A Detailed Analysis of the DNA Binding Properties and the Affinity Purification of the *Drosophila* Heat-Shock Transcription Factor

Thesis by

David J. Shuey

Submitted in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(submitted July 16, 2001)

Acknowledgements

I wish to thank the members of the Attardi and Parker labs, as well as all others at Caltech who helped to advance my professional development. I would particularly like to thank Carl Parker and the members of my committee, Giuseppe Attardi, Barbara Wold , and John Abelson for their seemingly bottomless well of support and patience. I truly do miss the energy and intellect at this place.

I dedicate this thesis to my parents, my brothers, and especially to Elaine, without whom this day may never have been realized.

Thanks to all - DJS

Abstract

The heat shock response presents an extremely attractive model to study the regulation of gene transcription in eucaryotes. I have focused the bulk of my research efforts on investigating the molecular determinants of protein-DNA interaction exhibited by the heat-shock transcription factor (HSTF) of *Drosophila*. The specific "contacts" made by the HSTF upon binding to the heat-shock element (HSE) were exhaustively determined using a variety of chemical and enzymatic probes for a number of HSTF-HSE complexes formed at both hsp70 and hsp83 gene promoters. During the course of these studies it is demonstrated that the HSTF appears to polymerize in a sequence-specific and template-directed manner on each of these promoters, a novelty in this class of regulators. Evidence suggesting that DNA bending may occur during this HSTF-DNA association on the hsp70 promoter is also presented. This observation represented the first report of a eucaryotic transcriptional activator exhibiting this property.

Significantly, during the course of these studies two novel technologies were advanced; namely the gel-based contact point method and sequence-specific DNA affinity purification methodology. These technical advancements and the HSTF-HSE interaction results will be discussed in the body of the thesis in light of their relevance to both heat-shock and global gene expression mechanisms.

Table of Contents

Chapter 1

Introduction

Introduction

The Heat-shock response

The Heat-shock or stress response is one of the most well studied regulatory systems in biology. Heat induced transcription at specific genetic loci was originally inferred by Ritossa in the early 1960's while analyzing chromosomal "puffing" in the polytene chromosomes of the *Drosophila melanogaster* salivary gland (1). Following heat-shock a specific set of these "puffs" were induced while subsequent experiments demonstrated that a number of pre-existing chromosomal "puffs" reduced in size (2,3). Thus it appeared that the heat-induced stress response could actually reorient the global gene expression pattern of *Drosophila.*

Curiously, quite a number of apparently disparate non-thermal challenges also seemed to elicit this response and further studies led to the conclusion that exposure of cells to any number of conditions or chemical agents could elicit a rapid induced expression of a specific set of highly expressed proteins, termed the heat-shock proteins or Hsp's (reviewed in 4,5). These inducing agents have been categorized by Cotto and Morimoto (6) into three classes; environmental, non-stressful, and pathophysiological. Examples of environmental stress response were the earliest studied and seem to produce the most robust response. Examples of environmental inducers would include heat-shock, exposure to heavy metals, and metabolic inhibitors. Non-stressful inducers seem to revolve around the cell division cycle and would include oncogenic states, growth factor stimulation and cellular differentiation.

Perhaps of most interest to the human disease condition would be the pathophysiological inducers which include viral and bacterial infection, inflammation, neuronal injury, cancer, aging, and a host of others pathological states (7).

Although the exact stress induction mechanism for each of these agents is not known, it seems that misfolded, denatured or aggregated cellular proteins are both common and prerequisite and appear to play the pivotal role in initiating the gene expression cascade of the stress response. The expression of the hsp's is thus believed to rescue the cell from the cytotoxic effects that arising from the stress induced abnormalities caused by protein translational, folding, aggregation and localization errors. It is also becoming clear that not all stresses can elicit a transcriptional response from all Hsp gene promoters and that different levels of stress often produce appropriately modulated expression responses (8). Eucaryotic cells seem to have exploited the flexibility of the heat-shock system and tailored it to generate a multiplicity of responses that can match and thus accommodate the diversity of induction agents and conditions.

Although this thesis will largely use the heat shock response as a transcriptional model system, a short introduction into the biology of the heat-shock proteins themselves is certainly warranted. The heat-shock proteins are classified by a molecular weight standard such that Hsp70 represents a heat-shock induced protein of approximately 70 Kd. The Hsp proteins have been shown to serve a critical

cytoprotective role and can also be grouped into three general categories; molecular chaperones that serve to assist in protein synthesis or folding pathways and can rescue misfolded intermediates, proteases that hydrolyze improperly folded or denatured proteins, and specific stress related proteins that can abate specific cellular challenges, including cellular translocation events (reviewed in 9).

The heat-shock response is a ubiquitous cell protection mechanism. A rudimentary version involving a stress induced sigma factor can even be found in procaryotic cells (10,11). In eucaryotes, this response has been remarkably conserved from yeast to fly to man and in each case at the transcriptional level seems to minimally consist of two basic molecular entities; the cis-acting Heat Shock Element (termed the HSE) and the trans acting Heat-Shock Transcription Factor (termed the HSTF). Thus the conservation of this inducible apparatus and the central role both the HSE and HSTF play in this regulatory molecular checkpoint make the heat-shock response an ideal system for inducible eucaryotic gene expression studies.

Additional layers of regulation embedded within the general heat-shock response are regularly emerging. The existence of multiple HSTF's that may exhibit markedly different oligomeric states, sub-cellular localization, and heat-shock activation responses (4). Different promoter HSE configurations also clearly play a key role in the recruitment and retention of active HSTF (12). In higher eucaryotes, HSTF often undergoes a cytosolic to nuclear relocalization and a concomitant inactive monomer to an active multimer transition upon exposure to stressful conditions (13). HSTF

also seems to possess an intramolecular repression domain that prevents this activation multimerization process (14,15). It also appears that differential phosphorylation events may play a critical role in the activation of certain HSTF's (16,17). Several groups have suggested that accumulation of the mammalian Hsp70 or Hsp90 proteins actually inhibits the HSTF-induced transcriptional activation by a direct feedback mechanism (8). Although only several of these modulatory phenomena will be discussed in detail below, the sophistication and variation generated by these and yet to be discovered components of the stress response is remarkable. It does appear clear, however, that the basic HSE and HSTF entities remain constants throughout and that many of the regulatory phenomena described above may serve a critical, but essentially complementary role as stress response induction "rheostats" in adjusting this system to react to a variety of stimuli.

Early molecular studies

I will briefly introduce some of the relevant early studies that resulted in the fundamental biochemical dissection of the heat shock response. This analysis will focus largely on work of the early-to-mid 1980's to coincide with the body of this thesis. Subsequent theoretical breakthroughs and technical advancements will be introduced in context within the Discussion section at the end of the thesis.

Early molecular analyses of the *Drosophila* HSE revealed that a 14 base pair (bp) element was commonly found in the "upstream" promoter regions of these heat-induced genes (18,19). The original HSE consensus sequence was defined as

CnnGAAnnTTCnnG where "n" can be any nucleotide. A number of studies also showed that this element could function in heterologous systems including stable mouse transformed cells (20) and transiently transfected monkey cells (21).

Comparisons of the growing list of these cis-acting elements revealed that the HSE's were often found to be imperfect, but almost always could be found in tandem arrangements with other HSE's or HSE "half-sites". The significance of these latter observations would await the discovery and isolation of the HSTF and would eventually force the redefinition of the HSE itself as a rather unique repeating cis-acting element consisting of repeating pentameric elements (22,23).

In the early 1980's an intense search was underway to identify the inferred trans-acting HSTF by a number of laboratories. Biochemical analysis by Parker (24) identified a partially purified protein component of heat-shocked *Drosophila* Kc cells that could both specifically bind to the upstream HSE elements on heat-inducible promoters and activate transcription of the hsp70 gene in a cell-free soluble *in vitro* transcription system. By employing a DNaseI footprinting technique it was demonstrated that the specific DNA binding activity was induced in heat-shocked cells. Additionally, this factor was shown to bind in an apparently contiguous topology with a previously described TATA-box binding factor from *Drosophila* (25) suggesting that this presumptive specific gene regulator may directly interact with the basal transcriptional machinery to promote specific gene transcription. This HSE binding protein factor was termed the HSTF in the spring of 1984. Subsequent

studies went on to show that this factor could also specifically bind to a number of other well characterized heat-inducible and HSE containing *Drosophila* hsp gene promoters (C.S. Parker unpublished observstions).

Nearly simultaneous *in vivo* chromatin exonuclease protection studies conducted by Wu revealed the apparent presence of protein bound to the 5' end of *Drosophila* hsp gene promoters (26). Significantly this pattern of nuclease resistance seemed to be heat inducible and appeared to consist of at least two protein binding domains. Subsequent higher resolution *in vitro* reconstituted chromatin nuclease protection and hypersensitivity studies employing the *Drosophila* hsp83 gene promoter region indicated that in fact two binding activities were involved (27). One was centered over the TATA-box region and was found to be constitutive, while the other was heat inducible and centered over the HSE region thus supporting the original hypothesis of Parker (24). Wu termed this binding protein the Heat-shock Activating Protein (HAP). It is important to note that subsequent publications by numerous groups have almost exclusively referred to this transactivator as the Heat-Shock Factor (HSF) perhaps in an apparent desire to shorten the original HSTF acronym. This factor will be referred to as HSTF in this thesis.

In 1984 recently developed P-element germline transformation methods in *Drosophila* facilitated an interesting study by Dudler and Travers (28) where they determined that full heat induction was observed when only 97 bp of upstream hsp 70 promoter DNA were included in the transgene vector. Interestingly, fly transformants

generated from vectors with only 68 bp of the hsp70 promoter DNA could not fully induce transcription. This suggested that a single HSE, present in the –68 bp construct, may not be sufficient for full induction but that two tandem HSE's, found in the –97 transgene, were.

Subsequent studies by Topol, et. al. confirmed these sequence requirements using a soluble HSTF dependent *in vitro* transcription approach (29). Interestingly, two previously undocumented HSE's, binding sites 3 and 4 located at approximately –175 and –245 respectively, were shown to bind the HSTF. This analysis demonstrated that deletion of these sites by removing all *Drosophila* sequences upstream from position –103 (Δ-103) resulted in an approximately 2-fold reduction in heat-shock gene transcription. Significantly, deletion to positions –73 and –60 showed almost identically low heat shock response levels (approximately 5-10% that of wild-type promoter constructs). Again it seems clear that a single HSE (Δ-73) is insufficient to facilitate a robust heat-shock transcriptional induction and these templates are thus essentially indistinguishable from a non-HSE containing construct (Δ-60). The potential cooperativity of HSTF binding to neighboring HSE sites of the *Drosophila* hsp70 promoter was measured in this study. These analyses demonstrated that tandem binding sites significantly stabilized the HSTF-HSE interaction. Immediate analogies to procaryotic models of gene induction were obvious, especially when this data was compared to the well established case of the cooperativity and resulting transcriptional cascades elicited during the lambda bacteriophage lysis-lysogeny decision (30). This positive HSTF cooperativity in HSE binding and it's importance

in heat-shock induction levels has since been confirmed by a number of groups (23, 31,32).

