View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Washington University St. Louis: Open Scholarship

Washington University in St. Louis

Washington University Open Scholarship

Biology Faculty Publications & Presentations

Biology

2-1990

Optimal heat-induced expression of the Drosophila hsp26 gene requires a promoter sequence containing (CT)n.(GA)n repeats

R Glaser

Graham H. Thomas
Washington University in St. Louis

Ester Siegfried
Washington University in St. Louis

Sarah C.R. Elgin
Washington University in St. Louis, selgin@wustl.edu

J Lis

Follow this and additional works at: https://openscholarship.wustl.edu/bio_facpubs

 Part of the Biology Commons

Recommended Citation

Glaser, R; Thomas, Graham H.; Siegfried, Ester; Elgin, Sarah C.R.; and Lis, J, "Optimal heat-induced expression of the Drosophila hsp26 gene requires a promoter sequence containing (CT)n.(GA)n repeats" (1990). *Biology Faculty Publications & Presentations*. 224.
https://openscholarship.wustl.edu/bio_facpubs/224

This Article is brought to you for free and open access by the Biology at Washington University Open Scholarship. It has been accepted for inclusion in Biology Faculty Publications & Presentations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

J. Mol. Biol. (1990) 211, 751–761

Optimal Heat-induced Expression of the *Drosophila hsp26* Gene Requires a Promoter Sequence Containing $(CT)_n \cdot (GA)_n$ Repeats

Robert L. Glaser[1]†, Graham H. Thomas[2]‡, Esther Siegfried[3]§, Sarah C. R. Elgin[2] and John T. Lis[1]

[1]*Section of Biochemistry, Molecular and Cell Biology Cornell University, Ithaca, NY 14851, U.S.A.*

[2]*Department of Biology and*
[3]*Molecular Biology and Biochemistry Program Division of Biology and Biomedical Sciences Washington University, St. Louis, MO 63130, U.S.A.*

(Received 23 June 1989; accepted 13 October 1989)

We report here the analysis of the sequence requirements for the heat-induced expression of the *Drosophila melanogaster hsp26* gene using germline transformation. Heat-induced expression is augmented fivefold by a homopurine/homopyrimidine region from −85 to −134 that is devoid of heat-shock elements but contains numerous (dC-dT)·(dG-dA) repeats. Sequences within this interval have been shown to assume a nuclease S_1-hypersensitive structure *in vitro*. In this paper, we extend those *in vitro* observations, demonstrating that the S_1-hypersensitive structure is triple-helical *H*-DNA formed by a symmetric (dC-dT)·(dG-dA) sequence. Thus, the sequences that form *H*-DNA *in vitro* are also required *in vivo* for optimal *hsp26* transcription. However, mutational analysis and diethylpyrocarbonate modification experiments in isolated nuclei suggest that the (dC-dT)·(dG-dA) sequence does not form *H*-DNA *in vivo* and argue against a role for *H*-DNA in the heat-induced expression of *hsp26*.

1. Introduction

The transcriptional induction of a discrete family of genes upon heat shock is a universal cellular response to heat stress (Schlesinger *et al.*, 1982; Atkinson & Walden, 1985; Nover, 1987). Extensive work on a number of model systems has revealed that this induction is mediated by the binding of a protein, heat-shock factor (HSF‖), to a specific sequence, the heat-shock element (HSE), located near the promoter of all heat-shock genes (for reviews, see Bienz & Pelham, 1987; Lis *et al.*, 1990). HSEs as upstream promoter elements are necessary and sufficient to make a gene heat inducible (Bienz & Pelham, 1986; Amin *et al.*, 1988; Xiao & Lis, 1988). These studies do not preclude, however, the involvement of other transcriptional elements that may be required for optimal heat-induced levels of transcription. Using germline transformation (Rubin & Spradling, 1982), we present evidence that optimal heat-induced expression from the *Drosophila melanogaster hsp26* gene requires, in addition to HSEs, a short region of sequence devoid of HSEs but containing numerous (dC-dT)·(dG-dA) repeats.

The dinucleotide repeat (dC-dT)·(dG-dA), as well as other homopurine/homopyrimidine repeats, occurs frequently in eukaryotic genomes (Birnboim *et al.*, 1979; Behe, 1987; Manor *et al.*, 1988), and is often found in the vicinity of gene promoters including that of the rat preproinsulin II gene and the SV40 72 bp repeat (Evans *et al.*, 1984), the human c-*myc* gene (Boles & Hogan, 1987), and the heat-shock genes of *Drosophila* (Mace *et al.*, 1983;

† Present address: Department of Embryology, Carnegie Institution of Washington, Baltimore, MD 21210, U.S.A.

‡ Present address: Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138, U.S.A.

§ Present address: Department of Genetics, Harvard Medical School, Boston, MA 02115, U.S.A.

‖ Abbreviations used: HSF, heat-shock factor; HSE, heat-shock element; SV40, simian virus 40; bp, base-pair(s); DEPC, diethylpyrocarbonate; CPRG, chlorophenol red/β-D-galactopyranoside.

0022-2836/90/040751-11 $03.00/0

© 1990 Academic Press Limited

Siegfried *et al.*, 1986). These repeats possess mirror image symmetry and, under appropriate conditions of high superhelicity and/or low pH, undergo a structural transition *in vitro* to form a unique structure called *H*-form DNA (for a comprehensive review, see Wells *et al.*, 1988, and references therein). This structure consists of a normal double helix to which a third strand is bound within the major groove *via* Hoogsteen base-pairing. The compliment of the Hoogsteen base-paired strand is single-stranded and S_1 nuclease-sensitive (see Fig. 4, below, for example).

Selleck *et al.* (1984) demonstrated the presence of several S_1-hypersensitive sites *in vitro* in DNA from the 67B locus of *D. melanogaster*, which contains four heat-shock genes including *hsp26*. At higher resolution, Siegfried *et al.* (1986) localized a small region of DNA responsible for the S_1-hypersensitive structure observed specifically in the *hsp26* promoter, and demonstrated that these sequences mediate this transition in a pH-dependent manner. In this paper, we extend these observations, demonstrating that this S_1-hypersensitive structure in the promoter of *hsp26* is indeed a triple helix.

In the case of the chicken *β^A-globin* gene (Larsen & Weintraub, 1982) and the chicken *α2(1) collagen* gene (Merlino *et al.*, 1983; McKeon *et al.*, 1984), S_1-hypersensitive sites formed *in vitro* by homopurine/homopyrimidine sequence elements appear to be correlated with S_1-hypersensitive sites observed *in vivo*. On the basis of these observations, it has been speculated that triple-helical DNA exists *in vivo* and may play a role in transcriptional regulation. However, a test of this hypothesis *in vivo* has been lacking. In this regard, we present experiments that directly test if the triple helix formed *in vitro* by sequences of the *hsp26* promoter exists *in vivo*, and if it is this structure that is contributing to the transcriptional activity of *hsp26*.

2. Experimental Procedures

(a) *DNA constructs*

The numerical assignment of nucleotides is based on the *hsp26* sequence and transcription start site as determined by Ingolia & Craig (1981). Constructs considered wild-type with respect to heat-induced expression contained upstream sequences to an *Xba*I site located at position −351. An *Xba*I–*Sal*I DNA fragment containing an *hsp26-lacZ* fusion gene with 350 bp of upstream sequence was subcloned from pMC1871.26 (Glaser *et al.*, 1986) into vector pΔZX (Xiao & Lis, 1988) making pX^{u}S26Z. An *Xho*I–*Sal*I fragment from pX^{u}S26Z was then subcloned into the transformation vector Car20T (called c70T1 by Xiao & Lis, 1988) forming the transformation plasmid cP-351. The 5′ junction sequences are GTCGAGGGGG-GATCC TCTAGA.

