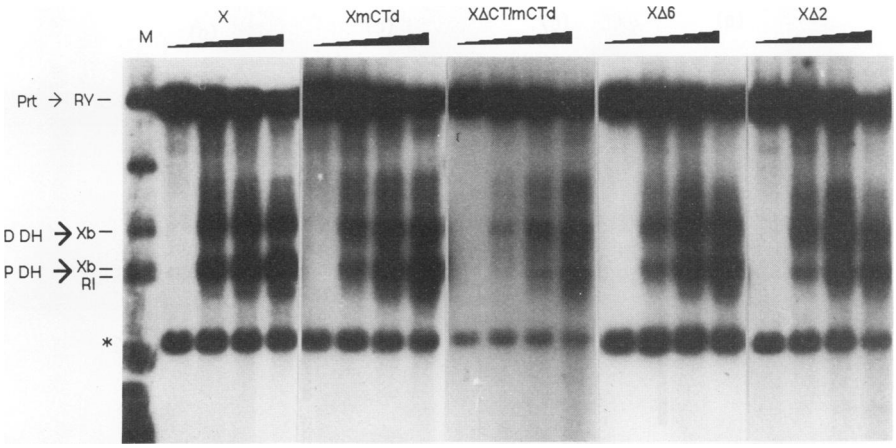

FIG. 4. DH site formation of transgenes with alterations or deletions in sequences containing $(CT)_n$ or HSE. Nuclei were isolated from non-heat-shocked larvae and treated briefly with DNase I, and the purified DNA was digested to completion with *Eco*RV. In addition, these samples were further digested to completion with *Sma*I, which does not cleave within the parental *Eco*RV fragment of interest but does cut an unrelated DNA fragment into the observed band indicated by an asterisk. The samples were subjected to indirect end-labeling analysis using the 1.1-kb *lacZ* fragment as a probe (Fig. 1). DH sites are indicated by arrows labeled P DH (proximal DH site) and D DH (distal DH site). The parental band, which is created by *Eco*RV digestion of DNA not cleaved in nuclei by DNase I, is indicated by an arrow labeled Prt. Transgenes analyzed are identified above the appropriate lanes. Wedge bars represent increasing amounts of DNase I used. M, markers; RI, *Eco*RI; RV, *Eco*RV; X, *Xba*I.

site indicated that deletion of a 33-bp sequence containing HSE 1-2 (CarXΔ2) reduces the accessibility to 73% of that of CarX; a 6-bp deletion which destroys HSE 6 (CarXΔ6) reduces the accessibility to 86% of that of CarX (Fig. 3c). Thus, deletion of either of the two HSEs has a minor effect on the formation of DH sites prior to heat shock induction.

The regulatory region (−351 to −52) is orientation independent and can tolerate a helical twist. The *Xba*I fragment (−351 to −52) of *hsp26* contains all of the necessary regulatory elements upstream of the TATA box for heat shock response, with an HSE/$(CT)_n$ unit located on each side of a DNA segment that is associated with a nucleosome in vivo (65). To test whether this regulatory unit is orientation dependent, we studied the heat shock inducibility of a transgene in which the entire *Xba*I fragment was inverted with respect to its native orientation (CarFX). As shown in Fig. 6, the heat shock-induced expression of CarFX is not significantly different from that of CarX, indicating that the inversion of the *Xba*I fragment does not appear to significantly affect the heat shock inducibility of the transgene.

The arrangement of the two HSEs also appears to be quite flexible in terms of their relative location on the DNA helix; a transgene containing a 5-bp insertion ($CarX^5$), which would change the relative rotational alignment between the two HSEs or between the proximal $(CT)_n$ element and the proximal HSE by half a helical turn, has no dramatic effect on heat inducibility; the heat-induced expression remains 87% of that of CarX (Fig. 6). Both transgenes CarFX and $CarX^5$ form wild-type DH site patterns, as assayed by DNase I digestion of the chromatin in isolated nuclei (data not shown). Thus, it appears that the arrangement of the regulatory sequences upstream of the TATA box is functionally symmetrical. The proximal and the distal HSEs, which are separated by about 300 bp of DNA, are not sensitive to their relative location on the surface of the helix in this promoter. The latter result is consistent with a previous analysis showing that while the interaction of two closely spaced HSEs is sensitive to rotational alignment, that of elements spaced more than 80 bp apart is not; presumably the greater separation allows sufficient flexibility to compensate for the rotational change (10).

