Genomic footprinting of the *hsp70* and histone H3 promoters in *Drosophila* embryos reveals novel protein–DNA interactions

Janet A. Weber and David S. Gilmour*

Department of Molecular and Cell Biology, Center for Gene Regulation, The Pennsylvania State University, University Park, PA 16802, USA

Received March 6, 1995; Revised and Accepted June 16, 1995

ABSTRACT

The transcriptional potential of the *hsp70* heat shock gene promoter is established prior to induction by stress. It has been shown previously that the TBP subunit of TFIID is associated with the TATA element and that RNA polymerase II is paused downstream from the transcription start site. In order to identify new interactions involved in establishing this potentiated state, a detailed analysis of the molecular architecture of a single copy of the *hsp70* promoter was performed. A suitably marked promoter was stably integrated using P-element-mediated transformation so as to overcome any ambiguity that might be associated with analyzing the five copies of the endogenous gene. Genomic footprinting using DNase I revealed two previously unidentified interactions. First, the GAGA element located at –120 is protected by protein. Secondly, the pattern of DNase I cleavage in the vicinity of the transcription start is found to bear significant similarity to the pattern associated with binding of purified TFIID. Noting that purified GAGA factor and TFIID interact similarly with the *hsp70* and H3 promoters, the architecture of the endogenous H3 promoter was analyzed to determine what interactions might be needed to establish a potentiated state containing a paused polymerase. Despite the detection of TFIID and GAGA on the H3 promoter, no paused polymerase is evident. In addition, no proteins appear to interact with the transcription start. These results suggest that the GAGA factor and TFIID are not sufficient to establish a potentiated state containing paused polymerase and that TFIID interactions downstream from the TATA element could be important for pausing.

INTRODUCTION

Work in recent years has established that chromatin structure plays a dominant role in transcriptional regulation. Biochemical studies have led to the notion that packaging DNA with histones should lead to transcriptional repression in many circumstances

(1). *In vivo* studies are largely consistent with this model (2). Mechanisms by which the cell overcomes or prevents a repressive chromatin structure are likely to be a key element in regulating gene expression. The *hsp70* heat shock gene from *Drosophila* has emerged as a model system for how a cell might prevent nucleosomes and other chromosomal proteins from interfering with the activation process. Under normal growth conditions the gene is inactive. Within 60 s of a heat shock treatment the genes are fully induced in virtually all cells. Transcriptional activation requires the association of heat shock factor (HSF) with sequences located upstream of the transcription start (3). HSF resides in the cell in a latent form and acquires DNA binding activity in response to a variety of stress treatments (4).

Several promoter interactions have been identified that are likely to be involved in establishing the transcriptional potential of the *hsp70* promoter under normal growth conditions when the gene is not expressed. The promoter region is hypersensitive to DNase I digestion in isolated nuclei (5), suggesting that the promoter region could be free of nucleosomes (6,7). RNA polymerase II has initiated transcription, but paused at a point 17–37 nucleotides (nt) downstream from the transcription start (8, and references therein). Genomic footprinting analysis with potassium permanganate has revealed the presence of the TBP subunit of TFIID on the TATA element in tissue culture cells (9). The presence of TBP is essential if RNA polymerase II is to be recruited.

Analyses of mutant promoters have identified two regions of the promoter that are important for establishing the transcriptional potential. These mutations decrease the level of paused polymerase (10). A mutation in a GAGA element located at –70 causes a 5-fold reduction in the level of RNA polymerase that pauses on a promoter construct spanning the region from –89 to +62. Other GAGA elements are present in the *hsp70* promoter region, but their contributions *in vivo* have not been assessed. The factor that binds the GAGA element, called the GAGA factor, has recently been shown to disrupt nucleosome structure in reconstituted chromatin in a way that depends on ATP hydrolysis (11). Hence, it is possible that the GAGA factor disrupts the nucleosome *in vivo* to render the promoter accessible to other proteins. A second region of the promoter has been implicated by 3' deletions. The stepwise deletion of sequences from +62 to –12 caused a graded decrease in the level of paused polymerase associated with the transformed

* To whom correspondence should be addressed

constructs (10). We have recently shown that much of this region downstream from the TATA element contributes significantly to the binding of purified TFIID (12,13). Hence, these 3' deletions may be interfering with the association of TFIID.

A rudimentary model for potentiation of the *hsp70* promoter can be formulated from the current data. Binding of TFIID and the GAGA factor might provide the foundation for recruiting the general transcription factors and RNA polymerase II to the promoter under normal growth conditions. The polymerase initiates transcription, but pauses by some unknown mechanism. This organization leaves the HSF binding sites exposed, so that upon stress, activated HSF can bind without any need to reorganize the chromatin structure.

To gain more insight into how the *hsp70* promoter is set up in the cell, we have analyzed the molecular architecture of an *hsp70* promoter construct that has been transformed into flies on a P-element. Genomic footprinting with DNase I was used on this promoter for the first time to seek out novel interactions. We were particularly interested in the region around the transcription initiation site, as this has been found to be important in TFIID binding (13). This region has been largely overlooked in genomic footprinting studies of other genes. Finally, we extended our analysis to the H3 promoter, since previous results suggested that this promoter would share many of the same interactions associated with *hsp70* (13,14). Despite the finding that GAGA factor and TFIID are associated with these two promoters *in vivo*, we find that only *hsp70* has a paused polymerase. We conclude that TFIID and GAGA factor are important for setting up the promoters, but that they are not sufficient to give a paused polymerase.