Transformation plasmid cP-52 was created from cP-351, which was cut with *Xba*I, which cleaves at −350 and −52, and the staggered ends made flush using the Klenow fragment of DNA polymerase (United States Biochemical Corporation). A *Not*I linker was then added by linker tailing (Lathe *et al.*, 1984). The 5′ junction sequences for this and the deletions described below is GGATCCTCTAGGCGGCCGC NNNNN.

The 5′ deletion series was constructed from pX^{u}S26Z. Plasmid pX^{u}S26Z linearized with *Xho*I located in polylinker sequences at the 5′ end of the *hsp26* promoter was treated for various times with exonuclease III (Promega), followed by treatment with nuclease S_1 (Boehringer-Mannheim Biochemicals). Plasmid DNA with various amounts of deleted sequence was gel-purified and treated with the Klenow fragment of DNA polymerase to repair the deleted ends. After another gel purification, *Not*I linkers were added by linker tailing and the subsequent *Not*I-containing deleted plasmids were purified. After preliminary restriction analysis, selected deletions were subcloned as *Not*I–*Sal*I fragments into KS plasmids (Stratagene) for analysis by dideoxy chain termination sequencing (Sanger *et al.*, 1977). *Not*I–*Sal*I fragments of these same deletions were also subcloned into cP-52 plasmids cut with *Not*I and *Sal*I. This digest of cP-52 removes the resident *hsp26-lacZ* gene, so alternative deletions can be cloned in its place. Subcloning deletions into *Not*I–*Sal*I-cut cP-52 created the transformation plasmids cP-272, cP-135, cP-114 and cP-85. Transformation plasmid cPΔCT·GA is cP-351 into which the Δ41.1 deletion (Siegfried *et al.*, 1986) was placed by first removing the wild-type *Xba*I fragment from −48 to −351 and replacing it with an *Xba*I fragment containing the internal deletion. This internal deletion removes sequences from −134 to −85 and inserts the sequence CCTCGAG.

Transformation plasmid cPC·G was constructed using the protocol of Lyamichev *et al.* (1987). Briefly, cPΔCT·GA was cut at the position of the internal deletion with *Xho*I. The staggered *Xho*I termini were filled in using the Klenow fragment of DNA polymerase. Terminal transferase (International Biotechnologies, Inc.) was used to add poly(dC) to a sample of the linear DNA and poly(dG) to another sample of linear DNA. DNA molecules containing comparable extents of poly(dC) and poly(dG) addition were mixed, heated and annealed, forming heteroduplexes, which were transformed directly into *Escherichia coli*. Numerous cPC·G plasmids were characterized by high-resolution restriction analysis, and a construct that recreated wild-type spacing (±5 bp) was used for transformation. Transformation plasmid cPri was constructed like cPC·G but a "random insert" of an appropriately sized fragment of salmon sperm DNA degraded with DNase I was inserted into the *Xho*I site of the internal deletion in cPΔCT·GA. Finally, transformation plasmid cPT/C107 was constructed by oligonucleotide-directed mutagenesis using the gapped duplex, amber selection protocol of Kramer & Fritz (1987). The oligonucleotide used for mutagenesis was 5′ GAAGAGAAGAGGGAGAACGTGCAC 3′, which contains an A to G transition at position −107 (underlined in sequence of oligonucleotide). After creation of the point mutant in bacteriophage M13, an *Xba*I fragment caring the mutation was subcloned into cP-351 from which the wild-type *Xba*I fragment had been deleted.

(b) Drosophila *transformation and CPRG assays*

The *hsp26-lacZ* constructs were introduced into *Drosophila* by P element-mediated germline transformation (Rubin & Spradling, 1982) as described (Simon *et al.*, 1985). Genomic Southern blot analysis (Southern, 1975) was done on all 81 transformant lines to determine the copy number of inserted constructs. We found that 75 of the lines had single inserts and 6 of the lines had 2 inserts. Individual transformant lines were maintained as non-homozygous inbred stocks.

The chlorophenol red/β-D-galactopyranoside (CPRG) assays to determine levels of β-galactosidase activity were modified from methods described by Simon & Lis (1987). Individual males from lines to be analyzed were outcrossed to the $Adh^{fn6}cn;ry^{502}$ injection stock. $rosy^{+}$ female progeny from this cross, all of which were heterozygous for the P-element insertion, were subjected to a 2 h heat shock at 36·5°C with a graded increase in temperature from 29°C to 36·5°C for the first 20 min. After heat shock , duplicate sets of 5 females were homogenized in 50 mM-K_2PO_4 (pH 8·15), 1 mM-$MgCl_2$, 0·25 mM-phenylmethylsulfonyl fluoride, debris was spun out, and a 7-μl sample of the supernatant was added to 1 ml of 1 mM-CPRG in homogenization buffer. After incubation at 37°C for 3 to 5 h (during which time the assay was linear), the A_{574} was determined. The values for the duplicate samples were averaged, and this number was adjusted for copy number when necessary. For Figs 1(a) and 5(b), these values were standardized to the mean of the cP-351 values done in parallel. Standard errors of the mean (S.E.M. $=\sigma/\sqrt{n}$) were calculated.

(c) *Oligonucleotide binding assay*

Radioactively labeled oligonucleotide probes specific for either the homopurine or homopyrimidine strand of the *hsp*26 promoter were isolated from plasmid hsp26S/X (Gilmour *et al.*, 1988). hsp26S/X was cut with *Xba*I cleaving the *hsp26* promoter at position −52. The linearized plasmid preparation was split into 2 portions, one to be radiolabeled using kinase and the other to be radiolabeled by filling in the ends with the Klenow fragment of DNA polymerase. The homopurine strand was selectively labeled at the 5′ end by standard methods (Maniatis *et al.*, 1982) using calf intestinal alkaline phosphatase followed by bacteriophage T4 polynucleotide kinase and [γ-^{32}P]ATP. The homopyrimidine strand was selectively labeled by filling in the 3′ ends with Klenow enzyme and [α-^{32}P]dNTPs by standard methods (Maniatis *et al.*, 1982). The end-labeled DNAs were each separated from unincorporated nucleotides by chromatography on Sephadex G-50 spun columns (Pharmacia) then extracted with phenol and precipitated with ethanol. Each DNA was then cut with *Apa*LI, cleaving the *hsp26* promoter at −118. The end-labeled DNA fragments encompassing the region from −52 to −118 were isolated on 8 M-urea/6% polyacrylamide gels. Note that these oligomers are predominantly homopurine or homopyrimidine, but do contain sequences that flank the homopurine/homopyrimidine stretch.

For the binding analysis, plasmid 88B13, which contains the *hsp26* promoter region (Cartwright & Elgin, 1982), as either a supercoiled plasmid or linear restriction fragment (*Bam*HI digestion), was combined with either the labeled homopurine strand-specific probe or the labeled homopyrimidine strand-specific probe in nuclease S_1 buffer at pH 5 (50 mM-sodium acetate, 200 mM-NaCl, 0·1 mM-$ZnSO_4$, 0·5% (v/v) glycerol). After incubation at room temperature for 20 min, the complexes were analyzed by electrophoresis through agarose gels in 100 mM-sodium citrate (pH 5) at 12°C, with continuous circulation for 18 to 24 h. The gels were stained with ethidium bromide for photography, and then dried for autoradiography.