Thesis synopsis

Clearly a challenging and energetic field was rapidly evolving and it was in this time frame that I joined the Parker laboratory. Our immediate focus was to conduct a detailed investigation of the HSTF-HSE interaction in hopes of revealing possible clues to the molecular mechanisms of both the heat-shock specific and hopefully more generally applicable properties of regulated transcriptional induction in eucaryotes. These studies will be introduced below, presented in detail in Chapter 1-3 and finally discussed at the end of this thesis. A second and equally urgent goal of the laboratory was to develop a more facile and efficient HSTF isolation method that could enable a more rigorous biochemical description of these phenomena and possibly facilitate the eventual expression cloning of the HSTF cDNA. This effort will also be briefly introduced, completely described in Chapter 4 and reviewed in the Discussion section. A brief introduction to Chapters 1-4 follows.

In Chapter 1 the development of a novel protein-DNA contact point method utilizing a gel-separation technique is presented. Employing this technology we performed a detailed analysis of the HSTF-DNA interactions at the *Drosophila* hsp70 gene promoter. The expected dyad symmetry of the contacts was observed and an unexpected and interesting contact point incongruity was also revealed. These phenomena are discussed in detail.

In Chapter 2 we attempted to address the question of whether HSTF may alter the tertiary structure of the DNA template upon specific binding to the HSE using a gel-shift methodology and inference from other systems. Results are also discussed in detail below. Significantly, during the course of this study an apparent intermediate HSTF-hsp70 intermediate was discovered, inspiring a more detailed analysis of this HSTF-DNA interaction and another *Drosophila* HSE containing region; the hsp83 promoter.

In Chapter 3 a higher resolution gel system is employed to fractionate intermediate HSTF-HSE protein DNA complexes. The contact point studies of both the hsp 70 and hsp83 gene promoters strongly suggest that the HSTF can actually "polymerize" in a sequence directed fashion at these HSE containing gene promoters, representing a completely novel protein-DNA interaction phenomenon for a eucaryotic regulator.

In Chapter 4 the development of a sequence-specific DNA affinity purification methodology is described. This technique was employed to purify the HSTF to apparent homogeneity from both yeast and *Drosophila* nuclear extracts. Further studies demonstrated that these activators were of the same approximate molecular weights and that the *Drosophila* HSTF activated HSE dependent transcription in a reconstituted *in vitro* system.

The final discussion of this data will follow Chapters 1-4 and will be split into two

parts (parts 1 and 2). Part 1 will focus on reviewing the results and technical advancements described in Chapters1-4 in a temporally contextual manner and will include our contemporary interpretations. Part 2 will reevaluate this data, where necessary, to more precisely fit the evolving models of the molecular nature of the HSTF-HSE interaction that have emerged from studies conducted following the completion of this work.

The Appendix consists of a reprinted study (33) describing the partial purification and characterization of the Human mitochondrial RNA polymerase. This work was performed under the encouraging supervision of Dr. Guiseppe Attardi prior to the studies presented herewith. Thematic and spatial considerations will unfortunately preclude any further discussion of this work in the thesis presentation.

Chapter 2

Binding of *Drosophila* heat-shock gene transcription factor to the hsp 70 promoter; Evidence for symmetric and dynamic interactions.

David J. Shuey and Carl S. Parker

Journal of Biological Chemistry (1986) June 15; volume 26, number 17, pages 7934-7940.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 261, No. 17, Issue of June 15, pp. 7934-7940, 1986
Printed in U.S.A.

Binding of *Drosophila* Heat-shock Gene Transcription Factor to the hsp 70 Promoter

EVIDENCE FOR SYMMETRIC AND DYNAMIC INTERACTIONS*

(Received for publication, October 30, 1985)

David J. Shuey‡ and Carl S. Parker

From the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

A *Drosophila* heat-shock gene transcription factor (HSTF) has been shown to bind to three domains upstream from the TATA homology on a hsp 70 gene. The domain closest to the TATA homology consists of two contiguous binding sites with different binding affinities. Occupancy of the TATA homology proximal site (site 1) coordinates HSTF binding to the neighboring site (site 2) in a cooperative manner (Topol, J., Ruden, D. M., and Parker, C. S. (1985) *Cell* 42, 527–537). We have used alkylation interference and protection experiments to determine which residues within the binding sites are closely contacted by the HSTF. The contacts inferred from these studies included the residues present in the consensus sequence found in all HSTF binding sites and exhibit rotational symmetry, suggestive of a multimeric HSTF. By employing a gel electrophoresis separation technique we were able to resolve two protein-DNA complexes consisting of site 1 occupancy (complex A) and sites 1 and 2 occupancy (complex B). Analysis of these discrete species reveals that a subset of contacts within site 1 change upon HSTF binding to site 2, suggesting that a conformational change in the protein-DNA complex occurs. Implications for the activation of heat-shock gene transcription are discussed.

The *Drosophila* heat-shock genes are transcriptionally activated by a mechanism that utilizes, in part, a heat-shock gene-specific transcription factor (HSTF[1]) (2). It is known that the HSTF binds to several sites on the hsp 70 promoter (1) as well as to all of the other *Drosophila* heat-shock genes examined.[2] Present within all of the high affinity binding sites is a rotationally symmetric consensus sequence: C__GAA__TTC__G. Several laboratories have shown that this sequence can confer heat inducibility when placed in an analogous position upstream of heterologous promoters (3, 4). Maximal activation of the hsp 70 gene in *Drosophila*, however, appears to require more than a single binding site. Dudler and Travers (5) have shown that sequences within approximately 100 base pairs from the start point of transcription are required for transcriptional activation of the *Drosophila* hsp 70 gene *in vivo*. Within this 100-base element it has been determined that two HSTF binding sites are present, both of which must be present for maximal transcription of the hsp 70 genes *in vitro*. These two sites are occupied by a cooperative interaction where a high affinity site (site 1) is first occupied at low levels of HSTF followed by efficient HSTF binding to the critical second site (site 2). The cooperative binding to site 2 has been suggested to serve as a molecular switch that turns on the hsp 70 genes upon activation (1).

In this study we have investigated the spatial distribution of the HSTF on binding sites 1 and 2 by performing a series of chemical DNA interference and protection experiments. These experiments have identified the purine and phosphate residues present within the two binding sites that are critical for HSTF binding. We have learned that many of the purine contacts occur within the consensus sequence and are found in a symmetrical arrangement. In addition, cooperative binding to site 2 results in a series of interesting transitions in the contacts made by the HSTF within the binding domain. These contact transitions suggest that conformational changes in the protein-DNA complex arise as a result of site 2 occupancy. The potential significance of these changes in the activation of the hsp 70 genes is discussed.

MATERIALS AND METHODS

Isolation of Drosophila HSTF—Heat-shock transcription factor was partially purified from nonshocked *Drosophila* K_c cells as previously described (2). In addition, phenylmethylsulfonyl fluoride was included at 1 mM during cell homogenation and nuclear lysis. We estimate that the HSTF used in these experiments represents approximately 10% of the total protein. Several HSTF preparations were used in this study, each yielding very consistent results.

All data presented in this report were generated with HSTF purified from nonshocked *Drosophila* K_c cells. We have, however, investigated the possibility that heat-shocked K_c cells harbor a structurally distinct HSTF that may interact with the promoter differently. Methylation interference experiments were performed on each strand with HSTF isolated from heat-shocked cells. The interference properties of this factor were found to be indistinguishable from those employing nonshocked factor.[3]

Cloned DNA Template—All binding experiments described utilized the 101-base pair gel-purified *Hin*dIII-*Bam*HI insert of plasmid aDm 3110 including *Drosophila* sequences between −11 and −103. This recombinant was constructed by ligating a *Hin*dIII-*Sal*I fragment of a 5′ Bal 31 deletion construct: 5′-Δ-103 (1) into a *Hin*dIII-*Sal*I restricted pUC9 vector (6).

DNA templates were labeled at the 5′ end with [γ-^{32}P]ATP and polynucleotide kinase at the *Hin*dIII site (coding strand) or the *Bam*HI site (noncoding strand). Following digestion with the second restriction enzyme, DNA fragments were gel purified and recovered from DE-81 membranes as previously described (1). Specific activities

* This research was supported by a grant from the National Institutes of General Medical Sciences and in part by funds from American Cancer Society Grant NP-482 (to C. S. P.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by National Institutes of Health Predoctoral Traineeship Grant GM 07616.

[1] The abbreviations used are: HSTF, heat-shock gene-specific transcription factor; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

[2] C. S. Parker and J. Topol, manuscript in preparation.

[3] D. Shuey, unpublished data.

were generally 2–5 × 10^6 cpm/pmol.

Alkylation Interference Experiments—Labeled DNA fragments were methylated with dimethyl sulfate or ethylated with ethylnitrosourea essentially as described by Sakonju and Brown (7). Approximately 0.03 pmol (100,000 cpm) of the alkylated templates were individually incubated with various concentrations of purified HSTF (0–1 μg) in a 15-μl reaction consisting of 20 mM Hepes, pH 7.6, 40 mM KCl, 5 mM $MgCl_2$, 4% Ficoll, 3% glycerol, and 50 μg/ml pBR322 plasmid DNA restricted with *Hin*fI. Following incubation at 22 °C for 5 min the reaction mixtures were loaded onto a 1.5-mm thick 2.5% vertical agarose gel and electrophoresed at 200 V for 60–75 min in 0.25 × TBE (25 mM Tris-OH, 25 mM boric acid, 0.25 mM EDTA). Bands representing unbound, complex A, and complex B species were identified by autoradiography and excised from the gel. DNA was extracted from the gel slices by incubating them at 90 °C in 150 μl of TE (10 mM Tris, pH 8.1, 1 mM EDTA) for 30 min followed by extraction with 70 °C phenol. The aqueous phase was extracted again with phenol and once with ether. All volumes were then adjusted to 200 μl with TE. Base cleavage was carried out by the addition of 10 μl of 2 N NaOH and subsequent incubation at 90 °C for 30 min. Samples were chilled on ice, and 2 μg of salmon sperm DNA was added, and the DNA was precipitated with 3 volumes of 90% ethanol, 0.1 M NaOAc. The pellets were washed with 70% ethanol, dried, and resuspended in formamide-loading dyes. Cleavage products were displayed on 12% acrylamide, 6 M urea sequencing gels using G + A and C + T specific reactions as markers (8).

DNase I Footprinting—Binding conditions were identical to those described above. Following a 5-min incubation at 22 °C, DNase I (Worthington) was added to each reaction to a final concentration of 1 μg/ml. After 30 s the digestion was stopped by the addition of EDTA to 10 mM and the mixture was applied directly onto an agarose gel. Labeled DNA was eluted, denatured, and run on a sequence gel as described above, omitting the base hydrolysis step.

Methylation (Dimethyl Sulfate) Footprinting—Increasing concentrations of HSTF (0–1 μg) were incubated with unalkylated labeled templates in the binding conditions described above. After 5 min, dimethyl sulfate was added to a final concentration of 50 mM and the incubation was continued at 22 °C for 10–15 min. Reactions were terminated by loading the mixtures directly onto an agarose gel. Protein-DNA complex gel analysis, DNA recovery, cleavage, and display were performed as described above.