The *hsp26* regulatory region (−351 to −52) is dependent on a $(CT)_n$ sequence to direct a minimal promoter. Given previous results (22, 41) and those presented here concerning the *Xba*I fragment (−351 to −52) of the *hsp26* promoter region, which contains multiple HSEs and $(CT)_n \cdot (GA)_n$ repeats, we examined whether this *Xba*I fragment could direct heat shock-dependent expression of a minimal promoter with the same dependency on $(CT)_n$ sequences. To this end, we made construct CarX-70Z, in which the *Xba*I fragment (−351 to −52) of *hsp26* was placed upstream of the TATA box of a *Drosophila hsp70-lacZ* fusion gene (69) at −50 of *hsp70* (Fig. 7) (sequences downstream of −50 of *Drosophila hsp70* do not contain any necessary HSEs, and the promoter is not heat shock inducible). The heat shock-induced expression of transgene CarX-70Z is approximately 50% of that of a *hsp70* wild-type control transgene, cP70Z. [cP70Z is a *hsp70-lacZ* construct containing *hsp70* sequences from −89 to +260, including HSE 1, HSE 2, and a 7-bp $(GA)_n \cdot (CT)_n$ sequence located between the two HSEs (69).] The heat shock expression of CarXΔCT-70Z, in which the proximal $(CT)_n$ sequences were deleted in the background of CarX-70Z, is severely reduced, to 12% of that of CarX-70Z (Fig. 7), indicating that $(CT)_n$ sequences contribute to the heat shock inducibility of CarX-70Z. Chromatin structure analysis using DNase I indicated that the DH site pattern in transgene CarX-70Z resembled that of CarX; upon deletion of the $(CT)_n$ sequence in transgene CarXΔCT-70Z, the DH sites, particularly the proximal one, are dramatically reduced (data not shown). These results indicate that in regulating heat shock-dependent expression of a minimal promoter, the *Xba*I fragment shows the same dependency on $(CT)_n$ sequences as that seen in the *hsp26* gene.