(d) *DEPC treatment of naked DNA*

Procedures for treatment of naked DNA with diethylpyrocarbonate (DEPC) were based on methods described by Hanvey *et al.* (1988). A 100-μl sample of 0·4 M-Mes buffer (pH 5·5), 1 mM-EDTA or (pH 6·5), 1 mM-EDTA was shaken vigorously with 10 μl DEPC for 2 min and centrifuged for 30 s in a microfuge. An 85-μl sample of the aqueous phase was added to 3·75 μg of supercoiled plasmid 88B13 in 15 μl of TE buffer (10 mM-Tris·HCl (pH 7·4), 1 mM-EDTA) and incubated for 5 min at 23°C. The reaction was halted by the addition of 11 μl of 3 M-sodium acetate (pH 7) plus 280 μl of ethanol. The DNA was dissolved in TE buffer and precipitated twice more. The DNA was then digested to completion with *Dra*I, extracted twice with phenol/chloroform/isoamyl alcohol (49 : 49 : 2, by vol.), once with ether and precipitated with ethanol. The DNA was dissolved in 50 μl of 1 M-piperidine and heated at 50°C for 45 min. The mixture was frozen and lyophilized, then dissolved in 50 μl of water and lyophilized 3 further times. The generation of markers, and running, blotting and probing of the sequencing gels was all as described (Thomas & Elgin, 1988; Church & Gilbert, 1984). The probe used in this study was transcribed from *Eco*RI-linearized p9B1 (Thomas & Elgin, 1988) using bacteriophage T3 RNA polymerase.

(e) *DEPC treatment of isolated nuclei*

Embryos were collected as described (Elgin & Miller, 1978; Thomas & Elgin, 1988). Nuclei were isolated as described (Thomas & Elgin, 1988) except that 50 mM-Hepes buffer was substituted for 15 mM-Tris buffer throughout in order better to control the pH in the presence of DEPC. After the sucrose gradient, the nuclei were dissolved in 50 mM-Hepes (pH 7·4), 60 mM-KCl, 15 mM-NaCl, 5 mM-$MgCl_2$, 0·1 mM-EGTA, 1 mM-dithiothreitol, 0·1 mM-phenylmethyl sulfonyl fluoride, 1 M-sucrose at room temperature to an apparent A_{550} of 0·1 after dilution of 1 : 300. From 3 ml of nuclei, 250 μl were removed as a control and the rest was mixed with a further 3 ml of the same buffer, which had been vortex mixed with 600 μl of DEPC and centrifuged. Samples (500 μl) were removed at appropriate times and the DNA was purified as described (Thomas & Elgin, 1988; Wu *et al.*, 1979). A 15 μg sample of DNA was digested to completion with *Dra*I, extracted twice with phenol/chloroform/isoamyl alcohol (49 : 49 : 2, by vol.), once with ether, and precipitated twice with ethanol. The DNA was dissolved in 150 μl of 1 M-piperidine and heated at 50°C for approx. 3 h, frozen and lyophilized. The lyophilization was repeated 3 further times, dissolving the DNA in 50 μl of water each time. The generation of markers, and running, blotting and probing of the sequencing gels was identical with procedures used for naked DNA described above.

(f) S_1 *hypersensitive site mapping*

A 7 μg sample of each plasmid DNA was precipitated with ethanol and dissolved in nuclease S_1 buffer (50 mM-sodium acetate, 200 mM-NaCl, 1 mM-zinc acetate, 0·5% glycerol) at a concentration of 100 ng/μl. Nuclease S_1 in 1 μl of buffer was added to a final concentration of 6 units/μg DNA or 36 units/μg DNA, and the samples were incubated for 15 s at 25°C. The reactions were terminated by addition of 30 μl of 30 mM-Tris base and 50 μl of phenol. After further extractions with phenol the DNA was precipitated with ethanol and dissolved in restriction buffer and cleaved with *Hin*dIII and *Eco*RI. The *Hin*dIII 3′ end was selectively labeled by filling in the staggered ends using the Klenow fragment of DNA polymerase,

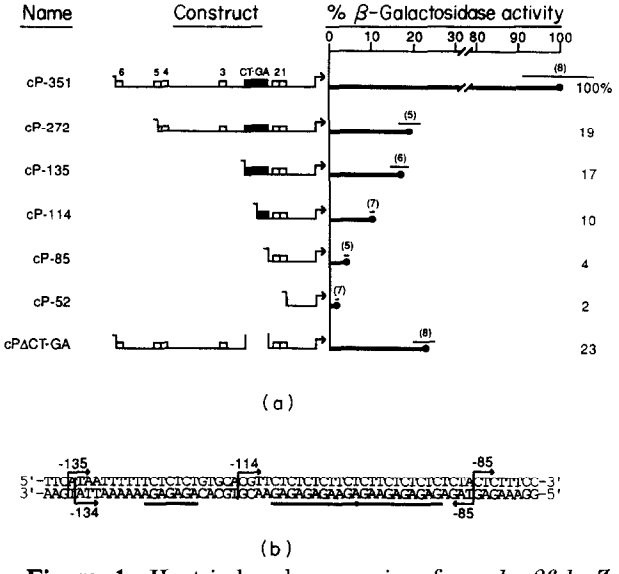

Figure 1. Heat-induced expression from *hsp26-lacZ* constructs in transgenic flies. Germline transformants containing *hsp26-lacZ* genes with 5′ or internal deletions were analyzed for levels of heat-induced β-galactosidase activity. (a) The deletion endpoints relative to the location of both the CT·GA region (filled box) and HSEs (numbered open boxes) as originally proposed by Pelham (1982). For a contemporary discussion of functional HSE structure, see Lis *et al.* (1990). The level of β-galactosidase is expressed as a percentage of the mean level of cP-351. The S.E.M. for each determination is indicated as an error bar with the number of independent transformant lines used for each determination shown in parentheses. The actual mean percentages for the constructs are listed to the right of the graph. β-Galactosidase activity from the injection stock $Adh^{fn6}cn;ry^{502}$ was not subtracted, and was at approximately 2% of wild-type, thus the level of β-galactosidase activity observed for cP-52 was not significantly above background. (b) The sequence within the CT·GA region. Endpoints of 5′ deletions and of the internal deletion are indicated above and below the sequence, respectively. The homopurine/homopyrimidine mirror repeat and GAGA sequences are underlined. The dot on the longer line designates the center of the homopurine/homopyrimidine mirror repeat.

[α-^{32}P]dGTP and non-radioactive dATP (Maniatis *et al.*, 1982). The DNA fragment containing single-stranded S_1 cleavages on the new end-labeled purine strand was gel-purified and separated on 8 M-urea/7% polyacrylamide gels that were then subjected to autoradiography. Densitometric analysis was performed on appropriate exposures of the gels using a BioRad model 620 video densitometer.

3. Results

(a) *Optimal heat-induced expression requires sequence elements in addition to HSEs*

We analyzed a series of progressive 5′ deletion mutations of an *hsp26-lacZ* fusion gene reintroduced into *Drosophila* by P element-mediated germline transformation (Fig. 1; Rubin & Spradling, 1982). Expression from the *hsp26-lacZ* gene was determined by measuring the level of β-galactosidase activity in whole-animal extracts after a two hour heat shock using chlorophenol red/β-D-galactopyranoside as substrate. This same strategy has been used to characterize the regulatory region of the *Drosophila hsp70* gene (Xiao & Lis, 1988). In addition, Simon & Lis (1987) have demonstrated that levels of heat-induced β-galactosidase activity are directly correlated with levels of steady-state RNA.