MATERIALS AND METHODS

Plasmids

All of the *hsp70* subclones were originally derived from the plasmid 132E3, which contains two tandem copies of the *hsp70* gene from locus 87C1 (15). The *Xba*I fragment spanning the region from -254 to +1847 was subcloned into the *Xba*I site in pUC13 and the resulting plasmid was called *hsp70/87C*. A hybridization probe specific for the endogenous gene was derived from this clone.

The *hsp70* promoter transformed into the flies was introduced on a plasmid called 70ZT (-194/+84). A generalized map of this construct is shown in Figure 1. This plasmid was prepared in several steps, a detailed description of which is available upon request. First, the transformation vector Carnegie 20 (16) was modified to generate a general vector called Car20-lacZT.2. Car20-lacZT.2 contains the *Escherichia coli* β -galactosidase (β -gal) gene from plasmid pMC1871 (17), beginning at codon 7 and extending downstream to a *Bam*HI site. A synthetic oligonucleotide containing a translation start and a ribosome binding site was placed upstream of the β -gal sequence. The sequence of the region upstream of the β -gal coding region was GTCGACACGCGGCCGACAATGCTATTGGAATCGATCC, where the first six underlined nucleotides represent a unique *Sal*I site in Car20-lacZT.2 and the underlined triplet represents the translational initiation codon for β -gal. The *hsp70* transcription termination region derived from 132E3 as a 250 bp *Xho*I-*Sal*I fragment was inserted downstream of the *E. coli* β -gal sequences.

To generate 70ZT (-194/+84) the region of the *hsp70* promoter spanning from -194 to +84 was inserted into the *Sal*I site of

Figure 1. Sequence elements in the *hsp70* promoter and transformation vector. (A) The top portion of the figure shows the sequence elements identified in the *hsp70* promoter. These are a TATA box, four HSEs and several GAGA factor binding sites. The -89 position is just upstream of HSE 2. The lower portion of the figure shows the transformation vector. The direction of transcription for *Xdh* (*rosy*) and β -gal are indicated by arrows. Approximately 400 bp separates the 3'-end of *Xdh* from the *Sal*I site. P represents the location of the P-element ends. Also shown are the *Eco*RV sites used in the Southern blot analysis. (B) Sequences of the primers used in LMPCR and the complementary DNA strands are shown for the endogenous promoter (*hsp70*), the transformed promoter (-194/+84) and the H3 promoter. The first primer in each series anneals closest to the 3'-end of the complementary strand. Positions relative to the transcription start are given above the complementary strands. Annealing temperatures for each series are 60, 63 and 63°C for *hsp70*, 52, 45 and 49°C for -194/+84 and 60, 45 and 52°C for H3. Buffer conditions are given in Materials and Methods.

Car20-LacZT.2. The source of this promoter fragment was a subclone called *hsp70* -190 XBS, which has an *Xho*I site upstream and a *Sal*I site downstream of the *hsp70* promoter. The promoter originated from the 132E3 clone containing *hsp70* genes from 87C1 (15).

Transformed fly lines

Transformed lines carrying *hsp70* promoter constructs were generated using P-element-mediated transformation (16). A mixture containing transformation vector DNA and a plasmid coding for transposase (p13 wings clipped, kindly provided by

Susan Abmayr) was injected into embryos of *cn⁻; ry⁵⁰⁶* flies which carry mutations in two eye color genes, *cinnabar* and *rosy*. (The product of the *rosy* gene is xanthine dehydrogenase, Xdh.) Transformants were identified by eye color and independent lines confirmed by back-crossing to *cn⁻; ry⁵⁰⁶*. Insertion of the transgene as a single copy and verification of independent lines was confirmed by Southern blotting (data not shown).

Promoter hypersensitivity and levels of transcription, both basal and induced, for the -194/+84 independent lines were determined at the beginning of this study. The two lines chosen had identical patterns of hypersensitivity and comparable levels of expression. All lines were maintained on standard cornmeal/yeast food at 25°C. For embryo collection separate stocks of flies were set up in containers with removable trays. All embryos represent 0–16 h overnight collections. They were flash frozen in liquid nitrogen and stored at -80°C until processed. For heat shock embryos on trays were placed in humidified containers at 37°C for 2 h, then rapidly collected and flash frozen.

Nuclei isolation and treatment

Nuclei were isolated from dechorionated embryos following the protocol outlined in Thomas and Elgin (18), except that the buffer used for resuspension and treatment did not contain calcium (1× DNase I buffer is 60 mM KCl, 15 mM NaCl, 0.5 mM DTT, 15 mM Tris-HCl, pH 7.4, 8.6% sucrose, 3 mM MgCl₂). Treatment of nuclei with DNase I (Boehringer Mannheim) was for 3 min at 21°C and was stopped by the addition of EDTA to 12.5 mM. Naked DNA standards were prepared by isolating DNA in bulk from untreated nuclei and treating it under the same conditions but in the presence of 100 µg/ml acetylated bovine serum albumin. Treatment of nuclei with 25 mM potassium permanganate was performed for 30 s at 0°C and was stopped by the addition of EDTA and β-mercaptoethanol to 25 mM and 0.2 M, respectively (9). Naked DNA standards were prepared by using DNA purified from equal aliquots of untreated nuclei and treating them with 25 mM potassium permanganate for 30 s at 0°C. The observed yield of DNA from nuclei isolated from 100 mg embryos was ~20 µg.