Computer Graphics—The HSTF-DNA contacts shown as dotted surfaces which depict the van der Waals radius of either guanine N-7 or adenine N-3 were projected in three dimensions by Stephen L. Mayo at the Caltech Materials Simulation Facility (directed by Professor William A. Goddard, funded in part by the Energy Conversion and Utilization Technology Project of the Department of Energy). An Evans and Sutherland PS-300/DEC VAX 11/780 was used to display interactive color graphics in real time, using the BIOGRAF macromolecular modeling and analysis program written by S. L. Mayo, B. D. Olafson, and W. A. Goddard. The crystallographic coordinates of B-DNA established by Arnott and Huskins (9) were adapted for visual display by S. L. Mayo.

RESULTS

Isolation of HSTF-DNA Complexes—To determine the residues critical for HSTF binding we employed the chemical modification procedure developed by Gilbert and co-workers (10–13). This procedure identifies contacts between a sequence-specific DNA-binding protein and its DNA target by partially alkylating the DNA and determining if the binding protein can specifically bind to the modified DNA fragment. Those alkylated residues that prevent binding are considered to be critical contacts. This approach requires that specifically bound DNA be separated from unbound DNA in order to determine the important contacts for protein binding. Methods previously used to separate protein-DNA complexes from unbound DNA include nitrocellulose filtration (13) and antibody precipitation (7, 14). The nitrocellulose filtration procedure was initially attempted with the HSTF but failed to resolve specific protein-DNA complexes from nonspecific protein-DNA interactions. By using a modification of the protein-DNA complex gel electrophoresis technique initially developed by Garner and Revzin (15) and Fried and Crothers (16), we have resolved HSTF-DNA complexes from unbound DNA fragments.

Shown in Fig. 1 are the results obtained by the addition of increasing amounts of partially purified HSTF to a 101-base pair end-labeled DNA fragment containing two contiguous HSTF-binding sites. All binding reactions contained a 400-fold mass excess of pBR322 plasmid DNA. As described previously (1) two slowly migrating protein-DNA complexes are observed. At low HSTF concentrations a protein-DNA complex (designated *complex A* in Fig. 1) is first observed followed by a more slowly migrating complex (designated *complex B* in Fig. 1) formed at higher HSTF concentrations. That specific protein-DNA interactions are responsible for the formation of these complexes was determined by subjecting the protein-DNA complexes to DNase I digestion prior to agarose gel electrophoresis. The DNA in the complexes was excised and eluted from an agarose gel, denatured, and applied to a standard sequencing gel (see "Experimental Procedures"). In *panel B* of Fig. 1 the DNase I cleavage pattern obtained for unbound DNA (*lane 1*), complex A (*lane 2*), and complex B (*lane 3*) is shown. The DNase I footprint clearly shows that complex A has HSTF bound only to site 1 (the TATA homology-proximal binding site) while complex B has HSTF bound to both sites 1 and 2.

This two-step gel procedure consisting of a native agarose gel of the protein-DNA complex followed by a denaturing acrylamide gel of the eluted DNA was employed in each of the chemical modification experiments described below. We have analyzed only those complexes that formed at lower HSTF concentrations. This was done because saturating levels of HSTF resulted in a low level of nonspecific binding in

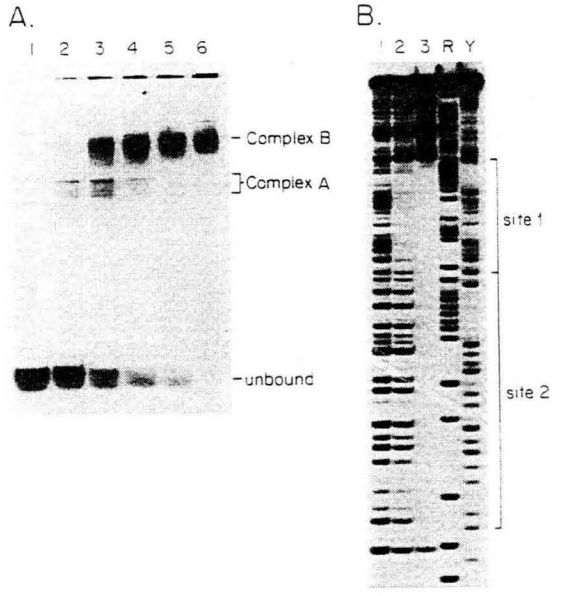


FIG. 1. **HSTF-DNA complex gel and DNase footprint analysis.** *A*, autoradiogram of a dried analytical complex gel showing complexes A and B. *Lanes 1–6* represent 0.01, 0.02, 0.05, 0.1, and 0.2 μg of partially purified HSTF. A constant 0.005 pmol of template DNA was used in each binding reaction. *B*, DNase I cleavage pattern of template DNA eluted from a preparative complex gel. *Lane 1*, minus protein; *lane 2*, complex A; *lane 3*, complex B. The fragment used consisted of the plasmid aDm 3110 insert labeled at the 5′ end of the coding strand. *Lanes R* and *Y* are chemical cleavage reactions of purines and pyrimidines, respectively.

complex B, masking the specific interactions. Presumably this is due to the presence of contaminating DNA-binding proteins in the HSTF preparations. All experiments described in this report employed the same DNA fragment (see "Experimental Procedures" for details).

Interference of HSTF Binding by Purine Methylation—Double-stranded DNA methylation with dimethyl sulfate is specific for the N-7 position of guanine (located in the major groove) and the N-3 position of adenine (located in the minor groove). Alkylation of either base creates a positive charge on the purine ring, rendering the modified residue base labile.

A partially methylated template DNA was incubated under standard binding conditions with subsaturating levels of HSTF and subsequently electrophoresed on a native agarose gel. The DNA recovered from the protein-DNA complex gel was denatured, subjected to base cleavage, and displayed on a sequencing gel shown in Fig. 2. *Panel A* shows the results obtained when the coding (top) strand was labeled and *panel B* shows the results obtained for the noncoding (bottom) strand. *Lanes 1* and *5* are control reactions where no HSTF was added, showing the dimethyl sulfate-induced cleavage pattern. The DNA eluted from complex A, complex B, and the unbound fragment are shown in *lanes 2* and *6*, *3* and *7*, *4* and *8*, respectively. *Lanes R* and *Y* represent purine and pyrimidine specific cleavage products, respectively, for each template. *Arrows* denote residues that, when methylated, interfere with specific HSTF binding. Such residues are identified by the absence of a band in the protein-DNA complex, because their methylation prevents specific HSTF binding. All guanines within the consensus sequence (with the exception of G-62 in complex A) are strong contacts. The interference properties of two residues, in particular, are altered in the transition from complex A to complex B. Base G-55 of the coding strand is only an important contact for site 1 occupancy in complex A. This contact is no longer essential for site 1 occupancy when site 2 is bound by HSTF to form complex B (compare *lanes 2* and *3* in *panel A* of Fig. 2). Conversely, base G-62 of the noncoding strand is a critical contact for complex B formation but not complex A formation. This is surprising because this residue is present within the consensus sequence of site 1. These observations suggest that a conformational change may occur to the HSTF-DNA complex upon cooperative HSTF binding to site 2 (see below for further results and discussion).

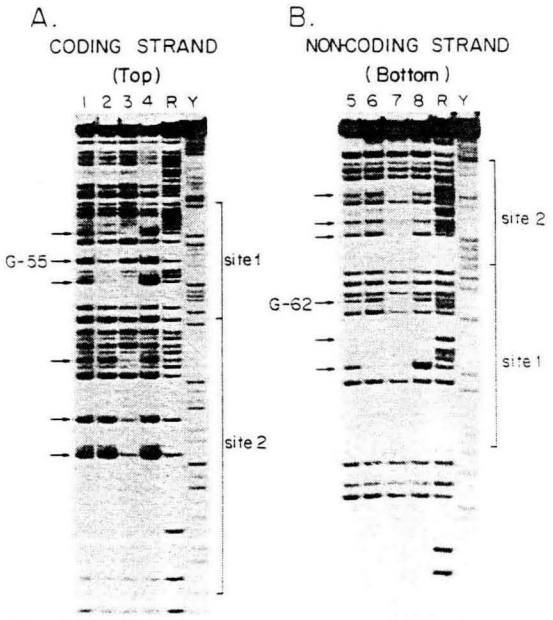


FIG. 2. **HSTF methylation interference.** Methylation-specific cleavage products of template DNA eluted from preparative complex gels. *Lanes 1* and *5*, minus protein; *lanes 2* and *6*, complex A; *lanes 3* and *7*, complex B; and *lanes 4* and *8*, unbound fraction. *Lanes R* and *Y* represent chemical cleavage of purines and pyrimidines in each case. Approximate DNase I footprint boundaries are shown in *brackets*. *Arrows* denote critical contacts. Designated bases G-55 and G-62 represent contacts within site 1 that differ in complex A and complex B lanes.

The unbound methylated DNA fragments derived from experiments employing subsaturating levels of HSTF are shown in *lanes 4* and *8*. These fragments are enriched with those residues critical for HSTF binding to site 1. We note that those residues critical for site 2 occupancy are not significantly enriched in the unbound fraction. Stable HSTF binding to site 2 is dependent upon prior HSTF binding to site 1; thus methylated residues that prevent binding to site 1 also prevent binding to site 2. Methylation of critical residues within site 2 has no effect on site 1 occupancy and will, therefore, appear in complex A. For these reasons a significant enrichment of site 1 but not site 2 contacts in the unbound fraction of DNA fragments is expected.

Interference of HSTF Binding by Phosphate Ethylation—To gain insights into how the HSTF is positioned on the phosphate backbone of the binding site we performed a series of ethylation interference experiments. The reagent used in these studies was ethylnitrosourea, which reacts primarily with phosphates forming a neutral base-labile phosphotriester. The experimental design for these experiments was essentially as described above for the methylation interference studies. The effects of prior ethylation of the DNA fragment

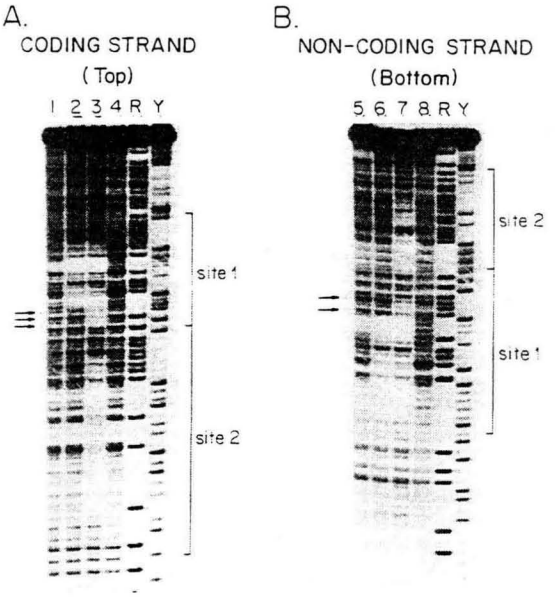


FIG. 3. **HSTF ethylation interference.** Ethylation-specific cleavage products from a preparative complex gel. Each lane is identified as described in Fig. 2. *Arrows* denote contacts within site 1 boundaries only detected in the complex B lanes.

containing sites 1 and 2 on HSTF binding are shown in Fig. 3. *Lanes 1* and *5* are the ethylated DNA fragments derived from control reactions where no HSTF was added. Complex A, B, and the unbound DNA fragments are shown in *lanes 2* and *6*, *3* and *7*, *4* and *8*, respectively. The cleared regions in complex A and complex B represent those phosphate residues that, when ethylated, prevent HSTF binding. Particularly worth noting are the strong phosphate contacts surrounding the consensus sequence central TTC elements in both sites 1 and 2 (*lanes 2* and *3* in *panel A* and *lanes 6* and *7* in *panel B* of Fig. 3). The position of these essential phosphate contacts suggests that the conserved bases may have important minor groove binding determinants. Consistent with this observation we can detect partial interference of site 1 occupancy by methylation of the A-56 residue in the minor groove (Fig. 2, *lanes 5–7*). As described above for the methylation interference experiments, the interference of certain residues within site 1 is dependent on which protein-DNA complex is analyzed. *Arrows* denote those phosphate residues within site 1 that only make contacts in complex B.