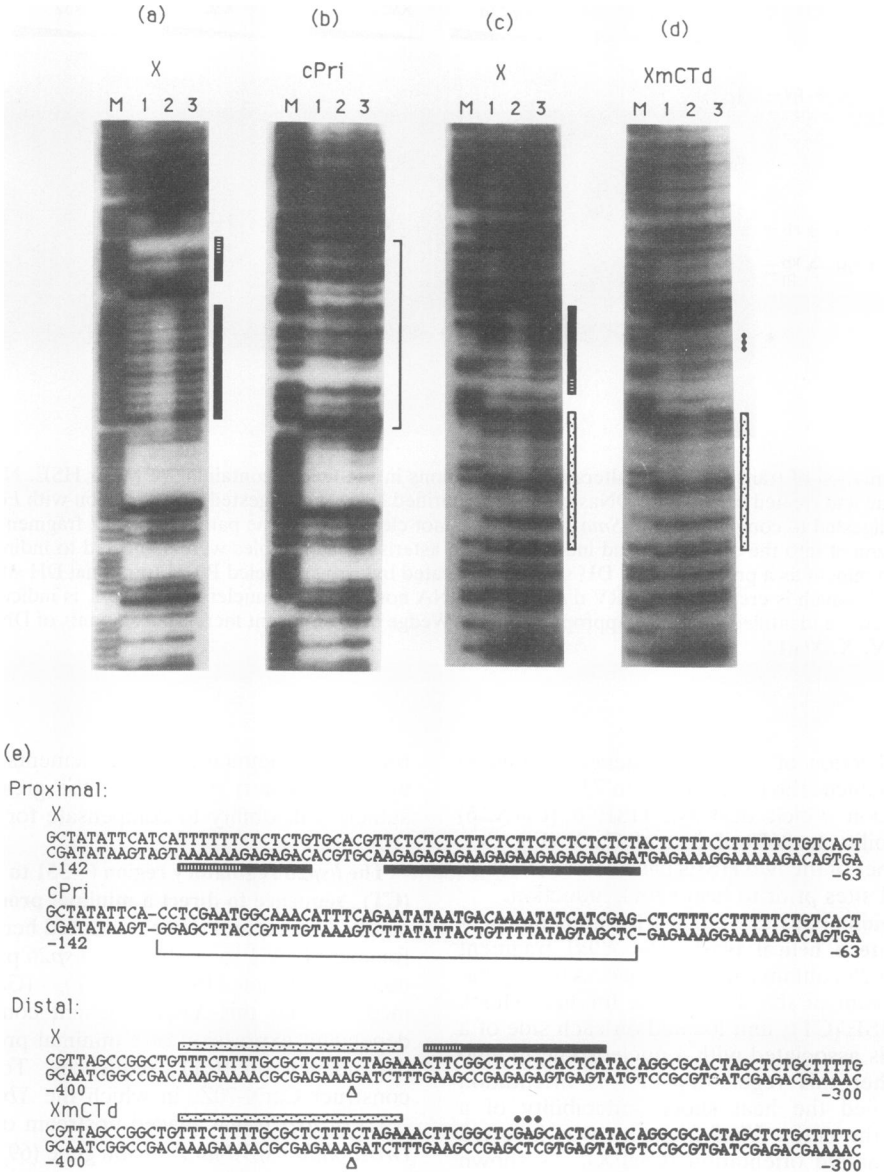

FIG. 5. Binding of the GAGA factor to the $(CT)_n$ sequences. (a) DNase I footprinting analysis of the recombinant GAGA factor on proximal promoter regions of construct CarX. Binding reactions were performed as described in Materials and Methods. Lanes: M, Maxam-Gilbert purine cleavage reactions; 1, no protein added to the binding reaction; 2, recombinant GAGA factor added to the binding reaction; 3, recombinant GAGA factor added, with 50 ng of $(CT)_n \cdot (GA)_n$ oligonucleotide (instead of *Hae*III-digested *E. coli* DNA) included as a competitor in the binding reaction. Strong footprints are indicated by solid vertical bars; hatching indicates uncertainty in delimiting the footprints. (b) DNase I footprinting analysis of the recombinant GAGA factor on proximal promoter regions of construct cPri. The bracket indicates the sequence replaced by 47 bp of salmon sperm DNA. Lanes are numbered as in panel a. (c) DNase I footprinting analysis of the recombinant GAGA factor on distal promoter regions of construct CarX. A strong footprint is indicated by the solid bar, hatching indicates uncertainty in delimiting the footprint, and a weak footprint is indicated by a stippled bar. Lanes are numbered as in panel a. (d) DNase I footprinting analysis of the recombinant GAGA factor on distal promoter regions of construct CarXmCTd. The three diamonds indicate the base pairs which have been mutated in this construct relative to CarX. Lanes are numbered as in panel a. (e) Summary of the DNase I footprint analysis on CarX, cPri, and CarXmCTd. Strong footprints are indicated by solid bars, weak footprints are indicated by stippled bars, and hatching indicates uncertainty in delimiting a footprint. The bracket below the cPri sequence indicates those base pairs derived from salmon sperm DNA. The deletion of 21 bp (−371 to −352) from the distal promoter region is indicated by a delta below the distal CarX and distal CarXmCTd sequences. The three diamonds above the CarXmCTd sequence indicate the three base pairs which have been mutated in this construct relative to CarX.

DISCUSSION

We have examined the chromatin structure and gene expression of a series of *Drosophila hsp26-lacZ* fusion transgenes containing deletions or mutations of the regulatory elements in the promoter region of *hsp26*. Our results indicate that $(CT)_n \cdot (GA)_n$ repeats play an integral role in the formation of the appropriate chromatin structure of *hsp26* and in its heat shock-induced expression. In contrast, the HSEs are essential for transcription activation but have a relatively minor role in the formation of the appropriate chromatin structure.

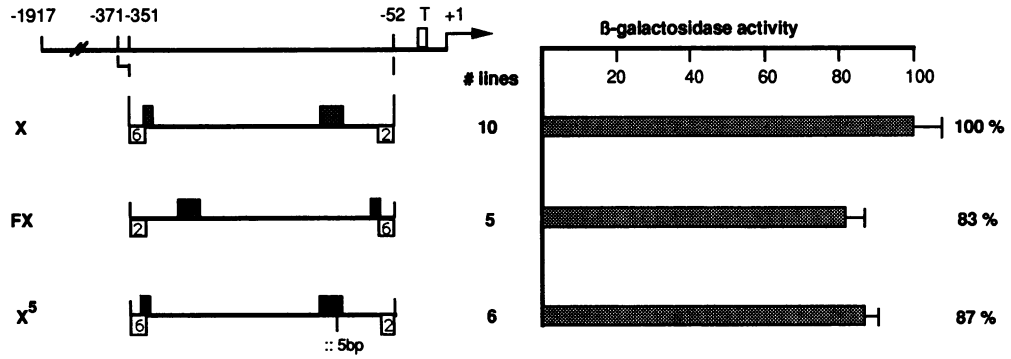

FIG. 6. Flexibility in arrangement of regulatory elements at the promoter region of the *hsp26* gene. Constructs CarX, CarFX, and CarX^5 are schematically presented at the left. Functional HSEs (HSE 1-2 and HSE 6) and sequences containing $(CT)_n \cdot (GA)_n$ repeats are as indicated in Fig. 2. T indicates the TATA box. The number of independent lines used to determine the average β-galactosidase activity is shown. The graph shows levels of heat shock-induced expression obtained by measuring the relative level of β-galactosidase activity, with that of CarX set at 100%.