From this deletion analysis, we identified three regions of DNA that contribute to *hsp26* heat-induced transcription. Two of the intervals, −351 to −272 and −85 to −52, contain HSEs 6, and 1 and 2, respectively (Fig. 1). This analysis and the mutational analyses conducted by Cohen & Meselson (1985), Pauli *et al.* (1986) and Simon & Lis (1987) suggests a functional role for HSEs 1, 2 and 6. This is consistent with the genomic footprinting data of Thomas & Elgin (1988), who demonstrated the heat-induced formation of DNase I-protected regions over these same HSEs *in vivo*. In addition, the absence of contribution from sequences between −272 and −135 (Fig. 1) is consistent with the observation that, while this interval does contain several incomplete HSEs (designated as 3, 4 and 5 in Figs 1(a) and 5(b); Simon & Lis, 1987; Xiao & Lis, 1988), they are unoccupied after heat shock in nuclei (Thomas & Elgin, 1988).

The third region that contributes to the heat-induced expression of the *hsp26-lacZ* gene is located between −135 and −85 (Fig. 1). We will call this the CT·GA region because of the numerous (dC-dT)·(dG-dA) repeats in this interval (see Fig. 1(b)). A 5′ deletion of sequences to −135 is expressed at 17% of wild-type; the reduction likely reflecting the loss of HSE 6 (Cohen & Meselson, 1985; Pauli *et al.*, 1986; Simon & Lis, 1987; Thomas & Elgin, 1988). A further deletion to −85, removing the CT·GA region, reduces expression to 4% of wild-type, a four- to fivefold reduction relative to the −135 deletion. We also analyzed a gene containing an internal deletion of the CT·GA region (−134 to −85, Fig. 1(b)) and again observed a four- to fivefold reduction in heat-induced expression, in this case from 100% to 23%. The reduction in expression that we observed for the internal deletion was not a consequence of changing the spacing between elements still on the construct, since restoration of correct spacing with random DNA had no effect (see Fig. 5(b), below, construct cPri).

The significant contribution of the CT·GA region to heat-induced expression was unexpected. The CT·GA region contains no HSEs as determined by sequence analysis (Xiao & Lis, 1988), yet contributes as much activity to the total heat-induced expression of the *hsp26* gene as a functional HSE (Fig. 1(a), compare cP-272 to cPΔCT·GA).

(b) *Homopurine/homopyrimidine sequence in promoter forms* H-*DNA* in vitro

The same CT·GA region identified above as playing a role in heat-induced expression has been

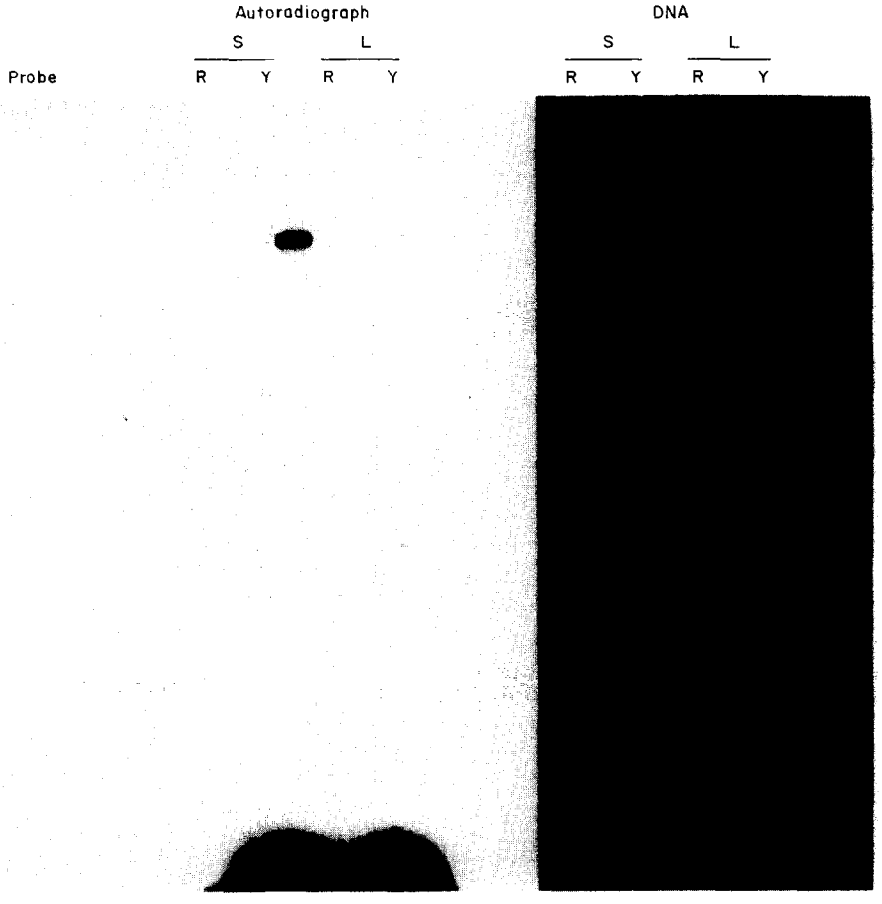

Figure 2. Complex formation of oligonucleotides with supercoil and linear DNA containing CT·GA region. Radiolabeled oligonucleotides from either the pyrimidine-rich strand (Y) or the purine-rich strand (R) of the CT·GA region were incubated under conditions of low pH with plasmid 88B13 as either supercoiled DNA (S) or as linear DNA (L). Linear DNA was generated by *Bam*HI digestion, which generates 2 DNA fragments, the larger of which contains the CT·GA region. Complexes were then run on agarose gels. The autoradiograph (left-hand panel) and ethidium bromide-staining pattern (right-hand panel) are illustrated.

identified as the region required for formation of a pH-dependent, S_1-hypersensitive structure *in vitro* (Siegfried *et al.*, 1986). If the structure identified by Siegfried *et al.* (1986) is triple-helical *H*-DNA formed by the homopurine/homopyrimidine mirror repeat that resides in this region, the purine strand of the repeat should be single-stranded and available for hybridization (Lyamichev *et al.*, 1986; Hanvey *et al.*, 1988; Johnson, 1988; Htun & Dahlberg, 1989; and see Fig. 4, below). Lyamichev *et al.* (1988) used strand-specific oligonucleotide probes to demonstrate the availability of the purine strand of homopurine/homopyrimidine mirror repeats under low pH conditions. We isolated labeled oligonucleotide probes specific for binding to either the purine or pyrimidine strand of the homopurine/homopyrimidine mirror sequence in the promoter of *hsp26*, and incubated these probes with DNA containing the CT·GA region. Only the pyrimidine probe bound to the plasmid DNA, suggesting that the purine strand of the structure was indeed single-stranded (Fig. 2). In addition, no binding to linear DNA (Fig. 2), was observed consistent with data demonstrating that *H*-DNA formation at pH 5 requires that the DNA be negatively supercoiled (Htun & Dahlberg, 1988). We have demonstrated that the pyrimidine oligonucleotide bound only at pH 5 and not pH 7·5 (data not shown), consistent with the pH dependence of S_1 hypersensitivity observed by Siegfried *et al.* (1986).