It must be stressed that this reagent is not completely phosphate specific (17) and may be reacting with ring nucleophiles, particularly on guanines (compare minus protein, *lanes 1* and *5*, to purine markers, *lane R*). Thus, all "phosphate" contacts immediately 5′ of guanine residues are considered ambiguous. Alkylation interference results are summarized in Fig. 5A.

Methylation (Dimethyl Sulfate) Footprint Experiments—A complementary set of data for the purine interference experiments was obtained by identifying residues protected from dimethyl sulfate methylation by prebound HSTF. The bases previously identified that, when methylated, interfere with HSTF binding are expected to be protected from alkylation in these studies. As before unbound DNA and complexes A and B were eluted and analyzed on a sequencing gel (shown in Fig. 4). Those residues protected from methylation by bound HSTF appear as the absence of bands in the complex A (*lanes 2* and *5*) and complex B (*lanes 3* and *6*) lanes. Some bases exhibit greatly enhanced reactivity in the presence of bound HSTF possibly due to increased ring nitrogen nucleophilicity and/or dimethyl sulfate concentration created by a hydrophobic protein domain. Such enhancements indicate a close approach by the HSTF protein to DNA.

The dimethyl sulfate footprinting pattern is consistent with the methylation interference results because all previously determined close contacts are also strongly protected from methylation by bound HSTF (compare Fig. 5, *A* and *B*). As previously observed in the interference studies, certain bases within site 1 have altered methylation properties when examined in complex A and complex B. For example, on the bottom strand G-62 is protected only in complex B and not complex A as observed in the interference experiments. In addition, G-63 on the bottom strand and G-64 on the top strand are enhanced or protected, respectively, only in complex B. Residues G-55, G-64, and G-66 of the top strand in complex B failed to show interference yet are protected from methylation by bound HSTF. This apparent discrepancy can be explained by steric exclusion of the larger dimethyl sulfate molecule not elicited with a smaller previously incorporated methyl group.

Another interesting feature is the striking 2-fold rotational symmetry within each site exhibited in the methylation foot-

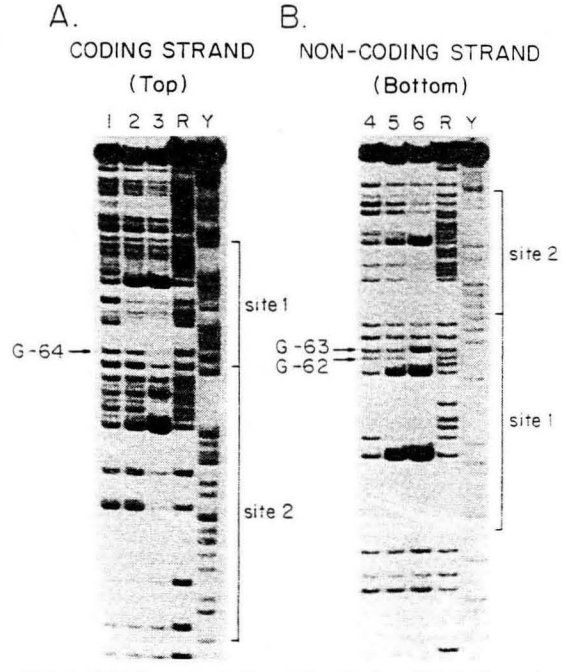


FIG. 4. **HSTF methylation (dimethyl sulfate) footprint.** Methylation-specific cleavage products of template DNA eluted from preparative complex gels. *Lanes 1* and *4*, minus protein; *lanes 2* and *5*, complex A; *lanes 3* and *6*, complex B; *lanes R* and Y, chemical cleavage of purines and pyrimidines, respectively. Approximate DNase I footprint boundaries are shown in *brackets*. Site 1 bases G-64, G-63, and G-62 display altered methylation patterns in complex A and complex B lanes.

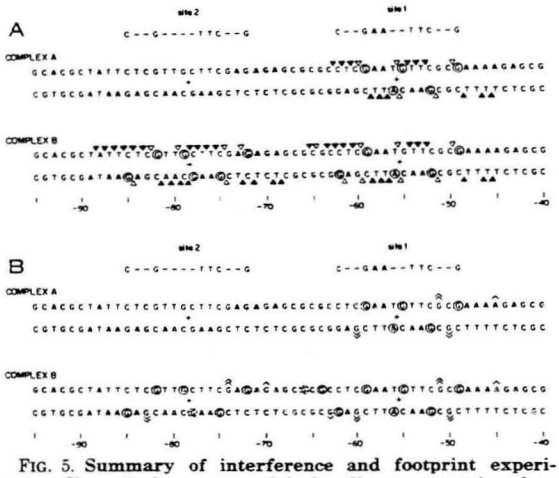


FIG. 5. **Summary of interference and footprint experiments.** Shown is the sequence of the hsp 70 promoter regions from −40 to −95 base pairs. The top strand represents the coding strand and is written in a 5′ to 3′ direction. Diagrammed above are the consensus sequence matches for sites 1 and 2. *A*, summary of alkylation interference. Data for complexes A and B. *Circled residues* represent base contacts identified by methylation interference. *Solid triangles* represent phosphate contacts, while *open triangles* are ambiguous and may represent phosphate and/or base ethylation interference events. *B*, methylation footprint summary for complexes A and B. *Circled residues* denote those bases protected from methylation by bound HSTF. *Dotted circles* represent partial protection. *Single* and *double carats* identify sites of weak and strong methylation enhancement, respectively. The "+" in each case denotes the 2-fold symmetry axis.

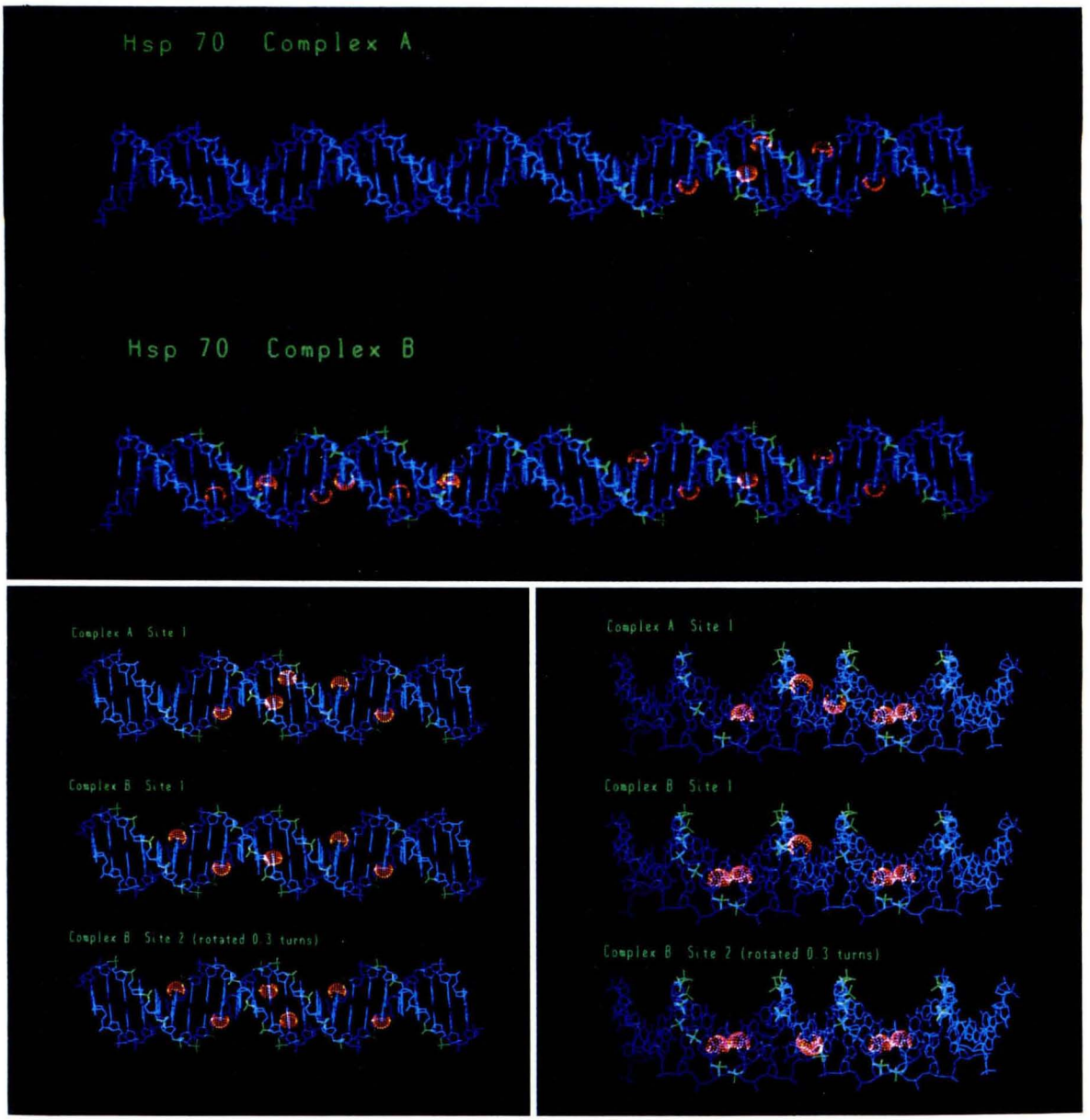


FIG. 6. **Three-dimensional computer graphics of the HSTF-binding surfaces.** The B-form of double-stranded DNA is represented by the *blue lines*. Contacts with guanosine (N-7) and adenine (N-3) are indicated by *orange stippled spheres* which outline the van der Waals radii of the ring nitrogens. Phosphate contacts are represented by *green lines* that highlight the phosphate groups. *Top*, HSTF contacts in site 1 of complex A and sites 1 and 2 in complex B. Site 1 is on the right side. *Bottom left*, the image of each individual binding site is aligned vertically. Binding site 2 of complex B has been rotated 108° to demonstrate the similarities of the contacts with binding site 1. *Bottom right*, the DNA images shown in *B* were transposed by tilting the double helix out from the plane of the page on the right side approximately 40°. This allows a direct view of the major and minor grooves.

print pattern. The interference data also display this characteristic symmetry suggesting the possibility of a multimeric HSTF (see below for discussion). In Fig. 5, *A* and *B*, the symmetry axis is noted with a "+" and aligns centrally within each consensus sequence.