Analysis of genomic DNA

A portion of each sample was digested with the appropriate enzymes and run on agarose gels in 1× Tris-borate-EDTA buffer. The gels were treated with hydrochloric acid (0.12 M), then sodium hydroxide (0.4 M) and blotted onto GeneScreen Plus filters overnight in 0.4 M sodium hydroxide. Filters were rinsed in 0.2 M Tris-HCl, pH 7.5, 2× SSC and air dried. The hybridization mixture contained 5.4× SSC, 9% dextran sulfate, 45% formamide, 0.9% SDS and probes prepared by random priming (19) with a specific activity of at least 10⁸ c.p.m./µg. Hybridizations proceeded overnight at 42°C. Filters were washed in 2× SSC, 0.5% SDS at room temperature and in 0.1× SSC, 0.1% SDS at 55°C.

The protocol for genomic footprinting is essentially the original ligation-mediated PCR method (LMPCR) outlined by Mueller and Wold (20). DNA purified from nuclei or naked DNA was digested either with *Hae*II (for histone H3), with *Xba*I (for endogenous *hsp70*), or with *Bst*NI and *Nsi*II (for the transformed gene) prior to performing LMPCR. A portion was removed to confirm digestion and the remainder phenol extracted and ethanol precipitated (DNase I-treated samples) or precipitated directly

(potassium permanganate-treated samples). The potassium permanganate samples were resuspended in 10% piperidine, heated at 90°C for 30 min and then ethanol precipitated twice. All samples were resuspended in 1 mM EDTA, 10 mM Tris-HCl, pH 7.5, so that there were 100 ng/3 µl (for H3) or 500 ng/3 µl (for *hsp70* or the transformed gene).

Extension of the first genomic primer was performed using Sequenase (USB, version 1.0) following annealing (see Fig. 1B for details). Annealing buffer consisted of 3 µl DNA in 50 mM NaCl, 38 mM Tris-HCl, pH 7.7, and 0.3 pM primer in a volume of 15 µl. Sequenase extension buffer was 6 mM MgCl₂, 6 mM DTT, 62 µM dNTPs, 200 µM EDTA, 40 mM NaCl, 0.5 mM thioredoxin, 32 mM Tris-HCl, pH 7.5, in a final volume of 24 µl. Samples were adjusted for ligation overnight at 15°C using T4 DNA ligase (USB). We had an alternate version of the long linker primer synthesized which had a calculated *T_m* closer to those of the primers used for the transformed gene. The sequence of this primer is 5'-GCGGTGATTATAAAGATCTGAATTC-3'. The short linker primer had the same sequence as that described by Mueller and Wold (20). Following ligation the samples were heated to 70°C, ethanol precipitated using sodium acetate and tRNA and the pellets were resuspended in water on ice. A mixture containing 10× PCR buffer, MgCl₂, dNTPs, primers and *Taq* polymerase (Promega) was added to each sample (final amplification conditions were 50 mM KCl, 10 mM Tris-HCl, pH 9, at 25°C, 0.1% Triton X-100, 5 mM MgCl₂, 400 µM dNTPs, 10 pM each second genomic primer and long linker primer and 2 U *Taq* polymerase in a volume of 100 µl). Primers and annealing temperatures are given in Figure 1B. Each amplification program consisted of equal numbers of cycles with 1 min denaturation and 2 min annealing at the indicated temperatures and either 3 or 4 min extension time at 72°C, for a total of 20 cycles. Labeling of the amplified material was accomplished by the addition of 0.5 pM kinased third genomic primer (in 20 µl amplification buffer without MgCl₂) for four cycles of annealing and extension, with an extension time of 10 min. After the labeling cycles the samples were extracted with chloroform to remove the mineral oil layer, the volume expanded using 2.8 volumes of 260 mM sodium acetate, 10 mM Tris-HCl, 4 mM EDTA, pH 7.4, 0.5 µg yeast tRNA and ethanol precipitated. Each sample pellet was resuspended after a wash with 75% ethanol in sequencing loading buffer and run on a sequencing gel. Positions of bands were mapped by making three comparisons. First, restriction cut markers were included on all gels to provide reference bands. Second, G/A sequencing markers were also included to provide finer resolution. The final assessment of positions comes from comparing the pattern of DNase I cutting detected by LMPCR with the pattern of cutting that we routinely observe for end-labeled *hsp70* DNA fragments.