To demonstrate precisely that it is the homopurine/homopyrimidine mirror repeat within the CT·GA region that forms *H*-DNA, the structure was analyzed by diethylpyrocarbonate modification. DEPC specifically carboxyethylates N-7 residues of accessible adenine residues, which are then susceptible to cleavage by piperidine (Herr, 1985; Johnson & Rich, 1985). Under the conditions used here, i.e. bacterial levels of superhelicity and pH 5, the adenine residues specifically of the 5′ half of the mirror repeat should be single-stranded and available for carboxyethylation by DEPC (Htun & Dahlberg, 1989; Fig. 4). Supercoiled plasmids containing the CT·GA region were treated with DEPC, cleaved with restriction enzymes, heated in the presence of piperidine, and the resulting DNA fragments were analyzed by high-resolution indirect end-labeling (Fig. 3(b)). In agreement with *H*-DNA

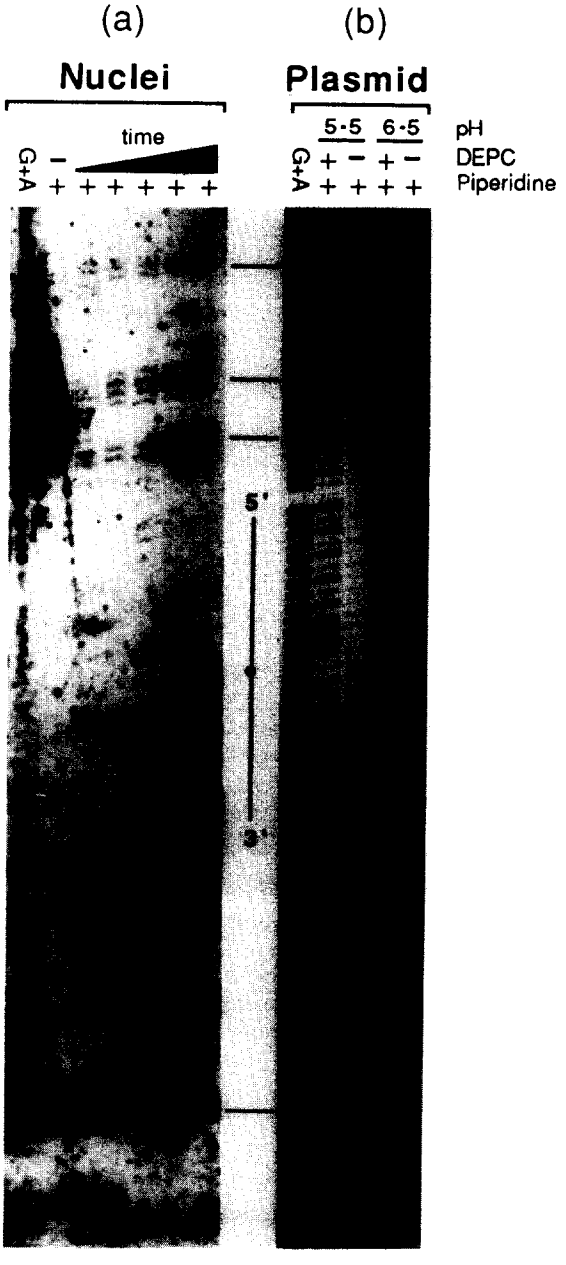

Figure 3. DEPC modification of sequences in CT·GA region *in vitro* and in nuclei. (a) Isolated nuclei were exposed to DEPC for 5, 10, 15, 20 and 25 min, after which the DNA was isolated, restriction cut, treated with piperidine and analyzed by indirect end-labeling. (b) Naked DNA was reacted with DEPC at pH 5·5 or pH 6·5 and processed by the same method as used for the genomic DNA isolated from nuclei. G+A sequencing markers were used for both experiments. The homopurine/homopyrimidine mirror repeat is indicated by the vertical line with the polarity of the purine strand indicated. Regions that contain distinct but low levels of accessible adenosine residues both *in vivo* and *in vitro* are indicated by horizontal lines.

structure, a pH-dependent accessibility of adenosine residues exclusively in the 5′ half of the mirror repeat was observed. From these high-resolution data, and from the previously discussed oligonucleotide binding results (Fig. 2), we conclude that, under

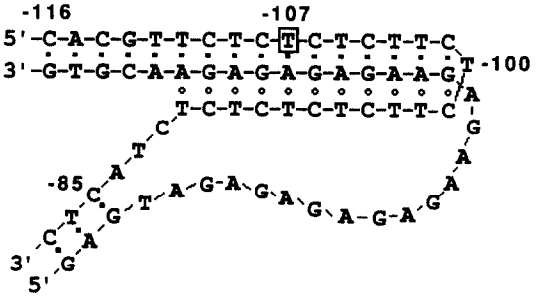

Figure 4. Proposed *H*-form structure for the homopurine/homopyrimidine mirror repeat. This is the structure predicted to form *in vitro* by the homopurine/homopyrimidine mirror repeat in the *hsp26* promoter at bacterial superhelicity and a pH of 5. Watson–Crick base-pairs are indicated by filled dots, and Hoogsteen base-pairs by open dots. The nucleotide mutated in cPT/C107 is boxed. Denaturation of base-pairs at the junction of the *H*-form and the *B*-form DNA accounts for the reactivity of adenosine residues outside the triple-helix structure (Htun & Dahlberg, 1988).

the appropriate experimental conditions, the homopurine/homopyrimidine mirror repeat located in the promoter of the *hsp26* gene does form triple-helical, *H*-DNA *in vitro*. The structure predicted from this analysis is illustrated in Figure 4.

(c) *Lack of correlation between ability to form* H-*DNA* in vitro *and transcriptional activity* in vivo

We wanted to differentiate between two of the simplest models to explain the contribution of the sequences in the CT·GA region to the heat-induced expression of the *hsp26* gene. One hypothesis, drawn from the data discussed above, is that the fivefold stimulation of expression mediated by the CT·GA region is a direct consequence of the ability of the homopurine/homopyrimidine mirror repeat within this region to form a triple helix *in vivo*, and it is the formation of this dramatic secondary structure and not the primary sequence *per se*, that is critical for function. The alternative hypothesis is that the CT·GA region contains distinct sequence elements that bind sequence-specific proteins, which then act to stimulate transcription. In this latter hypothesis, *H*-DNA does not form *in vivo*, and the ability of these sequences to form such a structure *in vitro* is without consequence *in vivo*. The analysis of two mutations, as described below, begins to address the question of which of these models is most appropriate to explain the role of the CT·GA region in *hsp26* expression.

Perfect mirror symmetry within homopurine/homopyrimidine repeats is a critical factor for the stability of the triple-helical structure (Mirkin *et al.*, 1987; Hanvey *et al.*, 1989). We introduced a T to C transition at position −107 (T/C107) in the *H*-DNA structure. This mutation disrupts the mirror symmetry, but keeps the sequence homopurine/homopyrimidine (Fig. 4). We reasoned that this change should reduce *H*-DNA stability, while being a rela-