DISCUSSION

Computer-aided graphics provide a powerful way to display in three dimensions the surfaces contacted by a sequence-specific DNA-binding protein. We have used these graphics

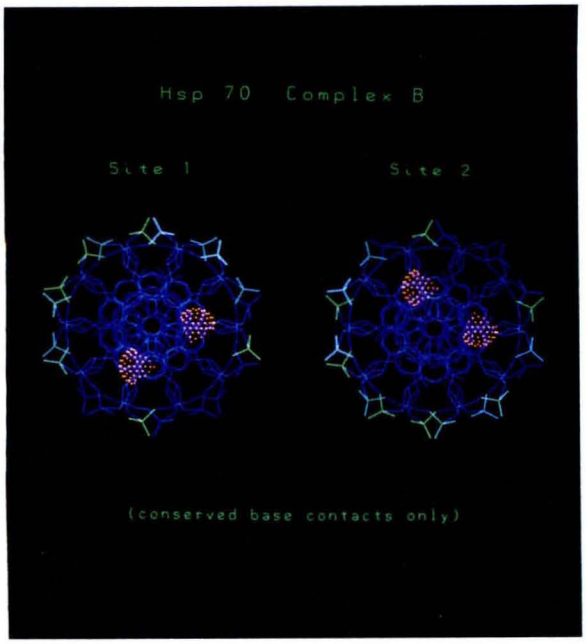


FIG. 7. **End views of binding sites 1 and 2 in complex B.** The image shown is an end view looking down the helical axes of the separate binding sites. For clarity, only the phosphate and consensus sequence guanosine contacts are shown.

to exhibit the HSTF-DNA contacts in the two binding sites of the TATA homology proximal domain of the hsp 70 promoter. As shown in Figs. 6 and 7 the sites of purine residue methylation that interfere with HSTF binding are indicated by the *orange stipples* which outline the surfaces of the ring nitrogens. The phosphate residues that, when ethylated, interfere with HSTF binding are indicated by the *green lines*. The B-form DNA double helix is represented by the *blue lines*. (For simplicity the dimethyl sulfate footprint data is omitted from the graphics display.)

As previously described a hyphenated palindromic consensus sequence has been observed in all of the HSTF high-affinity binding sites consisting of: C _ _ G A A _ _ T T C _ _ G. In complex A three of the four guanosine contacts occur in the consensus sequence: G-49 and G-59 on the top strand and G-52 on the bottom strand (see Fig. 5). As shown in Fig. 6, these three consensus guanosine contacts occur on the same surface of the DNA. A minor groove adenine contact occurs near the center of rotational symmetry on the same side of the double helix as the consensus sequence guanosine contacts. This adenosine residue is surrounded by a cluster of phosphate-backbone contacts lending further support to a direct contact with the HSTF. The fourth guanosine interaction occurs on the opposite side of the helix (G-55 on the top strand). The level of interference occurs on the opposite side of the helix (G-55 on the top strand). The level of interference observed by methylation of this base is, although extremely reproducible, never complete and only interferes with complex A formation. When both sites are occupied this residue is no longer critical for binding and may, therefore, not be directly contacted by the HSTF but destabilize the interaction by an indirect means.

In complex B, we observe interactions in the major groove of sites 1 and 2 in a similar arrangement observed for site 1 of complex A. The contacts in site 2 include all four consensus guanosines: G-72 and G-82 on the top strand, and G-75 and G-85 on the bottom strand. In addition, a consensus guanosine (G-62) of site 1 is also contacted (Figs. 5*A* and 6). The binding surface in site 2 is rotated approximately 108 degrees from that of site 1 (for B-form DNA, Fig. 7). Correcting for this rotation in the graphics display (Fig. 6, *bottom panels*), one finds remarkable similarities in the topology of sites 1 and 2 in complex B. Two other guanosine contacts occur near the center of symmetry of site 2: G-79 on the top strand and G-78 on the bottom strand. These major groove interactions are on the opposite side of the helix from the consensus contacts. Because these guanosine residues are not present in the higher affinity site (site 1), their spatial distribution relative to the conserved contacts suggests that they may interfere with site 2 occupancy by a more indirect means. It is important to note that the combination of alkylation interference and protection studies have identified only the surfaces to which the HSTF binds. The chemical nature of the contacts described in this report is not revealed by the experiments performed.

The interactions described above display marked rotational symmetry about the center of the consensus sequence at each site (Fig. 5). The distribution of these contacts is suggestive of a symmetrical binding protein, perhaps a multimeric HSTF. This speculation is supported by the well established prokaryotic examples of symmetrical contacts made by multimeric DNA-binding proteins on their target sequences (18).

The most unexpected and potentially the most biologically relevant observations were made by comparing the interference and footprinting properties of site 1 in each of the two discrete complexes. Upon cooperative binding to site 2 (complex B) several changes occur in the distribution of purine and phosphate residue contacts within site 1. A new major groove consensus guanosine contact appears within binding site 1 (G-62 on the bottom strand) while a partial interference contact (G-55 on the top strand) is lost. In addition, an essentially continuous pattern of phosphate contacts spans the two binding sites, largely on one face of the DNA helix. The dimethyl sulfate footprints support the interference studies and demonstrate that the HSTF makes a close approach to the double helix between the two binding sites in complex B (Fig. 5*B*). Collectively, these observations suggest that a conformational change in the complex occurs upon HSTF binding to site 2. One possibility is that the HSTF contacts these residues with flexible protein domains that can interact with DNA. A second, but not mutually exclusive, possibility is that the DNA bends between the two binding sites so that these residues now interact directly with surfaces on the HSTF to further stabilize the protein-DNA interactions. Recently we have obtained evidence that the HSTF induces a bend in the DNA upon binding to this domain of the hsp 70 promoter.[4] The final solution to these structural questions will only come from future x-ray crystallographic studies.

That the conformational changes described above may be important for hsp 70 gene activation is suggested by *in vivo* and *in vitro* transcription studies of deletion templates. Using P-element-mediated transformation of hsp 70 genes, Dudler and Travers (5) have shown that deletion of sequences between −68 and −97 upstream from the start point of transcription results in at least a 100-fold reduction in hsp 70 gene activation *in vivo*. Similarly, Topol *et al.* (1) have observed that sequences from −73 to −103 are required for maximal transcription *in vitro*. This study has also demonstrated that the presence and occupancy of site 1 is insufficient for optimal transcriptional activation *in vitro*. Therefore,

[4] D. J. Shuey and C. S. Parker, manuscript in preparation.

it appears that occupancy of site 2 by the HSTF results in the maximal activation of the hsp 70 promoter. Thus, a clear correlation exists between cooperative binding at site 2, apparent conformational changes, and activation of transcription. It is not clear how these protein-protein and protein-DNA interactions function mechanistically to activate transcription. One possibility is that HSTF in the "active" state (complex B) presents a complementary surface for interaction with other components of the transcriptional machinery (RNA polymerase, the TATA binding factor, etc.). These protein-protein and protein-DNA contacts may also be influenced by the conformation of the promoter itself. It is interesting to speculate that bending of the DNA by positive activators of transcription may be a general phenomenon. Support for this idea has come from Wu and Crothers (19) who have suggested that CAP binding to *lac* promoter DNA induces a bend or a kink in the DNA. Clearly, other transcriptional activators must be examined before any generalization can be made.

An examination of the consensus sequences and DNase I protection boundaries present on the other heat-shock genes reveals that there are potentially two contiguous binding sites.[2] The distance between the rotationally symmetric units on hsp 83, hsp 22, and possibly hsp 26 is, however, 20 base pairs and not 23 as for the hsp 70 gene. This would result in two adjacent HSTF molecules on the same surface of the double helix (approximately 10 Å closer to each other than on the hsp 70 promoter, assuming B-form DNA). It has not yet been determined experimentally whether occupancy of both presumptive sites is required for maximal transcription of these heat-shock genes or whether binding to these sites is cooperative. It is important that this information be obtained before a uniformly applicable model can be formulated to explain the mechanism of HSTF action in the activation of heat-shock gene transcription.

Acknowledgments—We thank the members of the Parker Lab, Eric Davidson, Norman Davidson, and Stewart Scherer, for helpful comments on the manuscript. We are indebted to Stephen Mayo and William Goddard for their valuable time and expertise in preparing the computer-aided graphics presented in this report. We thank Phoebe Ray for her significant aid in the preparation of this manuscript.

REFERENCES

1. Topol, J., Ruden, D. M., and Parker, C. S. (1985) *Cell* **42,** 527–537
2. Parker, C. S., and Topol, J. (1984) *Cell* **37,** 273–283
3. Pelham, H. R. B. (1982) *Cell* **30,** 517–528
4. Mirault, M.-E., Southgate, R., and Delwart, E. (1982) *EMBO J.* **1,** 1279–1285
5. Dudler, R., and Travers, A. A. (1984) *Cell* **38,** 391–398
6. Vieira, J., and Messing, J. (1982) *Gene* **19,** 259–268
7. Sakonju, S., and Brown, D. D. (1982) *Cell* **31,** 395–405
8. Maxam, A., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
9. Arnott, S., and Huskins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* **47,** 1504–1509
10. Gilbert, W., Maxam, A., and Mirzabekov, A. (1976) in *Control of Ribosome Synthesis* (Kjeldgaard, N. C., and Maaloe, O., eds), pp. 139–143, Munksgaard, Copenhagen
11. Ogata, R. T., and Gilbert, W. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 5851–5854
12. Johnsrud, L. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 5314–5318
13. Siebenlist, U., and Gilbert, W. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 122–126
14. Jones, K. A., and Tjian, R. (1984) *Cell* **35,** 155–162
15. Garner, M. M., and Revzin, A. (1981) *Nucleic Acids Res.* **9,** 3047–3060
16. Fried, M., and Crothers, D. M. (1981) *Nucleic Acids Res.* **9,** 6505–6525
17. Sun, L., and Singer, B. (1975) *Biochemistry* **14,** 1795–1802
18. Pabo, C. O., and Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53,** 293–321
19. Wu, H.-M., and Crothers, D. M. (1984) *Nature* **308,** 509–513

Chapter 3

Bending of promoter DNA on binding of heat shock transcription factor.

David J. Shuey and Carl S. Parker

Nature (1986) October 8; volume 323, pages 459-461.

Reprinted from Nature, Vol. 323, No. 6087, pp. 459-461, 2 October 1986

Bending of promoter DNA on binding of heat shock transcription factor

David J. Shuey & Carl S. Parker

Division of Chemistry 147-75, California Institute of Technology, Pasadena, California 91125, USA

The *Drosophila* heat-shock transcription factor (HSTF) has been shown to bind to three domains of the heat shock protein 70 gene (*hsp 70*) control region[1,2]. The most critical of these for transcriptional activation appears to be the one closest to the TATA-homology region[2-4]. This domain, spanning sequences from −40 to −95, consists of two contiguous HSTF binding sites (sites 1 and 2) that are occupied in a cooperative manner[2] (see Fig. 1). Recent alkylation interference and protection studies suggest a conformational change occurs in the protein–DNA complex at site 1 upon sequential HSTF binding at site 2 (ref. 5). We report here that HSTF binding to a single site or to both contiguous sites results in the introduction of a specific DNA bend within this domain of the *hsp 70* promoter.