In vitro footprinting

With the exception of the changes described below, TFIID was immunopurified from a phosphocellulose fraction with anti-TBP monoclonal antibody and the binding of DNA to the immunoprecipitate was performed as previously described (12,13). For the current work the DNA was not radiolabeled prior to binding. The *hsp70* DNA was generated by PCR from the plasmid that was originally injected into the flies to generate the transformed lines; this provided a fragment that spanned from -194 to +84. The H3 DNA was produced by restriction digestion of plasmid DNA

containing the H3 promoter with *Ava*I; this provided a fragment that spanned the region from -130 to +102. Approximately 2 ng each fragment and 1 μ g *Hae*III-cut *E.coli* DNA were combined and then incubated for 1 h at room temperature with immobilized TFIID. The total volume of the reaction was 100 μ l and the binding buffer consisted of 10 mM HEPES, pH 7.6, 5 mM $MgCl_2$, 1 mM DTT, 1 mM PMSF, 80 mM KCl, 10% glycerol. After the 1 h incubation the unbound material was collected and treated with DNase I to provide an unbound control. The Sepharose was washed several times with the binding buffer and then suspended in 50 μ l binding buffer supplemented with 250 ng *Hae*III-cut *E.coli* DNA. Five microliters of DNase I (freshly diluted to 0.01 U/ μ l in 50 mM NaCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 20 mM Tris, pH 7.6, 100 μ g/ml BSA) was added to the slurry of Sepharose and incubated for 40 s at room temperature. The DNase I digestion was stopped by adding 100 μ l 0.5% SDS, 10 mM EDTA. DNA was purified by phenol extraction and ethanol precipitation. The pattern of DNase I cuts for the H3 and the *hsp70* DNA fragments was evaluated by LMPCR, as described for genomic DNA.

RESULTS

We focused our analysis on the molecular architecture of a promoter that spans the region from -194 to +84. Our biochemical characterization of TFIID-*hsp70* promoter interactions had indicated that the region from -194 to +84 is sufficient for association of purified TFIID (12). Sequence-specific interactions are restricted to the region between -40 (the TATA element) and +33 (12,13). Although previous work had shown that 5' deletions to -89 still mediated heat shock induction (21,22), we chose to include the region from -194 to -89 because this region contains several GAGA elements that have been shown to bind the GAGA factor *in vitro* (14), but have not been tested for function *in vivo*. We extended the promoter in the 3' direction to +84 because this region was shown to be sufficient for heat shock induction (23) and to extend just beyond the region required to give normal levels of paused polymerase (10). The entire region from -194 to +84 is sufficiently small that we can apply the LMPCR technique for genomic footprinting protein-DNA interactions in nuclei. Figure 1A summarizes the distribution of sequence elements that contribute to the interactions of heat shock factor, GAGA factor and TFIID and situates the promoter region within the context of the sequences that were transformed into the flies. Figure 1A also outlines the construct that was transformed into the flies. The -194 to +84 region of the *hsp70* promoter was fused to the *E.coli* β -gal gene. The β -gal sequences allow us to distinguish the structure and activity of the transformed promoter from that of endogenous *hsp70* genes. The *hsp70* termination region was placed downstream of the β -gal coding sequences so that a message of a discreet size would be produced.

The transformed promoter performs much like the endogenous genes

Two independent transformants were generated using P-element-mediated transformation. Southern blot analysis showed that each line contained a single copy of the integrated sequence and that the insertions were in different locations (data not shown). It was essential to show first that the region we chose to transform into flies exhibited properties of the endogenous gene. Evaluation of

expression of the *hsp70* promoter construct in these two lines by Northern blot analysis (data not shown) revealed no detectable basal transcription and high levels of induced transcription following heat shock. The -194/+84 transformed promoter, like the endogenous promoter, is properly regulated and strongly induced.

We sought evidence that the chromatin structure of the uninduced transformed gene matched that of the endogenous gene. The transformed promoter was analyzed to determine if it was assembled into a DNase I hypersensitive region. Nuclei were isolated from embryos of the non-heat shocked transformants and then digested with a range of DNase I concentrations. The DNase I hypersensitivity of the transformed promoter and the endogenous promoters was detected using appropriate probes and an indirect end-labeling method (5). As shown in Figure 2A, the region spanning the *hsp70* promoter in the endogenous gene was much more sensitive to DNase I cutting than were the surrounding regions. The hypersensitivity was very similar to that of the transformed gene shown in Figure 2B, implying that they had similar structures. (The presence of multiple parental *Bam*HI bands for the endogenous promoter is due to the presence of multiple copies with differing patterns of restriction sites.) Figure 2C provides a view of a larger region surrounding the *hsp70* sequences of the transformed gene. From this it is clear that the hypersensitivity was restricted to the *hsp70* sequences within the region from -194 to +84. This suggests that there are no interactions in neighboring sequences that are influencing the behavior of the transformed promoter.

Genomic footprinting with DNase I reveals interactions of the GAGA factor and TFIID

Close examination of the cutting pattern within the hypersensitive regions of both the transformed and endogenous promoters shown in Figure 2 clearly revealed patches of reduced sensitivity within the hypersensitive region that may reflect the binding of specific proteins. The protected patch just downstream of -50 could be due to TATA binding protein bound to the TATA element. The partially protected region upstream of -50 could be due to association of the GAGA factor.