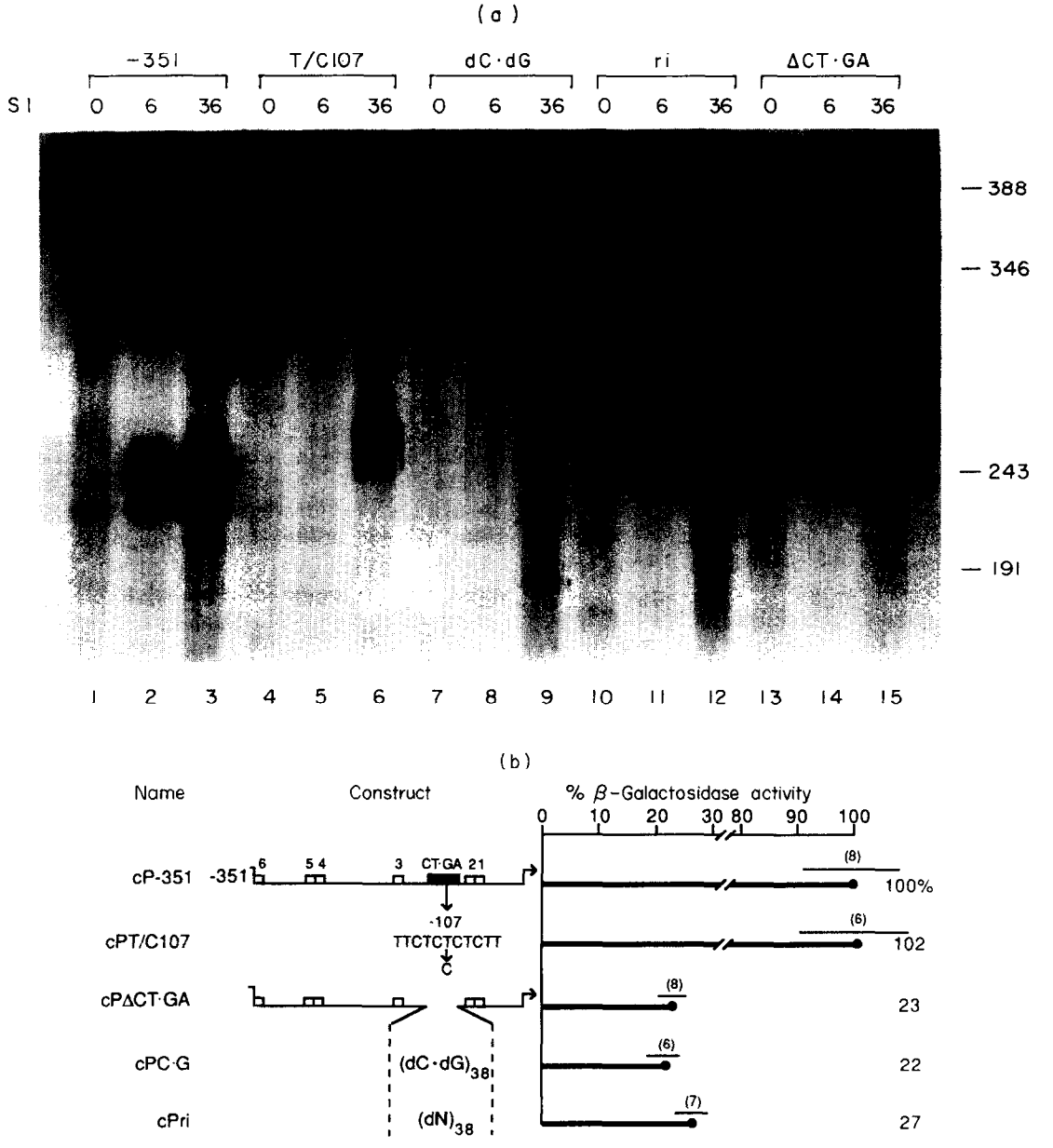

Figure 5. Comparison of *in vitro* structure and *in vivo* expression of mutations in CT·GA region. (a) Supercoil plasmid DNAs at pH 5 were treated with nuclease S_1 at 0, 6 or 36 units/μg DNA. The locations of the S_1 cleavages were mapped relative to the radiolabeled 3′ end of the purine strand. S_1-dependent fragments are highlighted with vertical black lines. Fragment sizes (in nucleotides) are listed on the right and were confirmed in repetitions of this experiment using dideoxy sequencing tracts run in parallel. Note that cPΔCT·GA contained a 42 nucleotide deletion causing the observed shift in mobility. (b) Expression of *hsp26-lacZ* genes containing the same mutations as analyzed and illustrated in (a). The analysis and presentation of the data are the same as Fig. 1(a).

tively minor sequence change with respect to a sequence-specific binding site, given the repetitive nature of the sequences within the CT·GA region. We analyzed the structure of this and other constructs discussed in this section by determining the position of nuclease S_1-hypersensitive sites on the purine strand. Analysis of wild-type DNA demonstrated the presence of eight to ten nuclease S_1-hypersensitive sites in the 5′ half of the mirror repeat (Fig. 5(a), lanes 1 to 3 and Fig. 4) consistent with an earlier analysis of this sequence by Siegfried *et al.* (1986). The T/C107 mutation resulted in both qualitative and quantitative changes to the structure (Fig. 5(a), lanes 4 to 6). The mutation resulted in an altered structure, as judged from the new location of nuclease S_1-hypersensitive sites. The nuclease S_1-accessible region is shifted in a 5′ direction by two to three nucleotides and the extent of accessible nucleotides is increased from eight to ten nucleotides in wild-type to 12 to 15 nucleotides in T/C107. While we do not know the exact nature of this new structure, it is clear from densitometry of gels like that shown in Figure 5(a) that there is, at a minimum, a twofold reduction in the frequency of

this new structure relative to wild-type. The formation of alternative but less stable structures upon introduction of asymmetric mutations in the mirror repeat has been demonstrated with other *H*-DNA-forming sequences (Hanvey *et al.*, 1989).

Analysis of the effect of the T/C107 mutation on the level of heat-induced expression in germline transformants revealed no influence of this mutation on transcription *in vivo* (Fig. 5(b)). This demonstrates that the ability of the homopurine/homopyrimidine sequence to stimulate heat-induced transcription *in vivo* is unaffected by a mutation that alters both the quantity and quality of *H*-DNA *in vitro*.

H-DNA can be formed by very different primary sequences as long as the sequence is homopurine/homopyrimidine and has mirror symmetry (Lyamichev *et al.*, 1987; Hanvey *et al.*, 1988). This allowed us to test whether a completely different sequence, one that would be unlikely to bind any of the same sequence-specific proteins but that could still form triple-helical DNA, could function to stimulate heat-induced transcription. The sequence poly[d(C·G)] was tested because this homopolymer would be unlikely inadvertently to create potential binding sites for putative proteins that might bind the wild-type (dC-dT)·(dG-dA) sequence. Sequences removed by the internal deletion of the CT·GA region were replaced with homopolymeric (dC)·(dG) using terminal transferase (Lyamichev *et al.*, 1987).

As anticipated, nuclease S_1-hypersensitive site analysis of this construct was complicated. While poly[d(C·G)] does form *H*-DNA (Lyamichev *et al.*, 1987; Kohwi & Kohwi-Shigematsu, 1988), the pattern of nuclease S_1-hypersensitive sites in poly[d(C·G)] *H*-DNA is distinct from that formed by other sequences. Rather than seeing the entire 5′ half of the guanine strand accessible to S_1, nuclease S_1-hypersensitive sites are observed at the center of the mirror repeat and at the junctions between the homopolymer and adjacent regions (Kohwi-Shigematsu & Kohwi, 1985; Hanvey *et al.*, 1988). It has been proposed that the general inaccessibility of the poly(dG) is a consequence of its close but non-specific association with the triple-helix structure and that only nucleotides at the apex of the triple helix where the purine strand is bent most severely are susceptible to nuclease S_1 (Kohwi & Kohwi-Shigematsu, 1988). It has long been observed that guanine polymers have a strong propensity to form aggregates making them inaccessible to S_1 nuclease (Vogt, 1973).

We observed two major nuclease S_1-hypersensitive sites in the poly[d(C·G)] construct (Fig. 5(a), lanes 7 to 9). The larger band corresponds to S_1 cleavage in the center of the (dG) polymer, and the smaller band corresponds to cleavage at the 5′ junction of the homopolymer and adjacent sequences, a pattern consistent with previous reports as discussed above. Diffuse bands are barely visible above background corresponding to S_1 cleavage in the 5′ half of the purine strand, while no such bands are visible corresponding to the 3′ half, suggesting that at a very low level the 5′ half of the purine strand is accessible. Quantitative calculations of poly[d(C·G)] *H*-DNA like those done for T/C107 were not possible, due to the complications discussed above. Nonetheless, a strong prediction can be made that the triple helix formed by the poly[d(C·G)] sequence would be more stable than that formed by the wild-type sequence. Both increasing length of the mirror repeat (Htun & Dahlberg, 1989) and increasing fraction of d(G·C) base-pairs (Hanvey *et al.*, 1988) contribute to the stability of the resulting triple helix. The poly[d(C·G)] construct has 38 bp of mirror symmetry with 100% d(G·C), while the wild-type sequence has 22 bp of mirror symmetry with 45% d(G·C).