It is well known that proteins can alter the conformation of DNA. Examples of stable DNA bending include nucleosomes in eukaryotes[6,7] and possibly DNA binding protein II complexes in prokaryotes[8]. More recently, DNA bending induced by site-specific DNA binding proteins has also been documented. Rosenberg *et al.* have obtained co-crystals of the restriction enzyme *Eco*RI with its recognition site that diffract to high resolution, revealing dramatic DNA kinks[9]. Prior to these studies, DNA bending had been postulated in an effort to maximize the proposed crystal complementarity of the structures of cro[10,11] and CAP[12] to their target sequences in B-form DNA. Wu and Crothers provided evidence supporting this predicted DNA bending for CAP bound at the *lac* operon using a protein-DNA complex gel assay[13]. This inferred bending was confirmed by direct electron microscopic visualization of CAP bound to the *lac* promoter element[14]. DNA bending has also been demonstrated by electron microscopy of size fractionated transcription complexes from HeLa cell extracts where a striking bend has been mapped near the site of initiation on a *Xenopus laevis* vitellogenin gene[15]. The functional significance of these DNA structural changes remains to be demonstrated. In this report we show that DNA bending occurs upon site-specific DNA binding of a transcriptional regulator in eukaryotes—HSTF of *Drosophila*.

The HSTF, purified from nuclei of heat-shocked *Drosophila* Kc cells, has been shown to be a positive activator of *hsp 70* transcription but not for a control actin gene in reconstituted *in vitro* transcription studies[1]. DNase I footprinting experiments demonstrate that the HSTF also binds to the 5′ control regions of many other *Drosophila* heat-shock genes[16]. Within each binding site, the heat-shock consensus sequence C--GAA--TTC--G can be found[2,3]. *In vivo*[3,4] and *in vitro*[2] transcription studies implicated the contiguous TATA-homology proximal to binding sites 1 and 2 as the most critical for transcriptional activation. Biochemical analysis has revealed that the HSTF binds cooperatively to these sites with a 10–20-fold higher intrinsic affinity for site 1 than for site 2 (ref. 2). Using a protein–DNA complex gel separation technique it is possible to resolve species consisting of HSTF bound to site 1 (complex A) and HSTF bound to both sites 1 and 2 (complexes B and B*)[2]. The precise HSTF stoichiometries in the three complexes are not known. We have previously postulated from a contact point symmetry argument that complex A has a HSTF dimer bound at site 1 while in complex B dimers occupy both sites[5]. Recent dimethyl sulphate interference and protection studies suggest complex B* consists of a dimer bound at site 1 and a monomer bound at the proximal half of site 2 (D.S., unpublished results). Support for this notion comes from the observation that complexes A and B* can be shifted to complex B at higher HSTF concentrations (Fig. 3*a* and unpublished results). Similar experiments demonstrate that the HSTF can also form at least three distinct complexes with the analogous promoter region of the *Drosophila* hsp 83 gene (D.S., manuscript in preparation).

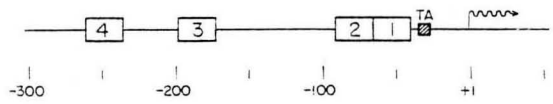


Fig. 1 The *hsp 70* promoter structure showing the positions of the four HSTF binding sites, the TATA-homology region (TA), and the start point of transcription (wavy arrow).

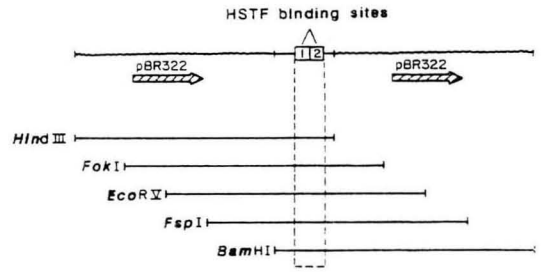


Fig. 2 Restriction fragments harbouring HSTF binding sites 1 and 2 of the *hsp 70* promoter used to investigate DNA bending. The 346-base pair (bp) direct repeats of pBR322 (wavy lines) flank a 101-bp stretch of *Drosophila* DNA containing the HSTF binding domain. Shown below are the fragments generated upon cleavage with the indicated restriction enzymes. The position of the binding sites relative to the ends of each fragment is shown with the dashed lines.

Methods. The clone aDm3111 was constructed by a two-step procedure. Plasmid pBR322 was restricted with enzymes *Bam*HI and *Hin*dIII and the 346 bp fragment was isolated by gel electrophoresis. This fragment was subsequently treated with phosphatase. A 10-fold molar excess of this DNA was ligated to 1 μg of the 101-bp gel-purified *Bam*HI/*Hin*dIII insert of plasmid aDm3110 containing HSTF binding sites 1 and 2. The ligation mixture was applied to an agarose gel and the 793-bp product was isolated. This fragment was subsequently inserted into a *Bam*HI/*Hin*dIII restricted pUC9 vector.

Recently, the protein–DNA complex gel separation technique was used to determine the critical contacts HSTF makes with its target sequence[5]. Interestingly, the contacts within site 1 differ when complexes A and B are analysed suggesting that a conformational change occurs to the HSTF-site 1 complex upon sequential HSTF binding at site 2. Thus, we wished to ask whether changes in DNA structure played a significant role in this apparent conformational change.

As discussed by Wu and Crothers[13], the mobility of identically sized DNA fragments containing a specific DNA bend is dependent upon the position of the bend relative to the ends of the fragment. In accordance with gel electrophoresis theory[17,18], when a bend is located near the end of the molecule it will migrate faster than a molecule with the bend centrally located. We constructed the clone aDm 3111, drawn in Fig. 2, by inserting a segment of the *hsp 70* promoter harbouring HSTF binding sites 1 and 2 between direct repeats of a pBR322 restriction fragment. Digestion with the indicated enzymes yields fragments of identical size and base pair composition with variant relative

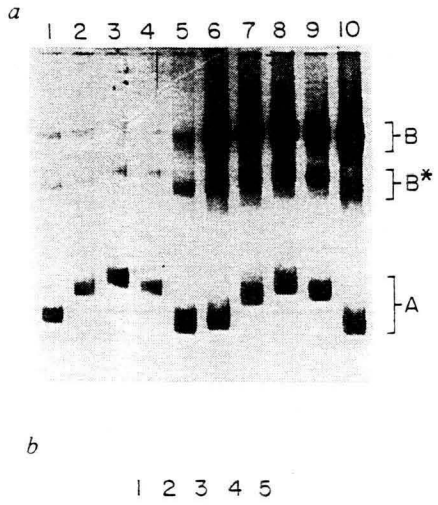


Fig. 3 *a*, HSTF-DNA complex gel analysis. Autoradiograph of a protein-DNA complex gel with the positions of complexes A, B and B* indicated at the right. Complexes formed with the labelled restriction fragments of enzymes *Hin*dIII (lanes 1 and 6), *Fok*I (lanes 2 and 7), *Eco*RV (lanes 3 and 8), *Fsp*I (lanes 4 and 9), and *Bam*HI (lanes 5 and 10) are shown. Lanes 1–5 contained 0.20 μg HSTF in the initial binding reaction while lanes 6–10 contained 0.60 μg of HSTF. *b*, Gel analysis of purified fragments in the absence of HSTF. Lanes 1–5 contain *Hin*dIII, *Fok*I, *Eco*RV, *Fsp*I, and *Bam*HI restriction fragments, respectively.

Methods. Plasmid aDM3111 was restricted with the indicated enzymes, treated with phosphatase, and labelled at the 5′ ends with [γ-^{32}P]ATP and polynucleotide kinase. The 447-bp fragments of each digest were gel-purified by elution onto DE-81 membranes. HSTF was purified from heat-shocked *Drosophila* Kc cell nuclei as previously described. Approximately 0.005–0.01 pmol of the DNA fragments (10,000 c.p.m.) were incubated with 0.2 or 0.6 μg HSTF in the standard binding conditions consisting of 20 mM HEPES (*p*H 7.6), 5 mM $MgCl_2$, 60 mM KCl, 4% Ficoll, 3% glycerol, 0.03 mM EDTA, and 50 μg ml^{-1} pBR322 plasmid DNA restricted with *Hin*fI. Following incubation at 22 °C for 5 min the samples were chilled on ice and applied directly to a vertical 2% agarose gel. Electrophoresis was carried out at 400 V for 4–6 h at 4 °C in 25 mM Tris, 25 mM boric acid and 1 mM EDTA. The gels were dried and exposed for autoradiography. In panel *b* electrophoresis was for 2 h only to ensure the unbound DNA fragments remained on the gel.

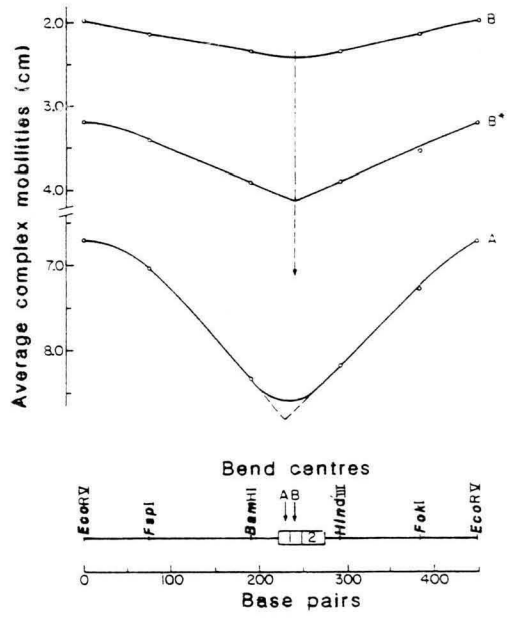


Fig. 4 Mobility plot of the HSTF-DNA complexes. Shown is a plot of the mobility of complexes A, B, and B* for each fragment used in this study against the map positions of the corresponding restriction sites. The bend centre of complex A was mapped by extrapolating the linear portions of the curve to a point on the DNA fragment. Complex B and B* bend centres were assigned as the midpoint of the identically migrating *Bam*HI and *Hin*dIII restriction fragments. The mobility data shown represent an average of five separate experiments.

positions of the binding loci. These five restriction fragments were end-labelled, gel-purified, and incubated with HSTF purified from heat-shocked *Drosophila* Kc cells. The binding conditions are essentially the same as those described previously for DNase I footprinting studies[2] with the addition of 4% Ficoll to facilitate gel loading. After a brief incubation at 22 °C the reactions are chilled and applied directly to an agarose gel at 4 °C. Following electrophoresis, the gels are dried and exposed for autoradiography. Figure 3*a* shows the results of such an experiment. Lanes 1–5 have 0.2 μg HSTF per binding reaction while lanes 6–10 contain 0.6 μg HSTF. As mentioned previously, complex A has HSTF bound to site 1 whereas complexes B and B* have HSTF bound to both sites (see below). The DNA fragments produced by *Hin*dIII (lanes 1 and 6) and *Bam*HI (lanes 5 and 10) cleavage, have HSTF binding sites close to the ends of the fragment. The DNA fragments produced by *Fok*I, *Eco*RV, and *Fsp*I harbour more centrally located HSTF binding sites. When the HSTF binds this set of fragments, complexes formed at terminal binding loci migrate faster than those with more centrally located sites indicative of a protein-induced site-specific DNA bend (Fig. 3*a*). It is apparent from these data that bending occurs in all three complexes (A, B and B*). In Fig. 3*b*, the DNA alone is electrophoresed to show that the fragments themselves show no observable bending. The unbound DNA migrates about twice as fast as complex A and is normally electrophoresed off the gel. During the course of this investigation a variety of gel systems, including acrylamide and acrylamide/agarose composite gels, were utilized. In each case the characteristic anomalous migration of the HSTF-DNA complexes was evident whereas the mobilities of the unbound DNA fragments were identical. The extreme size of these complexes, calculated to be minimum of 500 kilodaltons for complex B, dictates agarose gels to be the system of choice for this study.