The functional role of the $(CT)_n$ elements. As shown in Fig. 3 and 4, alteration of either of the $(CT)_n$ elements decreases both the nuclease accessibility of the DH sites and levels of heat shock inducibility; alteration of both sites has a dramatic effect, suggesting that these two $(CT)_n$ regions may work together in establishing DH sites. How do $(CT)_n$ elements act in the formation of the wild-type chromatin structure prior to heat shock induction? In vitro, $(CT)_n$ sequences have been shown to have two distinct properties: they can form an intramolecular triplex structure (H-form DNA) if sufficient in length, and they can serve as binding sites for GAGA factor. The proximal $(CT)_n$ region of *hsp26* can adopt an H-form DNA structure in a supercoiled plasmid in vitro at low pH (57, 64). However, our results indicate that such a structure may not be relevant to the activity described here. First, the 7 bp of $(CT)_n \cdot (GA)_n$ at the distal DH site of *hsp26* (this work) and 7 bp of $(GA)_n \cdot (CT)_n$ within the proximal regulatory region of *Drosophila hsp70* (41) both have functional roles in establishing DH sites. Such a short sequence cannot form a stable triplex structure in vitro (28). Second, GAGA factor binds to the native $(CT)_n$ sequences of *hsp26* specifically in vitro but fails to bind to the random sequence substituted for the proximal $(CT)_n$ element or to the mutated distal $(CT)_n$ element (Fig. 5). Taken together, these results suggest that the $(CT)_n$ sequences function in vivo in the formation of DH sites by serving as binding sites for GAGA factor.

How could GAGA factor, shown to be a transcription factor by in vitro assessment (5), influence chromatin structure? It is noted that before heat shock induction, the TATA boxes of both *hsp70* and *hsp26* are occupied by TATA-binding protein(s) (65, 72); however, the TATA box and downstream sequences of *hsp26* are not sufficient for the formation of DH sites (41). Preliminary evidence has suggested that GAGA factor binding at the proximal site may facilitate the binding of a TATA box-binding complex on the promoter (13). Alternatively, the GAGA factor may be involved in directing or limiting the deposition of nucleosomes or histone H1 within the regulatory region prior to gene activation, acting either during or after chromatin assembly following replication (12, 34).

Studies on the *Drosophila hsp70* gene have also suggested a role for GAGA factor in establishing the poised RNA polymerase II (37), a transcriptionally engaged RNA polymerase II molecule detected on the 5′ end of the promoter prior to heat shock induction (51). Alteration in the $(GA)_n \cdot (CT)_n$ repeats (7 bp of alternating GA located between HSE 1 and HSE 2 of *hsp70*) dramatically affects the level of poised RNA polymerase II prior to heat shock activation, whereas mutation in HSE 1 has only a minor

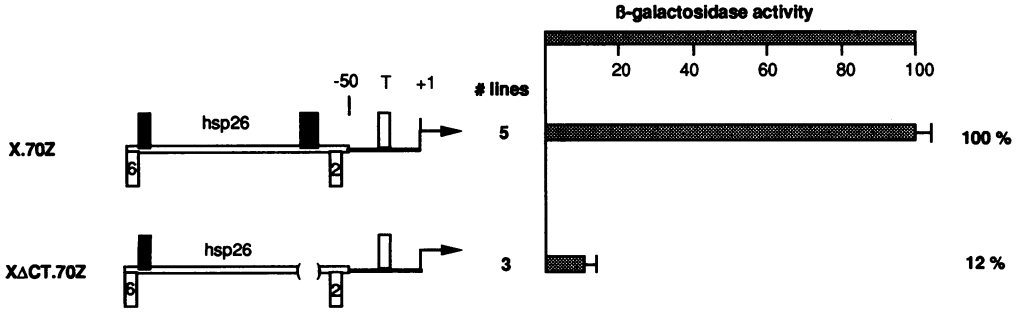

FIG. 7. Ability of the *Xba*I fragment to drive a minimal promoter. Constructs CarX-70Z (X · 70Z) and CarXΔCT-70Z (XΔCT · 70Z) are schematically presented at the left. *hsp26* sequences (−351 to −52) are represented as an open horizontal bar; *hsp70* sequences (−50 to +260) are represented as a solid bar. Functional HSEs (HSE 1-2 and HSE 6) and sequences containing $(CT)_n \cdot (GA)_n$ repeats of *hsp26* sequences are as indicated in Fig. 2. T indicates the TATA box. Parentheses on CarXΔCT-70Z indicate the deletion of the proximal $(CT)_n$ repeats. The number of independent lines used to determine the β-galactosidase activity is shown. The graph shows levels of heat shock-induced expression obtained by measuring the relative level of β-galactosidase activity, with that of CarX-70Z set at 100%.

effect on the level of poised RNA polymerase II (37); however, either of the alterations results in a dramatic reduction in heat shock-induced expression (37, 69). Previous experiments have shown that this 7-bp $(GA)_n \cdot (CT)_n$ region binds the GAGA factor specifically in vitro (20). Poised RNA polymerase II has also been detected on the 5′ sequences of *Drosophila hsp26* (52); a similar role of the $(CT)_n$ sequences of *hsp26* in establishing the poised polymerase can be envisioned. Poised RNA polymerase II may be an integral part of the preset chromatin structure which allows for rapid transcriptional activation.