We analyzed expression from *hsp26-lacZ* genes containing the poly[d(C·G)] sequence in transformant fly lines. Although the poly[d(C·G)] construct reintroduced a *H*-DNA-forming sequence into the deletion, the level of expression from cPC·G was indistinguishable from the level of expression from the original deletion, cPΔCT·GA as well as from constructs with only random DNA inserts, cPri (Fig. 5(b)). These results demonstrate that loss of expression *in vivo* due to the removal of sequences in the CT·GA region cannot be recovered by introducing alternative sequences that have the ability to form *H*-DNA *in vitro*.

(d) *Purine residues of homopurine/homopyrimidine mirror repeat in nuclei are not accessible to DEPC*

The results of both the point mutant T/C107 and the poly[d(C·G)] insertion demonstrate the lack of correlation between the ability of sequences within the CT·GA region to form *H*-DNA *in vitro* and the ability of those sequences to stimulate heat-induced expression *in vivo* (Fig. 5(a) and (b)). This result supports the conclusion that *H*-DNA formation is not involved in the regulation of transcription of the *hsp26* gene.

We were interested in knowing if formation of *H*-DNA was actually occurring *in vivo*, a question not addressed by the mutational analysis. One can imagine two possibilities; *H*-DNA is not formed *in vivo* or it is formed *in vivo* but has no obvious influence on transcription. To determine directly if sequences of the *hsp26* promoter form *H*-DNA *in vivo*, the same DEPC carboxyethylation assay used on plasmid DNA was applied to nuclei isolated from *D. melanogaster* embryos. Figure 3(a) clearly illustrates that *in vivo* there is no indication that any adenosine residues within the homopurine/homopyrimidine mirror repeat are accessible to DEPC. It is important to note that the pattern of fortuitously modified adenosine residues in the flanking sequences show a similar pattern of modification to that observed *in vitro*, indicating that the assay was working properly, and supporting the conclusion that the adenosine residues within the homopurine/homopyrimidine mirror repeat were not accessible *in vivo*. This result suggests that an *H*-DNA struc-

ture is not formed *in vivo* by sequences of the *hsp26* promoter.

4. Discussion

During analysis of sequences necessary for optimal heat-induced expression of the *Drosophila hsp26* gene, we discovered that sequence elements in addition to HSEs are responsible for fivefold stimulation of heat-induced expression. These additional sequences localized to a (dC-dT)·(dG-dA)-rich region of the promoter that had been demonstrated to form an S_1-hypersensitive structure *in vitro* (Siegfried *et al.*, 1986). Here, we demonstrate that this structure is triple-helical *H*-DNA. Could these sequences form *H*-DNA *in vivo* and mediate the stimulation of heat-induced transcription? Both mutational and structural analyses of the sequences in the (dC-dT)·(dG-dA)-rich region support the conclusion that *H*-DNA does not occur stably *in vivo*, and that the ability of these sequences to form such a stable structure *in vitro* is unrelated to their ability to stimulate heat-induced transcription *in vivo*.

Homopurine/homopyrimidine sequences with demonstrated or predicted ability to form *H*-DNA have been identified in the promoter regions of numerous genes in addition to *hsp26* (see Introduction). Except in the case of adenovirus, there has been no analysis of appropriate mutants to determine even if any of these homopurine/homopyrimidine sequences play a role in transcription, let alone to determine if these sequences mediate such a role by forming triple-helical DNA *in vivo*. In the case of the adenovirus late gene promoter, mutational analysis did implicate a homopurine/homopyrimidine sequence in the formation of a supercoil-dependent S_1 hypersensitive structure likely to be *H*-DNA (Kilpatrick *et al.*, 1986; Yu & Manley, 1986), but indicated that the ability to form this structure *in vitro* was not required for transcription *in vivo* (Yu & Manley, 1986).

We propose that some homopurine/homopyrimidine sequences that are required for transcription *in vivo* might consist of repeated arrays of binding sites for a sequence-specific transcription factor. The ability of the homopurine/homopyrimidine sequence to form *H*-DNA *in vitro* in these cases would be a consequence of the recognition sequence for the protein being itself homopurine/homopyrimidine and would not represent a functionally relevant characteristic of the sequence. We believe this may be the case for *hsp26* (see discussion below), and might be anticipated for other homopurine/homopyrimidine recognition sequences such as that for SP1 (GGGCGG; Dynan & Tjian, 1985).

Some known and potential *H*-DNA-forming sequences have been identified in eukaryotic DNA in spatial association with regions of DNA involved in functions other than transcription, such as recombination (Collier *et al.*, 1988) and replication (Lapidot *et al.*, 1989). Perhaps an *in vivo* role for *H*-DNA will be found that is related to these or some other processes. The high frequency and apparently non-random localization of homopurine/homopyrimidine sequences certainly suggest some *in vivo* function, but what those functions are and to what extent, if any, those putative functions require *H*-DNA formation *in vivo* remains an open question.

If *H*-DNA formation is not important for *hsp26* gene expression, how then do sequences in the CT·GA region mediate their effect on transcription? As mentioned above, we favor, at present, the model that a previously identified transcription factor, the GAGA factor (Biggin & Tjian, 1988), binds multiple sites in this region and is responsible for the enhanced transcription. Inspection of the CT·GA region of *hsp26* reveals three sequence elements that closely match the *Ubx* GAGA binding site consensus (C/AGAGAGAGC; Biggin & Tjian, 1988), the most-proximal two sites forming the homopurine/homopyrimidine mirror repeat (Fig. 1(b)). A highly purified preparation of protein(s) that bind the CT·GA region of *hsp26* has been obtained from *Drosophila* embryos, and a major component of this preparation is similar in size to the GAGA factor (Gilmour *et al.*, 1989). In addition, genomic footprinting done on intact nuclei (see Fig. 5A of Thomas & Elgin, 1988) reveals a region protected from DNase I beginning at the CT·GA region and extending upstream through the positioned nucleosome. Finally, consistent with a role for the GAGA-binding sites is the reduction in expression observed for cP-114 (Fig. 1(a)). This deletion does not remove sequences involved in *H*-DNA formation but does remove the GAGA element furthest upstream (Fig. 1(b)).

The substantial contribution of the CT·GA region to the heat-induced expression of *hsp26* was unexpected. There is little precedent for such a substantial contribution to heat-induced expression of non-HSE sequences outside the TATA and start-site sequences. In their mutational analysis of the yeast *YG100* gene, Slater & Craig (1987) presented data demonstrating a sixfold stimulation of heat-induced expression by a 25 bp interval that contains no HSEs, but does contain the sequence GAGAGAA. It would be interesting if yeast has a GAGA-like factor that acts in *YG100* in a manner analogous to that suggested for *hsp26*. Also, Bienz & Pelham (1986) demonstrated that optimal heat-induced expression of the *Xenopus hsp70* gene when transfected into mammalian cells requires both HSEs and a CCAAT box.

Like *hsp26*, the *hsp22* and *hsp70* genes of *Drosophila* have CT·GA sequences in their promoters (Mace *et al.*, 1983). We believe this warrants continued investigation to determine if they also require CT·GA sequence elements for optimal expression. In this regard, it is interesting to note that the same highly purified preparation that contains binding activity for the *hsp26* CT·GA region also binds a CT·GA sequence that resides between HSE1 and HSE2 of the *hsp70* gene (D. S. Gilmour, G. H. Thomas & S. C. R. Elgin, unpublished observation). Mutational analysis suggests

that this same *hsp70* GAGA sequence, located between HSE1 and HSE2, may play a role in heat-induced expression (Xiao, 1989).