To obtain a more detailed view of the protein interactions occurring on the promoter, we examined the DNase I cutting pattern at single nucleotide resolution. This genomic footprinting analysis could be performed on the transformant without ambiguity because the junction between the *hsp70* sequences and the β -gal sequences provided a specific starting point for the LMPCR technique that was used to detect the DNase I cuts. Figure 3B shows the genomic footprinting results for the transformed *hsp70* promoters in one of the fly lines. Factor binding is readily apparent on the GAGA element located at -120 and over the TATA element. The upstream boundary of the footprint is partially defined by the DNase I cut around -160. Unlike an *in vitro* footprinting analysis, however, sharp demarcation of the footprint on the upstream side is not anticipated, because the DNA flanking the promoter is presumably packaged with histones in a relatively DNase I-resistant form (see Fig. 2). There is also evidence for protection over the GAGA element located just downstream of HSE 2, but the paucity of DNase I cuts in this region tempers the strength of this conclusion.

We were particularly interested in analyzing the region downstream from the TATA element, because this region has not been scrutinized in other genes and because of the potential for

Figure 2. Southern blot analysis of DNase I hypersensitivity of endogenous and transformed promoters. Shown are: (A) endogenous *hsp70* promoters; (B and C) transformed promoters. The blots in (A) and (B) represent digestion of genomic samples with *Bam*HI and *Eco*RV. This is the same blot probed first with a β -gal-specific probe for the transformant, then stripped and reprobed for the endogenous promoter. The blots in (C) represent genomic samples digested with *Eco*RV and probed first with the β -gal-specific probe (left panel) then stripped and reprobed with a *Xdh*-specific probe (right panel). The positions marked alongside the blots were located by restriction digestion of plasmid DNA containing the endogenous or transformed gene with *Xba*I (–250), *Nru*I (–50), *Hind*III or *Bam*HI (–194), *Pst*I (+89) or *Pvu*II (+200). The probe for *hsp70* is a 1 kb *Pst*I fragment from the transcribed region of *hsp70*. The probe for β -gal is a 620 bp *Hpa*I fragment from the 5' coding region of that gene. The *Xdh* probe is a 1 kb *Eco*RV–*Hind*III fragment taken from Car20-lacZT.2.

overlapping interactions of TFIID and RNA polymerase II. There are clearly differences in the intensity of bands for the purified genomic DNA and DNA cut in nuclei (Fig. 3B). It appeared that certain aspects of the interaction bore significant resemblance to the footprints that we had recently observed for purified TFIID (compare Fig. 3B and C). For example, protection over the TATA element is clearly evident. However, we were disturbed that a hallmark of purified TFIID binding, a hypersensitive site at –44 on the non-transcribed strand, was absent from the genomic footprint. To provide footprinting data for the TFIID complex that could be compared directly with the genomic footprint, we used the LMPCR method to visualize the DNase I cutting pattern formed when purified TFIID was bound to the same *hsp70* promoter sequences as were being analyzed in the transformants.

Figure 3. DNase I footprinting of *hsp70* promoters. (A) The non-transcribed strand of the endogenous promoter. (B) The non-transcribed strand of the transformed promoter. DNA denotes naked DNA standards, while Nuclei indicates samples from nuclei. Nuclei were treated with 14 (A) and 15 U/ml (B) DNase I. The positions indicated for both promoters are –4, +1, +4, +15 and +28. Protection is evident for positions +1, +4 and +15, while sensitivity is shown in the region of –18 and at –4 and +28. (C) Interaction of immunopurified TFIID with the *hsp70* promoter on the non-transcribed strand from –50 to +50. The DNase I cutting pattern for the bound (B) and unbound fractions (F) are shown. M indicates restriction fragments amplified to serve as markers.

Figure 3C shows the resulting footprint produced by immunopurified TFIID. The protection produced by the purified protein is extensive and it clearly spans the region from the TATA element to +35. The hypersensitive site at –44 that had become a hallmark for us in other studies with end-labeled DNA (12) was noticeably absent from this material analyzed by LMPCR. Apparently, cuts at this site are not amplified; this type of idiosyncrasy is not uncommon for the LMPCR method (24).

When the cutting patterns for naked genomic DNA and the naked DNA used in the binding assay with TFIID were compared there was a virtually one-to-one correspondence. This confirms that a direct comparison between the footprints should yield valid information. When the footprint of purified TFIID was compared with the genomic footprint there were significant similarities. Certain bands associated with purified TFIID binding penetrate the footprinted region from –18 to +28 and the intensities of these bands relative to each other are identical to those observed in nuclei (Fig. 3B and C). As these contacts downstream of the TATA

element are not found in association with the purified TBP subunit (12), these contacts could be due to other subunits of TFIID.

For comparison with the transformed gene we also analyzed the endogenous *hsp70* genes. The primers and conditions for PCR amplification were chosen to enrich for the three copies of the *hsp70* gene located at the 87C locus; our transformed promoter was derived from one of these copies. Figure 3A shows the resulting genomic footprint. There was a strong resemblance to the transformed promoter. Protection over the GAGA elements located at -120 and the TATA element were clearly evident. In addition, protection of the GAGA element located at -70 was clearer for the endogenous genes than was evident for the transformed promoter. The relative intensity of several of the bands around the initiation site also resembled those of the transformed gene and the purified TFIID. Hence, although the comparison is being made to at least three copies of the endogenous gene, the transformant appears to be organized in a very similar manner.