It is noted that the effect of mutations in the proximal $(CT)_n$ element on chromatin structure and heat shock inducibility is more pronounced than the effect of mutation of the distal $(CT)_n$ element (Fig. 2a). The proximal $(CT)_n$, lying closer to the TATA box, could play a larger role in stabilizing the transcription complex. Alternatively, the greater effect of the proximal $(CT)_n$ sequence may simply be due to the relative length of the $(CT)_n \cdot (GA)_n$ repeats; the proximal $(CT)_n$ contains three $(CT)_n \cdot (GA)_n$ repeats, each 7 bp in length, whereas the distal $(CT)_n$ contains only one 7-bp $(CT)_n \cdot (GA)_n$ repeat. We also note that the mutation of the distal $(CT)_n$ element in CarXmCTd does not abolish the weaker binding of GAGA factor to a CT-rich region immediately upstream (Fig. 5c and d). This CT-rich region might contribute to the residual activity of CarXmCTd.

The GAGA factor and the HSF have distinct roles in chromatin structure and transcriptional activation of *hsp26*. The evidence presented indicates that GAGA factor, an abundant protein in *D. melanogaster*, is critical to establishing the chromatin structure of *hsp26* and does so prior to transcription activation by a mechanism independent of activating HSF. The chromatin structure in non-heat-shocked larvae is not dramatically affected upon deletion of either of the HSEs (Fig. 3 and 4), whereas the heat inducibility of the gene is dramatically affected (Fig. 2a). This finding is consistent with previous observations that HSEs are not occupied by HSF prior to heat shock induction in *D. melanogaster* (6, 65, 66, 72, 73). The results suggest that HSF is not required for the formation of the preset chromatin structure at *hsp26*. However, there is a minor effect on DNase I hypersensitivity upon deletion of a HSE, more pronounced for HSE 1-2 than for HSE 6. This effect might be caused by the spacing change from the deletion (33-bp deletion in CarXΔ2; 6-bp deletion in CarXΔ6).

A recent study of the yeast *hsp82* gene has shown that when the proximal HSE 1 is mutated, both basal and heat shock-induced transcription are significantly reduced, whereas the nuclease-sensitive region at the promoter is still maintained (38). The two systems differ in that in yeast cells, the HSF is bound to the HSEs constitutively (25, 33), while in *Drosophila* cells, the HSF is bound to the HSEs only after heat shock (6, 65, 66, 72, 73). Nonetheless, the observation that transcriptional activation can be uncoupled from the formation of appropriate chromatin structure clearly supports our conclusions from our study of *hsp26*.

Thus, GAGA factor and HSF appear to have distinct roles in gene regulation, the former setting up chromatin structure and the latter exploiting it. Is this preset structure necessary for gene activation? In all of the *hsp26-lacZ* transgenes examined, reduction in nuclease accessibility of the *Xba*I sites is always associated with reduced heat shock inducibility, even when the HSEs are intact and in their normal positions. Studies in progress indicate that in a mutant transgene in which the distal HSE and the distal $(CT)_n$ sequence are occupied by a nucleosome, the formation of both DH sites and heat shock inducibility of the gene are dramatically reduced (43). Thus, formation of the preset chromatin structure at this locus prior to heat shock, providing a structure in which the HSEs are always accessible, appears to be necessary for the subsequent gene activation by HSF upon heat shock induction. In vitro analysis supports the hypothesis that a preset structure is required at the heat shock loci; while some transcription factors, such as GAL4 and glucocorticoid receptor, can bind to DNA in nucleosomes, potentially destabilizing and remodeling that structure, the HSF apparently cannot (3, 63, 71).

The preset chromatin structure of *hsp26*. Previous work in this laboratory led to a model for the *hsp26* gene promoter in which the upstream regulatory sequences, such as HSE 6, are brought to promoter proximity by winding the intervening sequences on a precisely positioned nucleosome (65). The facts that the formation of the two DH regions are generally coordinated and that the *Xba*I fragment (−351 to −52) which contains all of the necessary HSEs and $(CT)_n$ sequences for expression is orientation independent support the looping model. Similar structures that facilitate protein-protein interactions by DNA bending or looping have been proposed for gene promoters in both prokaryotic and eukaryotic systems to enhance assembly of transcription complexes (8, 32, 47, 48) or replication complexes (44) (for a review, see reference 56). Similarly, Schild et al. have reported that assembly of a nucleosome in a region between the promoter and enhancer of the *Xenopus* vitellogenin B1 gene potentiates estrogen receptor-mediated transcription approximately 5- to 10-fold in vitro (55). On the basis of previous data and the evidence presented here, we propose that the looping structure at the promoter region of the *Drosophila hsp26* gene is a part of the preinitiation complex that is required for rapid transcriptional activation by HSF upon heat shock induction.

We conclude that in this instance, the GAGA factor plays a major role in setting up chromatin structure, presumably by interacting with the transcriptional machinery and/or interacting (or competing) with nucleosomes during chromatin assembly. In vitro competition and assembly experiments controlled for the presence of GAGA factor will provide a more explicit test of this model and of the mechanism of formation of the preset chromatin structure.