A plot of the average mobility data from five experiments versus a map of the corresponding restriction sites is shown in Fig. 4. Note that each fragment can be viewed as originating from the enzymatic linearization of a circular version of the *Eco*RV fragment shown. Extrapolation of the linear portions of the complex A curve map the bend centre to site 1 as expected. Accurate mobility data for complexes B and B* data are more difficult to obtain due to their extremely slow relative migration in this gel system. It is clear, however, that the *Bam*HI and *Hin*dIII restriction fragments migrate identically in complexes B and B* in contrast to complex A (Fig. 3*a*; compare lanes 5 and 6). Even when electrophoresis is carried out for a significantly longer period of time, the mobilities of these two fragments in complexes B and B* are indistinguishable (data

not shown). Therefore, the bend centre was assigned as the midpoint of these restriction sites (*Bam*HI and *Hin*dIII), and is shifted towards site 2 relative to the complex A bend centre. The observation that the bend centres in both complexes appear to be skewed towards the direction of transcription is interesting in that a similar asymmetric bend is found when the CAP dimer is bound to the *lac* promoter[13]. It is possible that asymmetric bend centres may in fact serve a biological role. It should be noted that this study uses larger DNA restriction fragments than those used in the CAP bending study and thus we expect a correspondingly larger error in mapping the bending loci. We estimate this error to be ±5–7 base pairs.

With the data available one can only speculate on the possible significance that a bend at the promoter might have. von Hippel and co-workers[19,20] have postulated that sequence-specific DNA binding proteins scan the template for their target sequence by a sliding mechanism. If this were the case a bend may somehow signal a sliding protein (perhaps RNA polymerase II) to slow or stop at that locus, facilitating a stable productive interaction. It is also conceivable that a DNA bend may bring the bound protein molecules into proximity with other regulatory factors enabling required protein–protein interactions to occur. That such interactions are important in gene transcription has been demonstrated by Ptashne and collaborators[21,22] who have shown that maximal activation of the λ promoter P_{RM} almost certainly involves protein–protein contacts between the phage repressor bound at *OR2* and RNA polymerase. Additionally, cooperative binding of λ repressor dimers bound at operators separated by five or six turns of the helix[23] is thought to occur by a DNA bending mechanism. Transcriptional repression of the *araBAD* operon of *Escherichia coli* may actually involve interactions of repressor molecules bound to operators 150 base pairs apart[24]. Recently, Takahashi *et al.*[25] provided evidence that the spatial distribution of the simian virus 40 promoter elements is critical, suggesting that eukaryotic gene activation may depend upon the three-dimensional topology of transcription complexes. An alternative and less intriguing prospect, however, is that DNA bending could serve no regulatory function and may simply represent the most energetically favourable DNA configuration for interaction with a given protein.

We thank the members of the Parker laboratory for helpful comments on the manuscript. This research was supported by NIH grant GM29430 and American Cancer Society grant NP-482. D.J.S was supported by predoctoral traineeship grant GM07616 from the NIH. C.S.P. is a Rita Allen Foundation Scholar.

Received February 24; accepted July 25 1986.

1. Parker, C. S. & Topol, J. *Cell* **37**, 273–283 (1984).
2. Topol, J., Ruden, D. M. & Parker, C. S. *Cell* **42**, 527–537 (1985).
3. Pelham, H. R. B. *Cell* **30**, 517–528 (1982).
4. Dudler, R. & Travers, A. A. *Cell* **38**, 391–398 (1984).
5. Shuey, D. J. & Parker, C. S. *J. biol. Chem.* **261**, 7934–7940 (1986).
6. McGhee, J. D. & Felsenfeld, G. *A. Rev. Biochem.* **49**, 1115–1156 (1980).
7. Richmond, T. J., Finch, S. T., Rushton, B., Rhodes, D. & Klug, A. Nature **311**, 532–537 (1984).
8. Tanaka, I., Krzysztof, A., Dijk, J., White, S. W. & Wilson, K. S. *Nature* **310**, 376–381 (1984).
9. Frederick, C. A. *et al. Nature* **309**, 327–331 (1984).
10. Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. *Nature* **298**, 718–723 (1983).
11. Weber, I. T. & Steitz, T. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3973–3977 (1984).
12. Matthew, J. B. & Ohlendorf, D. H. *J. biol. Chem.* **260**, 5560–5562 (1985).
13. Wu, H. M. & Crothers, D. M. *Nature* **308**, 509–513 (1984).
14. Gronenborn, A. M., Nermut, M. V., Eason, P. & Clore, G. M. *J. molec. Biol.* **179**, 751–757 (1984).
15. ten Heggeler, B. & Wahli, W. *EMBO J.* **4**, 2269–2273 (1985).
16. Parker, C. S. & Topol, J. (in preparation).
17. Lerman, L. S. & Frisch, H. L. *Biopolymers* **21**, 995–997 (1982).
18. Lumpkin, O. J. & Zimm, B. H. *Biopolymers* **21**, 2315–2316 (1982).
19. von Hippel, P. H., Bear, D. G., Morgan, W. D. & McSwigger, J. A. *A. Rev. Biochem.* **53**, 389–446 (1984).
20. Park, S. C., Wu, F. Y.-H. & Wu, C.-W. *J. biol. Chem.* **257**, 6950–6956 (1982).
21. Hochschild, A., Irwin, N. & Ptashne, M. *Cell* **32**, 319–325 (1983).
22. Hawley, D. K. & McClure, W. R. *Cell* **32**, 327–333 (1983).
23. Hochschild, A. & Ptashne, M. *Cell* **44**, 681–687 (1986).
24. Dunn, T. M., Hahn, S., Ogden, S. & Schleif, R. F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5017–5020 (1984).
25. Takahashi, K. *et al. Nature* **319**, 121–126 (1986).

Chapter 4

Novel protein-DNA interactions at *Drosophila* hsp83 and hsp70 gene promoters.

David J. Shuey and Carl S. Parker

(unpublished results)

NOVEL PROTEIN-DNA INTERACTIONS AT *DROSOPHILA* HSP83 AND HSP70 GENE PROMOTERS

David J. Shuey and Carl S. Parker

Division of Chemistry and Chemical Engineering

California Institute of Technology

Pasadena, California 91125

ABSTRACT

Many of the heat-shock protein (hsp) gene promoters contain multiple heat-shock elements (HSE's) that have been shown to bind the positively activating heat-shock transcription factor (HSTF). These HSE's often overlap and share consensus sequence residues with neighboring sites. A dramatic example of this arrangement is found at the *Drosophila* hsp 83 gene promoter. In this report we have investigated the spatial distribution of HSTF bound at this promoter, and extended our previous binding analyses at the *Drosophila* hsp70 promoter. In each case, native agarose gel electrophoresis was used to resolve three distinct HSTF-DNA complexes of increasing apparent molecular weights. Analysis of these complexes with enzymatic and chemical probes revealed that for the hsp83 promoter the central HSE (site 1) is occupied at low concentrations of HSTF, followed by the sequential binding to the

flanking HSE's (sites 2 and 3) and higher protein concentrations. Curiously, for the hsp70 promoter, occupancy seems to again nucleate with a bound dimer, but then appears to polymerize distally with incremental HSTF binding to again generate three discrete complexes of increasing apparent molecular weights. We postulate that a dimeric HSTF binds first at a high affinity site (site 1) with subsequent monomeric HSTF binding at site 2 and site 3 to form a stable complex consisting of four HSTF monomers bound to HSE elements of the hsp83 and hsp70 promoters. Stoiocheometry and spatial considerations suggest that HSTF may be capable of sequence-specific template-directed polymerization. Possible implications for the role of multiple HSE's in the activation of heat-shock genes are discussed.

INTRODUCTION

The heat-shock response of eucaryotes is highly conserved at both the cellular and molecular levels (1-3). A critical component of this response is the trans-acting heat-shock transcription factor (HSTF).This protein has been has been purified to apparent homogeneity from both yeast and Drosophila (4) and seems to have nearly identical properties. Recent work suggests the HSTF from human cells is also very similar (D. Shuey, K. Plaxco, C. Parker, unpublished observations). In addition to the HSTF conservation the cis-acting elements also share many common features. All heat-inducible promoters studied contain at least one copy of the heat-shock element

(C _ _ G A A _ _ T T C _ _ G) termed the HSE. Most of these promoters harbor multiple HSE's (2,3) which are often clustered or even overlapping. In vivo (5) and in vitro (6,7) studies demonstrate that HSE's require the presence of a second cis-acting element for full heat-inducible transcription activation. This second element can be another HSE (5,7) or in some cases a distinct transcription factor binding site (8,9). HSE's also seem to function in an orientation and a distance independent manner (1,3). HSE's thus have the properties of a bi-partiate heat-inducible enhancer.

The hsp 70 and hsp 83 genes code for two of the major heat shock proteins of *Drosophila* (1). Both are highly induced upon heat-shock, although the regulation of the hsp 83 gene appears to be somewhat more complex than some of the other heat-shock genes as it responds to developmental and hormonal cues (10). It is also easily induced by stress and possesses an interesting promoter arrangement of three overlapping HSE's (11). In an attempt to gain insights into the specific protein-DNA and protein-protein interactions involved in heat-shock gene activation we have employed a high resolution gel-based separation technique to investigate the physical distribution of the HSTF bound at the hsp 83 promoter. We have also utilized this method to extend our previous analysis (12) of interactive HSTF binding at the *Drosophila* hsp70 promoter to include a recently discovered intermediate binding complex (13).

METHODS AND MATERIALS

Drosophila HSTF

Heat-shock transcription factor was purified from *Drosophila* Kc cells as previously described (7). From the observed binding stoichiometry we estimate that HSTF represents approximately 5% of the total protein in the preparation used in this study.

Cloned DNA Templates

The Hsp83 binding experiments described employed the 177 bp EcoRI-HindIII restriction fragment of plasmid aDm 3112. This plasmid was constructed by inserting the 162 bp EcoRI-MspI restriction fragment of subclone aDm 4.46 (10) into an EcoRI-AccI restricted pUC9 vector. aDm 3112 contains sequences from -8 to -170 of the Drosophila hsp 83 promoter.

The Hsp70 HSTF binding experiments were conducted with the previously described template (12) that harbors the TATA proximal HSE sites 1 and 2.

HSTF-DNA Complex Analysis

DNA templates were labeled at the 5' end with [γ-^{32}P]ATP and polynucleotide kinase at the EcoRI site (top strand) or the HindIII site (bottom strand). Following digestion with the second restriction enzyme, DNA fragments were gel-purified and recovered

from DE-81 membranes as previously described (14). Specific activities were 1-2 x 10^6 cpm/pmol. Procedures employing native protein-DNA complex gels to determine DNAse I protection and dimethylsulfate interference and protection patterns have been described in detail elsewhere (12). In the present study we utilized 2% agarose gels and performed the electrophoresis at 4°C.