Paused polymerase is present on the transformed and endogenous promoters in nuclei preparations from *Drosophila* embryos

We were unable to detect any clear indication of a paused polymerase in the genomic footprints generated with DNase I. The paused polymerase has been visualized in *Drosophila* tissue culture cells by treating the intact cells or isolated nuclei with potassium permanganate (9). Potassium permanganate reacts preferentially with thymines located in single-stranded regions and such regions are found in the transcription bubble associated with a paused polymerase. Figure 4A and B shows the pattern of reactivity for the endogenous and transformed promoters respectively in nuclei isolated from embryos. The thymine residues located at +22 and +30 are clearly hyper-reactive to potassium permanganate. This corresponds precisely with results previously reported for *Drosophila* tissue culture cells and is consistent with a number of different studies that have mapped the paused polymerase to the region between +17 and +37 (see 8, and references therein). Direct analysis of the transcripts produced by the paused polymerase has previously shown that the polymerase pauses primarily at two locations. This heterogeneity may explain why the polymerase is not evident in the genomic footprint produced by DNase I.

Analysis of the H3 promoter suggests that interactions of the GAGA factor and TFIID with the promoter are not sufficient to assemble a paused RNA polymerase

The only interactions that were readily apparent in the genomic footprinting analysis of the *hsp70* promoter were those of the GAGA factor and TFIID. This raised the possibility that these two proteins might be the only two that must interact with specific sequences in the promoter to cause assembly of a paused RNA polymerase. We decided to test this hypothesis by examining the histone H3 promoter. A previous genomic footprinting analysis had indicated that a large GAGA element and the TATA element were occupied by protein *in vivo* (14). Hence, the histone H3 promoter appeared to be organized *in vivo* in a fashion that was similar to *hsp70*. Moreover, the interactions between purified TFIID and the H3 promoter *in vitro* were similar to those found for the *hsp70* promoter (13). Purified TFIID made extensive contact extending from the TATA element downstream to +35

Figure 4. A transcription bubble is detected on the transformed promoter. On the left (A) are two lanes showing potassium permanganate-treated naked DNA (D) and nuclei DNA (N) amplified to show the endogenous *hsp70* promoter. On the right (B) are two lanes from the same nuclei preparation amplified to show the transformed promoter.

and recognition of the promoter depended on specific interactions in the TATA element, the start site and at least two regions downstream from the transcription start.

A genomic footprinting analysis of the H3 promoter was performed to provide a view of interactions both upstream and downstream from the transcription start; previous analyses had focused on the GAGA element and the TATA element (14). Figure 5A shows clear evidence for both TFIID and GAGA interactions. Direct comparison between the genomic footprint and the footprint with purified TFIID reveals striking similarities in the vicinity of the TATA element. The TATA element itself is protected, but it is flanked on each side by hypersensitive sites and the accessible site at -4 is also evident. Unlike the *hsp70* promoter, however, there is little similarity in the interactions made at the transcription start and further downstream. This entire region appears to be accessible to DNase I in nuclei. In particular, a cut at +1 is strongly protected by purified TFIID, but no protection of this site is evident for the genomic footprint. A difference between the cutting patterns in naked DNA and nuclei can be discerned in the region between +20 and +30 (compare lanes 1 and 6). This suggests a specific protein-DNA interaction, but the cutting pattern does not match that with purified TFIID. As will be shown, the interaction is also unlikely to be a paused polymerase, so it remains to be characterized.

The genomic footprint shows that TFIID and GAGA factor are present. To determine if this is accompanied by a paused polymerase, we analyzed the sensitivity of the downstream region to potassium permanganate. Figure 6 shows that none of the thymines in this region are hyper-reactive, indicating that polymerase is not paused in this region. These DNA samples were

Figure 5. Interactions on the *Drosophila* H3 promoter. (A) DNase I footprinting of the H3 promoter in nuclei from non-heat shocked embryos. Nuclei were treated with 60, 80 or 100 U/ml DNase I. (B) DNase I footprinting of immunopurified TFIID on the H3 promoter. As for Figure 5, the upward pointing triangles indicate sensitivity to DNase I and the downward pointing triangles indicate protection. The positions indicated for the genomic footprint are -40, -18 and -4; those indicated for the TFIID footprint are -40, -18 and -4.

derived from the same ones analyzed for *hsp70* (see Fig. 4), indicating that the lack of hyper-reactive thymines is not the result of a flaw in the experimental manipulations. Rather, it appears that the presence of TFIID and the GAGA factor is not sufficient to assemble a paused polymerase.

DISCUSSION

This is the first genomic footprinting analysis to be performed on an *hsp70* promoter that has been transformed into *Drosophila*. A genomic footprinting analysis was undertaken to identify novel interactions that could contribute to the process that establishes the transcriptional potential of the *hsp70* promoter. Two protein interactions are evident. Binding at the GAGA element located at -120 indicates that the GAGA factor is bound in this region. There are other GAGA elements nearby, but clear protection over these elements is not observed. Direct comparison between the footprint with purified TFIID and the genomic footprint indicates that TFIID could be interacting with the TATA element and the initiator. Potassium permanganate footprinting indicates that polymerase is paused in the region encompassing +22 and +30, but no evidence for this interaction is apparent when DNase I is used as a probe. This may be because the polymerase appears to be paused primarily at two places, so a heterogeneous structure could result (8). This heterogeneity could also explain why the DNase I footprint around the transcription start site in nuclei is not as complete as it appears in the case with purified TFIID.