ACKNOWLEDGMENTS

We thank John T. Lis, Charles Giardina, David S. Gilmour, Iain L. Cartwright, and members of the Elgin laboratory for helpful discussions and critical reading of the manuscript. We are indebted to John T. Lis and his colleagues for providing plasmids Car20T, pΔZX, and pX^{u}S26Z, and P-element transformants carrying cP70Z. We thank Walter Soeller and Thomas Kornberg for the generous gift of plasmid pARGAGA prior to publication, and we thank Kathryn Miller for making available the *Drosophila* ovarian cDNA expression library. Technical assistance from Jo Wuller, Jingyuan Xu, Dalit Meyer, and Vessela Ivanova during various stages of this project is appreciated.

This work is supported by NIH grant GM31532 to S. C. R. Elgin. L. L. Wallrath is supported by National Service Award GM07232.

REFERENCES

1. **Almer, A., H. Rudolph, A. Hinnen, and W. Hörz.** 1986. Removal of positioned nucleosomes from the yeast *PHO5* promoter on *PHO5* induction releases additional upstream activating DNA elements. EMBO J. **5**:2689–2691.
2. **Amin, J., R. Mestril, R. Lawson, H. Klapper, and R. Voellmy.** 1985. The heat shock consensus sequence is not sufficient for *hsp70* gene expression in *Drosophila melanogaster*. Mol. Cell. Biol. **5**:197–203.