We thank Janis Werner for embryo microinjection and assistance with genomic Southern blots. This work was supported by National Institutes of Health grant GM25232 to J.L., National Institutes of Health grant GM31532 and National Science Foundation grant DCB8601449 to S.E. and National Research Service Award GM07232 to R.G.; E.S. was supported by the Division of Biology and Biomedical Sciences, Washington University.

References

Amin, J., Ananthan, J. & Voellmy, R. (1988). *Mol. Cell. Biol.* **8**, 3761–3769.

Atkinson, B. G. & Walden, D. B. (1985). Editors of *Changes in Eukaryotic Gene Expression in Response to Environmental Stress*, Academic Press Inc., Orlando, FL.

Behe, M. J. (1987). *Biochemistry*, **26**, 7870–7875.

Bienz, M. & Pelham, H. R. B. (1986). *Cell*, **45**, 753–760.

Bienz, M. & Pelham, H. R. B. (1987). *Advan. Genet.* **24**, 31–73.

Biggin, M. D. & Tjian, R. (1988). *Cell*, **53**, 699–711.

Birnboim, H. C., Sederoff, R. R. & Paterson, M. C. (1979). *Eur. J. Biochem.* **98**, 301–307.

Boles, T. C. & Hogan, M. E. (1987). *Biochemistry*, **26**, 367–376.

Cartwright, I. L. & Elgin, S. C. R. (1982). *Nucl. Acids Res.* **10**, 5835–5852.

Church, G. M. & Gilbert, W. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 1991–1995.

Cohen, R. S. & Meselson, M. (1985). *Cell*, **43**, 737–746.

Collier, D. A., Griffin, J. A. & Wells, R. D. (1988). *J. Biol. Chem.* **263**, 7397–7405.

Dynan, W. S. & Tjian, R. (1985). *Nature (London)*, **316**, 774–778.

Elgin, S. C. R. & Miller, D. W. (1978). In *The Genetics and Biology of* Drosophila (Ashburner, M. & Wright, T. R. F., eds), vol. **2a**, pp. 112–121, Academic Press, Orlando, FA.

Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. & Efstratiadis, A. (1984). *Nucl. Acids Res.* **12**, 8043–8058.

Gilmour, D. S., Dietz, T. J. & Elgin, S. C. R. (1988). *Mol. Cell. Biol.* **8**, 3204–3214.

Gilmour, D. S., Thomas, G. H. & Elgin, S. C. R. (1989). *Science*, **245**, 1487–1490.

Glaser, R. L., Wolfner, M. F. & Lis, J. T. (1986). *EMBO J.* **5**, 747–754.

Hanvey, J. C., Klysik, J. & Wells, R. D. (1988). *J. Biol. Chem.* **263**, 7386–7396.

Hanvey, J. C., Shimizu, M. & Wells, R. D. (1989). *J. Biol. Chem.* **264**, 5950–5956.

Herr, W. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 8009–8013.

Htun, H. & Dahlberg, J. E. (1988). *Science*, **241**, 1791–1796.

Htun, H. & Dahlberg, J. E. (1989). *Science*, **243**, 1571–1576.

Ingolia, T. D. & Craig, E. A. (1981). *Nucl. Acids Res.* **9**, 1627–1642.

Johnson, B. (1988). *Science*, **241**, 1800–1804.

Johnson, B. & Rich, A. (1985). *Cell*, **42**, 713–724.

Kilpatrick, M. W., Torri, A., Kang, D. S., Engler, J. A. & Wells, R. D. (1986). *J. Biol. Chem.* **261**, 11350–11354.

Kohwi, Y. & Kohwi-Shigematsu, T. (1988). *Proc. Nat. Acad. Sci., U.S.A.* **85**, 3781–3785.

Kohwi-Shigematsu, T. & Kohwi, Y. (1985). *Cell*, **43**, 199–206.

Kramer, W. & Fritz, H. (1987). *Methods Enzymol.* **154**, 350–367.

Lapidot, A., Baran, N. & Manor, H. (1989). *Nucl. Acids Res.* **17**, 883–900.

Larsen, A. & Weintraub, H. (1982). *Cell*, **29**, 609–622.

Lathe, R., Kieny, M. P., Skory, S. & Lecocq, J. P. (1984). *DNA*, **3**, 173–182.

Lis, J. T., Xiao, H. & Perisic, O. (1990). In *Stress Proteins in Biology and Medicine* (Morimoto, R., Tissières, A. & Georgopoulos, C., eds), Chapt. 16, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, in the press.

Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1986). *J. Biomol. Struct. Dynam.* **3**, 667–669.

Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1987). *J. Biomol. Struct. Dynam.* **5**, 275–282.

Lyamichev, V. I., Mirkin, S. M., Frank-Kamenetskii, M. D. & Cantor, C. R. (1988). *Nucl. Acids Res.* **16**, 2165–2178.

Mace, H. A. F., Pelham, H. R. B. & Travers, A. A. (1983). *Nature (London)*, **304**, 555–557.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Manor, H., Sridhara, R. B. & Martin, R. G. (1988). *J. Mol. Evol.* **27**, 96–101.

McKeon, C., Schmidt, A. & deCrombrugghe, B. (1984). *J. Biol. Chem.* **259**, 6636–6640.

Merlino, G. T., McKeon, C., deCrombrugghe, B. & Pastan, I. (1983). *J. Biol. Chem.* **258**, 10041–10048.

Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A. & Frank-Kamenetskii, M. D. (1987). *Nature (London)*, **330**, 495–497.

Nover, L. (1987). *Enzyme Microbiol. Technol.* **9**, 130–144.

Pauli, D., Spierer, A. & Tissières, A. (1986). *EMBO J.* **5**, 755–761.

Pelham, H. R. P. (1982). *Cell*, **30**, 517–528.

Rubin, G. M. & Spradling, A. C. (1982). *Science*, **218**, 348–353.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463.

Schlesinger, M. J., Ashburner, M. & Tissières, A. (1982). Editors of *Heat Shock From Bacteria to Man*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Selleck, S. B. Elgin, S. C. R. & Cartwright, I. L. (1984). *J. Mol. Biol.* **178**, 17–33.

Siegfried, E., Thomas, G. H., Bond, U. M. & Elgin, S. C. R. (1986). *Nucl. Acids Res.* **14(23)**, 9425–9444.

Simon, J. A. & Lis, J. T. (1987). *Nucl. Acids Res.* **15**, 2971–2988.

Simon, J. A., Sutton, C. A., Lobell, R. B., Glaser, R. L. & Lis, J. T. (1985). *Cell*, **40**, 805–817.

Slater, M. R. & Craig, E. A. (1987). *Mol. Cell. Biol.* **7**, 1906–1919.

Southern, E. M. (1975). *J. Mol. Biol.* **98**, 503–517.

Thomas, G. H. & Elgin, S. C. R. (1988). *EMBO J.* **7**, 2191–2201.

Vogt, V. M. (1973). *Eur. J. Biochem.* **33**, 192–200.

Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M. & Wohlrab, F. (1988). *Fed. Amer. Soc. Exp. Biol. J.* **2**, 2939–2949.

Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R. & Elgin, S. C. R. (1979). *Cell,* **16**, 797–806.

Xiao, H. (1989). Ph.D. thesis, Cornell University, Section of Biochemistry, Molecular and Cell Biology, Ithaca, NY.

Xiao, H. & Lis, J. T. (1988). *Science,* **239**, 1139–1142.

Yu, Y. & Manley, J. L. (1986). *Cell,* **45**, 743–751.

Edited by N. Sternberg