Figure 6. Paused polymerase is not detected on the H3 promoter. Potassium permanganate-treated naked DNA and genomic DNA taken from the same nuclei preparation as in Figure 4 amplified to show the H3 promoter.

The GAGA elements in the region between -120 and -160 have not been subjected to *in vivo* studies. The finding that the GAGA element at -120 is occupied by factors in nuclei is consistent with the hypothesis that binding of the GAGA factor in this region contributes to assembly of the *hsp70* promoter structure prior to heat shock induction. Further support for this has recently been provided by biochemical studies involving purified GAGA factor and reconstituted chromatin (11). GAGA factor was found to disrupt the nucleosome structure in the vicinity of its binding sites in chromatin that had been reconstituted in extracts from pre-blastoderm embryos. This disruption requires ATP and is most evident when the GAGA factor is included at the beginning of the chromatin assembly reaction. Deletion of the GAGA elements located between -120 and -160 severely attenuated the level of nucleosome disruption. We have recently found that deletion of the sequences between -194 and -89 shortens the upstream boundary of the hypersensitive region to ~-90 in transformed flies (Weber and Gilmour, unpublished observation). Hence, the GAGA element in the -120 to -160 region may function to extend the open chromatin structure to heat shock elements normally located upstream from this region.

Additional deletions indicate that the GAGA element at -70 also mediates levels of disruption by GAGA factor in reconstituted chromatin (11). We have detected binding on this element for the endogenous *hsp70* promoter, but the protection on the transformed promoter is not strongly evident in our analysis. Its contribution to setting up the *hsp70* promoter *in vivo* is indicated by the results of a mutational analysis of promoters spanning the region from -89 to +62 (10). In this context, mutation of the GAGA element at -70 caused a 5-fold decrease in the level of paused polymerase. Hence, it is likely that GAGA elements located in both the region from -120 to -160 and the region around -70 contribute to assembling the chromatin structure of the *hsp70* promoter.

The GAGA protein acts *in vitro* to reposition nucleosomes on reconstituted chromatin (11). In addition, mutations in the GAGA factor gene enhance position-effect variegation, apparently by failing to establish or maintain open chromatin (25). GAGA elements in the *hsp26* promoter are important in establishing the DNase I hypersensitive sites for that promoter (18,26). Since the heat shock factor does not participate in the formation of open chromatin on the uninduced promoter, these binding sites are essential only for induced expression (27,28). We expect that the GAGA factor will play a similar structural role on the *hsp70* promoter.

The second factor that is clearly evident in the genomic footprint is TFIID. The contact with the TATA element is consistent with other studies (9,29). The data presented here, however, is the first to make the comparison between the gene in *Drosophila* and TFIID purified from *Drosophila*. Our results also identify possible TFIID interactions in the vicinity of the transcription start. There is clearly a difference between the intensity of the DNase I cuts around the transcription start for purified genomic DNA and DNA in nuclei. That this is due to TFIID is suggested by the congruence between the DNase I cutting pattern seen for nuclei and for purified TFIID. This congruence is reproducibly observed for both the endogenous and transformed gene.

The paused polymerase is one of the most intriguing aspects of the potentiated *hsp70* promoter. The results of genomic footprinting suggested that DNA interactions with TFIID and GAGA factor might be sufficient to assemble the promoter with a paused polymerase. Indeed, this hypothesis is supported by studies of mutant promoters that have been transformed into flies (10). We have tested this hypothesis by analyzing the H3 promoter, with which GAGA and TFIID were known to interact *in vivo*. Both GAGA factor and TFIID are clearly evident on the H3 promoter and yet no paused polymerase is observed. Thus the mere presence of GAGA factor and TFIID is not sufficient to assemble a paused polymerase.

It appears that the combined interactions of GAGA factor and TFIID may be sufficient to prevent binding of a nucleosome, as the promoters of both *hsp70* and H3 are assembled into DNase I hypersensitive regions (5,31,32). What additional component might be important to recruit polymerase into the paused state? Unfortunately, no clues are provided by considering the collection of promoters known to exhibit a paused polymerase in *Drosophila* (9). In addition to heat shock genes, paused polymerase is detected on some non-heat shock genes and on some genes that lack a TATA element (33). A hypothesis that has recently emerged suggests that the phosphorylation state of the CTD of RNA polymerase II could be important. The CTD of the paused polymerase is predominantly dephosphorylated (34). This form of the CTD binds TBP (35), suggesting that the CTD might act as a flexible tether. Our genomic footprinting data provide support for an alternative hypothesis. Perhaps the conformation of the TFIID complex *in vivo* dictates whether or not pausing will occur. *In vivo* TFIID appears to interact with the transcription start for *hsp70*, but not for H3. It is intriguing that cross-linking analysis of isolated TFIID indicates that TFIID contacts the *hsp70* promoter at +46, a position located immediately downstream from the paused polymerase (Li and Gilmour, unpublished observation; 30). Specific DNA interactions could dictate particular conformations of TFIID that pause polymerase and others that do not.