3. **Archer, T. K., M. G. Cordingly, R. G. Wolford, and G. L. Hager.** 1991. Transcription factor access is mediated by accurately positioned nucleosomes on the mouse mammary tumor virus promoters. Mol. Cell. Biol. **11**:688–698.
4. **Bienz, M., and H. R. B. Pelham.** 1986. Heat shock regulatory elements function as an inducible enhancer in the *Xenopus hsp70* gene and when linked to a heterologous promoter. Cell **45**:753–760.
5. **Biggin, M. D., and R. Tjian.** 1988. Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. Cell **53**:699–711.
6. **Cartwright, I. L., and S. C. R. Elgin.** 1986. Nucleosomal instability and induction of new upstream protein-DNA associations accompany activation of four small heat shock protein genes in *Drosophila melanogaster*. Mol. Cell. Biol. **6**:779–791.
7. **Casadaban, M., A. Martinez-Arias, S. K. Shapira, and J. Chou.** 1983. β-Galactosidase gene fusions for analyzing gene expression in *Escherichia coli* and yeast. Methods Enzymol. **100**:293–308.
8. **Choi, O.-R. B., and J. D. Engel.** 1988. Developmental regulation of β-globin gene switching. Cell **55**:17–26.
9. **Cohen, R. S., and M. Meselson.** 1985. Separate regulatory elements for the heat-inducible and ovarian expression of the *Drosophila hsp26* gene. Cell **43**:737–746.
10. **Cohen, R. S., and M. Meselson.** 1988. Periodic interactions of heat shock transcriptional elements. Nature (London) **332**:856–858.
11. **Cordingly, M. G., A. T. Riegel, and G. L. Hager.** 1987. Steroid-dependent interaction of transcription factors with the inducible promoter of mouse mammary tumor virus *in vivo*. Cell **48**:261–270.
12. **Croston, G. E., L. A. Kerrigan, L. M. Lira, D. R. Marshak, and J. T. Kadonaga.** 1991. Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. Science **251**:643–649.
13. **Dietz, T. J.** 1990. The identification, isolation and characterization of a *Drosophila* TATA box binding activity. Ph.D. thesis. Washington University, St. Louis, Mo.
14. **Dudler, R., and A. A. Travers.** 1984. Upstream elements necessary for optimal function of the *hsp70* promoter in transformed flies. Cell **38**:391–398.
15. **Eggert, H., and R. S. Jack.** 1991. An ectopic copy of the *Drosophila ftz* associated SAR neither reorganizes local chromatin structure nor hinders elution of a chromatin fragment from isolated nuclei. EMBO J. **10**:1237–1243.
16. **Eissenberg, J. C., I. L. Cartwright, G. H. Thomas, and S. C. R. Elgin.** 1985. Selected topics in chromatin structure. Annu. Rev. Genet. **19**:485–536.
17. **Fascher, K.-D., J. Schmitz, and W. Hörz.** 1990. Role of trans-activating proteins in the generation of active chromatin at the *PHO5* promoter in *S. cerevisiae*. EMBO J. **9**:2523–2528.
18. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
19. **Gilmour, D. S. (Pennsylvania State University).** Personal communication.
20. **Gilmour, D. S., G. H. Thomas, and S. C. R. Elgin.** 1989. *Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters. Science **245**:1487–1490.
21. **Glaser, R. L., and J. T. Lis.** 1990. Multiple compensatory regulatory elements specify spermatocyte-specific expression of the *Drosophila hsp26* gene. Mol. Cell. Biol. **10**:131–137.
22. **Glaser, R. L., G. H. Thomas, E. S. Siegfried, S. C. R. Elgin, and J. T. Lis.** 1990. Optimal heat-induced expression of the *Drosophila hsp26* gene requires a promoter sequence containing $(CT)_n \cdot (GA)_n$ repeats. J. Mol. Biol. **211**:751–761.
23. **Glaser, R. L., M. F. Wolfner, and J. T. Lis.** 1986. Spacial and temporal pattern of *hsp26* expression during normal development. EMBO J. **5**:747–754.
24. **Granok, H., and S. C. R. Elgin.** Unpublished data.
25. **Gross, D. S., K. E. English, K. W. Collins, and S. Lee.** 1990. Basal-level expression of the yeast *HSP82* gene requires a heat shock regulatory element. J. Mol. Biol. **216**:611–631.
26. **Gross, D. S., and W. T. Garrard.** 1988. Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. **57**:159–197.
27. **Han, M., and M. Grunstein.** 1988. Nucleosome loss activates yeast downstream promoters *in vivo*. Cell **55**:1137–1145.
28. **Htun, H., and J. E. Dahlberg.** 1989. Topology and formation of triple-stranded H-DNA. Science **243**:1571–1576.
29. **Ingolia, T. D., and E. A. Craig.** 1981. Primary sequence of the 5′ flanking regions of the *Drosophila* heat shock genes in chromosome subdivision 67B. Nucleic Acids Res. **9**:1627–1642.
30. **Jack, R. S., and H. Eggert.** 1990. Restriction enzymes have limited access to DNA sequences in *Drosophila* chromosomes. EMBO J. **9**:2603–2609.
31. **Jack, R. S., P. Moritz, and S. Cremer.** 1991. Restriction enzymes permit quantitative determination of defined chromatin structures within the chromosome. Eur. J. Biochem. **202**:441–446.
32. **Jackson, J. R., and C. Benyajati.** 1992. *In vivo* stage- and tissue-specific DNA-protein interactions at the *D. melanogaster* alcohol dehydrogenase distal promoter and adult enhancer. Nucleic Acids Res. **20**:5413–5422.
33. **Jakobsen, B. K., and H. R. B. Pelham.** 1988. Constitutive binding of yeast heat shock factor to DNA in vivo. Mol. Cell. Biol. **8**:5040–5042.
34. **Kerrigan, L. A., G. E. Croston, L. M. Lira, and J. T. Kadonaga.** 1991. Sequence-specific transcriptional antirepression of the *Drosophila Krüppel* gene by the GAGA factor. J. Biol. Chem. **266**:574–582.
35. **Klemenz, R., and W. J. Gehring.** 1986. Sequence requirement for expression of the *Drosophila melanogaster* heat shock protein *hsp22* gene during heat shock and normal development. Mol. Cell. Biol. **6**:2011–2019.
36. **Kunkel, T. A., J. D. Roberts, and R. A. Zabour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
37. **Lee, H.-S., K. H. Kraus, M. F. Wolfer, and J. T. Lis.** 1992. DNA sequence requirements for generating paused polymerase at the start of *hsp70*. Genes Dev. **6**:284–295.
38. **Lee, M.-S., and W. T. Garrard.** 1992. Uncoupling gene activity from chromatin structure: promoter mutations can inactivate transcription of the yeast *HSP82* gene without eliminating nucleosome-free regions. Proc. Natl. Acad. Sci. USA **89**:9166–9170.
39. **Lis, J. T., H. Xiao, and O. Perisic.** 1990. Modular units of heat shock regulatory regions: structure and function, p. 411–428. *In* R. Morimoto, A. Tissières, and C. Georgopoulos (ed.), Stress proteins in biology and medicine. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
40. **Lu, Q., and S. C. R. Elgin.** Unpublished data.
41. **Lu, Q., L. L. Wallrath, B. D. Allan, R. L. Glaser, J. T. Lis, and S. C. R. Elgin.** 1992. A promoter sequence containing $(CT)_n \cdot (GA)_n$ repeats is critical for the formation of the DNase I hypersensitive sites in the *Drosophila hsp26* gene. J. Mol. Biol. **225**:985–998.
42. **Lu, Q., L. L. Wallrath, and S. C. R. Elgin.** 1993. Using *Drosophila* P-element mediated germline transformation to examine the chromatin structure and expression of *in vitro* modified genes, p. 333–357. *In* K. W. Adolph (ed.), Methods in molecular genetics, vol. 1. Academic Press, Orlando, Fla.
43. **Lu, Q., L. L. Wallrath, and S. C. R. Elgin.** Unpublished data.
44. **Miron, A., S. Mukherjee, and D. Bastia.** 1992. Activation of distant replication origins *in vivo* by DNA looping as revealed by a novel mutant form of an initiator protein defective in cooperativity at a distance. EMBO J. **11**:1205–1216.
45. **Pauli, D., A. Spierer, and A. Tissières.** 1986. Several hundred base pairs upstream of *Drosophila hsp23* and *26* genes are required for their heat induction in transformed flies. EMBO J. **5**:755–761.
46. **Piña, B., V. Bruggemeier, and M. Beato.** 1990. Nucleosome positioning modulates accessibility of regulatory proteins to the mouse mammary tumor virus promoter. Cell **60**:719–731.
47. **Ptashne, M.** 1986. Gene regulation by proteins acting nearby and at a distance. Nature (London) **322**:697–701.
48. **Qian, S., B. Varjavand, and V. Pirrotta.** 1992. Molecular anal-

ysis of the zest-white interaction reveals a promoter-proximal element essential for distant enhancer-promoter communication. Genetics **131:**79–90.

49. **Reik, A., G. Schutz, and A. F. Stewart.** 1991. Glucocorticoids are required for establishment and maintenance of an alteration in chromatin structure: induction leads to a reversible disruption of nucleosomes over an enhancer. EMBO J. **10:**2569–2576.
50. **Riddihough, G., and H. R. B. Pelham.** 1986. Activation of the *Drosophila hsp27* promoter by heat shock and by ecdysone involves independent and remote regulatory sequences. EMBO J. **5:**1653–1658.
51. **Rougvie, A. E., and J. T. Lis.** 1988. The RNA polymerase II molecule at the 5′ end of the uninduced *hsp70* gene of *Drosophila melanogaster* is transcriptionally engaged. Cell **54:**795–804.
52. **Rougvie, A. E., and J. T. Lis.** 1990. Postinitiation transcription control in *Drosophila melanogaster*. Mol. Cell. Biol. **10:**6041–6045.
53. **Rubin, G., and A. C. Spradling.** 1982. Genetic transformation of *Drosophila* with transposable element vectors. Science **218:** 348–353.
54. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
55. **Schild, C., F.-X. Claret, W. Wahli, and A. P. Wolffe.** 1993. A nucleosome-dependent static loop potentiated estrogen-regulated transcription from the *Xenopus* vitellogenin B1 promoter in vitro. EMBO J. **12:**423–433.
56. **Schleif, R.** 1992. DNA looping. Annu. Rev. Biochem. **61:**199–223.
57. **Siegfried, E., G. H. Thomas, U. M. Bond, and S. C. R. Elgin.** 1986. Characterization of a supercoil-dependent S1 sensitive site 5′ to the *Drosophila melanogaster hsp26* gene. Nucleic Acids Res. **14:**9425–9444.
58. **Simon, J. A., and J. T. Lis.** 1987. A germline transformation analysis reveals flexibility in the organization of heat shock consensus elements. Nucleic Acids Res. **15:**2971–2988.
59. **Soeller, W., and T. Kornberg (University of California, San Francisco).** Personal communication.
60. **Spradling, A. C.** 1986. P element-mediated transformation, p. 175–197. *In* R. D. Roberts (ed.), *Drosophila*: a practical approach. IRL Press, Oxford.
61. **Steinhaur, W. R., R. C. Walsh, and L. J. Kalfayan.** 1989. Sequence and structure of the *Drosophila melanogaster* ovarian tumor gene and generation of an antibody specific for the ovarian tumor protein. Mol. Cell. Biol. **9:**5726–5732.
62. **Straka, C., and W. Hörz.** 1991. A functional role for nucleosomes in the repression of a yeast promoter. EMBO J. **10:**361–368.
63. **Taylor, I. A. C., J. L. Workman, T. J. Schuetz, and R. E. Kingston.** 1991. Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. Genes Dev. **5:**1285–1295.
64. **Teare, J., and S. C. R. Elgin.** Unpublished data.
65. **Thomas, G. H., and S. C. R. Elgin.** 1988. Protein/DNA architecture of the DNase I hypersensitive region of the *Drosophila hsp26* promoter. EMBO J. **7:**7291–2201.
66. **Topol, J., D. M. Ruden, and C. S. Parker.** 1985. Sequences required for *in vitro* transcriptional activation of a *Drosophila hsp70* gene. Cell **42:**527–537.
67. **Travers, A. A.** 1992. The reprogramming of transcriptional competence. Cell **69:**573–575.
68. **Vinson, C. R., K. L. LaMarco, P. F. Johnson, W. H. Landschulz, and S. L. McKnight.** 1988. *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. Genes Dev. **2:**801–806.
69. **Xiao, H., and J. T. Lis.** 1988. Germline transformation used to define key features of heat-shock response elements. Science **239:**1139–1142.
70. **Wallrath, L. L., and S. C. R. Elgin.** Unpublished data.
71. **Workman, J. L., and R. E. Kingston.** 1992. Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. Science **258:**1780–1784.
72. **Wu, C.** 1984. Two protein-binding sites in chromatin implicated in the activation of heat shock genes. Nature (London) **309:**229–234.
73. **Zimarino, V., and C. Wu.** 1987. Induction of sequence-specific binding of *Drosophila* heat shock activator protein without protein synthesis. Nature (London) **237:**727–730.