The results described here provide the first application of genomic footprinting to the analysis of a promoter that has been

transformed into *Drosophila*. Future application of these approaches to mutant promoters should provide a deeper understanding of the function of specific components in the transcription process.

ACKNOWLEDGEMENTS

We would like to thank Susan Abmayr for her generous gift of plasmid p13wc. We also thank Robert Wienzierl and Robert Tjian for the monoclonal antibody against TBP. This work was supported by research grant NP-721 from the American Cancer Society, research grant GM47477 from the NIH and an American Cancer Society Junior Faculty Award to DSG.

REFERENCES

- Gilmour, D.S., Buchman, A.R. and Workman, J.L. (1994) In Conaway, R.C. and Conaway, J.W. (eds), *Transcription: Mechanisms and Regulation*. Raven Press, New York, NY, Vol. 3, pp. 515–533.
- Grunstein, M. (1990) *Trends Genet.*, **6**, 395–400.
- Lis, J. and Wu, C. (1993) *Cell*, **74**, 1–4.
- Rabindran, S.K., Haroun, R.I., Clos, J., Wisniewski, J. and Wu, C. (1993) *Science*, **259**, 230–234.
- Wu, C. (1980) *Nature*, **286**, 854–860.
- Karpov, V.L., Preobrazhenskaya, O.V. and Mirzabekov, A.D. (1984) *Cell*, **36**, 423–431.
- Nacheva, G.A., Guschin, D.Y., Preobrazhenskaya, O.V., Karpov, V.L., Ebralidse, K.K. and Mirzabekov, A.D. (1989) *Cell*, **58**, 27–36.
- Rasmussen, E.B. and Lis, J.T. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7923–7927.
- Giardina, C., Perez-Riba, M. and Lis, J.T. (1992) *Genes Dev.*, **6**, 2190–2200.
- Lee, H., Kraus, K.W., Wolfner, M.F. and Lis, J.T. (1992) *Genes Dev.*, **6**, 284–295.
- Tsukiyama, T., Becker, P.B. and Wu, C. (1994) *Nature*, **367**, 525–532.
- Emanuel, P.A. and Gilmour, D.S. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 8449–8453.
- Pumell, B.A., Emanuel, P.A. and Gilmour, D.S. (1994) *Genes Dev.*, **8**, 830–842.
- Gilmour, D.S., Thomas, G.H. and Elgin, S.C.R. (1989) *Science*, **245**, 1487–1490.
- Schedl, P., Artavanis-Tsakonas, S., Steward, R., Gehring, W.J., Mirault, M.-E., Goldschmidt-Clermont, M., Moran, L. and Tissieres, A. (1978) *Cell*, **14**, 921–929.
- Rubin, G.M. and Spradling, A.C. (1982) *Science*, **218**, 348–353.
- Casadaban, M.J., Martinez-Arias, A., Shapira, S.K. and Chou, J. (1983) *Methods Enzymol.*, **100**, 293–308.
- Thomas, G.H. and Elgin, S.C.R. (1988) *EMBO J.*, **7**, 2191–2201.
- Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
- Mueller, P.R. and Wold, B. (1989) *Science*, **246**, 780–786.
- Dudler, R. and Travers, A.A. (1984) *Cell*, **38**, 391–398.
- Simon, J.A., Sutton, C.A., Lobell, R.B., Glaser, R.L. and Lis, J.T. (1985) *Cell*, **40**, 805–817.
- Cohen, R.S. and Meselson, M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5509–5513.
- Garity, P.A. and Wold, B.J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1021–1025.
- Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H. and Karch, F. (1994) *Nature*, **371**, 806–808.
- Glaser, R.L., Thomas, G.H., Siegfried, E., Elgin, S.C.R. and Lis, J.T. (1990) *J. Mol. Biol.*, **211**, 751–761.
- Lu, Q., Wallrath, L.L., Allan, B.D., Glaser, R.L., Lis, J.T. and Elgin, S.C.R. (1992) *J. Mol. Biol.*, **225**, 985–998.
- Lu, Q., Wallrath, L.L., Granok, H. and Elgin, S.C.R. (1993) *Mol. Cell. Biol.*, **13**, 2802–2814.
- Wu, C. (1985) *Nature*, **317**, 84–87.
- Sypes, M.A. and Gilmour, D.S. (1994) *Nucleic Acids Res.*, **22**, 807–814.
- Samal, B., Worcel, A., Louis, C. and Schedl, P. (1981) *Cell*, **23**, 401–409.
- Worcel, A., Gargiulo, G., Jessee, B., Udvardy, A., Louis, C. and Schedl, P. (1983) *Nucleic Acids Res.*, **11**, 421–439.
- Rougvié, A.E. and Lis, J.T. (1990) *Mol. Cell. Biol.*, **10**, 6041–6045.
- O'Brien, T., Hardin, S., Greenleaf, A. and Lis, J.T. (1994) *Nature*, **370**, 75–77.
- Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D. and Aloni, Y. (1992) *Cell*, **69**, 871–881.