The methidium isopropyl EDTA (15) (MPE) footprinting was initiated with a 5 min HSTF-DNA binding reaction at 22°C. Following the 5 min binding reaction, MPE was [prepared as recommended (16)] added to these reactions to give a final concentration of 7.5μM. After a second 5 min incubation the reagent was activated (Fe^{+++} to Fe^{++} reduction) by and addition of DTT to 10 mM. This incubation continued for 10 min at 22oC and was then quenched by the addition of EDTA to 10 mM. The reactions were chilled on ice and immediately applied to a cold agarose gel. Subsequent complex separation and analysis was performed as for the DNase I footprinting studies (12).

Computer Graphics

The coordinates for the various sequences in the B form of DNA were generated using the program genNUC (N. Pattabiraman, UCSF) and displayed using the program MIDAS (17) on an Evans and Sutherland PS2 calligraphics color display system driven by a VAX 11/750 with an UNIX (Bell Laboratories) operating system. Van der Waals molecular surfaces were generated following the algorithm reported (18).

RESULTS

An examination of the protein-DNA complexes formed at the hsp83 promoter are presented in figure 1A. In this case a chilled, porous agarose gel system was employed to resolve specific complexes harboring apparently integral increasing units of bound HSTF. To accomplish this separation, the unbound template DNA was electrophoresed off the gel. Increasing amounts of purified HSTF form three distinct complexes of increasing molecular weights, termed complexes A,B, and C. To identify the specific HSE occupacies in these three complexes these complexes were subsequently analyzed following preparative isolation of enzymatically and chemically treated samples as previously described (12).

Figure 2 depicts DNaseI and MPE footprinting of the three protein-DNA complexes formed on the hsp83 promoter. Following reaction with either DNase I or MPE these complexes were eluted from the gel and the cleavage products were displayed on the sequencing gels as shown. It is clear that HSTF binds first centrally (complex A) followed by the sequntial binding to flanking sequences. Complex B seems to consist of this initial complex A binding unit with an additional HSTF binding unit bound to the distal adjacent site. Complex C appears to fully occupy the overlapping HSE region and extend the complex B footprints by incrementally binding to the TATA

proximal site. Interestingly, the sequential transition between complexes A, B, and C seem to not only extend the footprinting profiles as described, but also seem to stablize them as the borders of both the MPE and DnaseI footprints become much more definitive. See figures 4A and 4B for a summary of these footprint boundaries.

Further characterization of this apparent binding pattern was obtained with methylation interference and protection experiments (figure 3) and this contact point data was found to be consistent with the footprinting data described above. Initially, in complex A, the strong contacts are centered around the central HSE, especially guanosine residues G-72, G-69 and G-65. With increasing concentrations of HSTF producing complexes B and then C the critical contact set first is observed distally and the proximally. See for example residues G-82 and G-52. Clearly the former is critical for both complex B and C formation, while the latter seems to only affect the larger molecular weight complex C. Also as observed above, these contacts also seem to "firm up" upon complete HSE occupancy (see G-84 and G-85 of complex B vs. complex C). These data are also summarized in figure 4A and 4B.

The Hsp83 data presented above along with previously described intermediate complexes observed at the Drosphila Hsp70 gene promoter (12) prompted our revaluation of our previous analysis of this interaction. When high-resolution gel shift experimentnts were performed, it was clear that the originally described complex B could in fact be resolved into two distinct protein-DNA complexes (B and B*). DMS

methylation footprint analysis of these complexes suggest that complex A consists of HSTF binding to the TATA proximal HSE of the Hsp70 gene promoter (see G-49, G-52, and G-59) as previously postulated. Interestingly, resolved complexes B and B* exhibit marked differences in the methylation patterns seen at the distal end of the second HSE (see G-82 and G-85; arrows in figure 5B) suggesting that the B* complex represents a discrete intermediate between the more prevalent HSTF-HSE complexes seen in A and B.

DISCUSSION

We have used protein-DNA complex gels to investigate the spatial distribution of the Drosophila HSTF bound at the hsp 83 gene promoter. Curiously, a dimeric HSTF appears to bind to a central HSE followed by the ordered binding of monomers to the two flanking "half-sites". We have also analyzed the analogous complexes formed by HSTF on the hsp 70 promoter (12). As reported, complex A appears to consist of a HSTF dimer bound to site 1 and complex B consists of dimers at both sites 1 and 2. Subsequent studies revealed the presence of a less prevalent intermediate complex termed B* (13). Recent evidence suggests that this complex consists of a HSTF dimer bound at site 1 and a monomer bound to the proximal half of site 2 (data not shown). (In the original work complexes B and B* were not resolved due to the use of a slightly different gel system.) Additionally it has been shown that HSTF binding to

hsp 70 promoter is cooperative (7) and although not rigorously tested for hsp 83 gene promoter the protein/DNA complex transitions of Fig. 1 are indicative of cooperativity. Collectively, these observations imply some type of sequence-specific template-directed protein polymerization nucleated at the HSE. Examples of non-specific template-directed protein polymerization include . . .

Although there appear to be free HSTF monomers in solution, we have never observed a single monomer binding specifically and perhaps the binding energy of this interaction is too low to detect in these experiments. This data may force us to re-evaluate the definition of the heat-shock element. The minimal HSE seems to be composed of 2 symmetrical "half-elements" which we propose and each occupied by a HSTF monomer. There are many examples, however, of overlapping HSE's that could accommodate multiple HSTF monomers (1,2). The evidence presented above suggests that such interactions do indeed occur at the hsp 83 and hsp 70 gene promoters.

Interestingly, the HSTF has been shown to bend the hsp 70 promoter upon binding (13). As the proposed monomeric repeat unit of the HSE's within the hsp 83 promoter is 10 bp, DNA bending at this site would be additive, possibly generating some type of quaternary superstrucutre that could be recognizeable to the transcriptional machinery. Evidence suggests that heat-inducible transcription requires the presence of multiple HSE's (5,7). It has also been reported that a single HSE can interact with another distant unique factor binding site to augment the heat-

inducible transcription of a human hsp 70 gene (8,9).

Therefore it seems conceivable or even likely that protein-DNA superstructures, possibly involving the bending, wrapping, or looping of the DNA may form at these promoters facilitating the binding and initiation by RNA polymerase II.

Finally, the protein-DNA complex gel separation technique presented offers several advantages over existing methods. One can generate high quality reproducible contact point and footprinting data from relatively crude protein samples. Also, an antibody to the protein of interest is not required. Perhaps most importantly, it is possible to resolve discrete complexes formed at neighboring and interactive sites and analyze these protein-DNA interactions in detail.

The Computer Graphics Laboratory (Dr. R. Langridge, Director) is supported by NIH Grant RR 1081.

We thank Dr. N. Pattabiraman and Dr. Robert Langridge for their time and expertise in generating the color illustrations presented. We are indebted to Dr. John Lis for supplying us with the hsp 83 subclone aDm 4.46. We thank Dr. Keith Harshman for technical advice on MPE footprinting. MPE was was a generous gift of Dr. Peter Dervan. We also thank members of the Parker lab for their critical readings and

Phoebe Ray for her patient preparation of the manuscript.

REFERENCES

1. Munro S, Pelham H. What turns on heat shock genes? Nature. 1985 Oct 10-16;317(6037):477-8. PMID: 2995833

2. Craig EA. The heat shock response. CRC Crit Rev Biochem. 1985;18(3):239-80. PMID: 2412760

3. Bienz M, Pelham HR. Mechanisms of heat-shock gene activation in higher eukaryotes. Adv Genet. 1987;24:31-72. Review PMID: 3324701

4. Wiederrecht G, Shuey DJ, Kibbe WA, Parker CS. The Saccharomyces and Drosophila heat shock transcription factors are identical in size and DNA binding properties. Cell. 1987 Feb 13;48(3):507-15. PMID: 3100052

5. Dudler R, Travers AA. Upstream elements necessary for optimal function of the hsp 70 promoter in transformed flies. Cell. 1984 Sep;38(2):391-8. PMID: 6432342

6. Parker CS, Topol J. A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. Cell. 1984 May;37(1):273-83. PMID: 6722872

7. Topol J, Ruden DM, Parker CS. Sequences required for in vitro transcriptional activation of a Drosophila hsp70 gene. Cell. 1985 Sep;42(2):527-37. PMID: 4028160

8. Morgan WD, Williams GT, Morimoto RI, Greene J, Kingston RE, Tjian R. Two transcriptional activators, CCAAT-box-binding transcription factor and heat shock transcription factor, interact with a human hsp70 gene promoter. Mol Cell Biol. 1987 Mar;7(3):1129-38. PMID: 3561411

9. Wu BJ, Williams GT, Morimoto RI. Detection of three protein binding sites in the serum-regulated promoter of the human gene encoding the 70-kDa heat shock protein. Proc Natl Acad Sci U S A. 1987 Apr;84(8):2203-7. PMID: 3031671

10. Hackett RW, Lis JT. Localization of the hsp83 transcript within a 3292 nucleotide sequence from the 63B heat shock locus of D. melanogaster. Nucleic Acids Res. 1983 Oct 25;11(20):7011-30. PMID: 6314271

11. Gilmour DS, Lis JT. In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. Mol Cell Biol. 1985 Aug;5(8):2009-18. PMID: 3018544

12. Shuey DJ, Parker CS. Binding of Drosophila heat-shock gene transcription factor to the hsp 70 promoter. Evidence for symmetric and dynamic interactions. J Biol Chem. 1986 Jun 15;261(17):7934-40. PMID: 3711119

13. Shuey DJ, Parker CS. Bending of promoter DNA on binding of heat shock transcription factor. Nature. 1986 Oct 2-8;323(6087):459-61. PMID: 3020432

14. Parker CS, Topol J. A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell. 1984 Feb;36(2):357-69. PMID: 6537904

15. Van Dyke MW, Dervan PB. Methidiumpropyl-EDTA.Fe(II) and DNase I footprinting report different small molecule binding site sizes on DNA. Nucleic Acids Res. 1983 Aug 25;11(16):5555-67. PMID: 6225070

16. Hertzberg RP, Dervan PB. Cleavage of DNA with methidiumpropyl-EDTA-iron(II): reaction conditions and product analyses. Biochemistry. 1984 Aug 14;23(17):3934-45. PMID: 6435669

17. Jarvis, L., Huang, C., Ferrin, T. & Langridge, R. (1984) MIDAS-Molecular Interactive Display and Simulation Program, Computer Graphics Laboratory, University of

California, San Francisco, revised 1985, 1986.

18. Bash, P. A., Pattabiraman, N., Huang, C., Ferrin, T. E. & Langridge, R (1983) Science 222, 1325-1327

FIGURE LEGENDS

FIG. 1

(A) HSTF-DNA complex analysis of the hsp 83 gene promoter. Shown is an autoradiogram of a dried analytical gel showing the three distinct protein-DNA complexes A, B and C. The unbound DNA was electrophoresed off the gel. Approximately 0.01 pmol of template DNA was incubated with increasing concentrations of HSTF (0, 0.025, 0.05, 0.1, 0.2, 0.4 mg) in lanes 1-6, respectively. **(B) Diagram of the hsp 83 gene promoter** showing the relative positions of the overlapping HSEs (1-3), the TATA box (TA), and the start point of transcription (wavy line).

FIG. 2

DNase I and MPE footprinting of complexes A, B and C. Following reaction with DNase I or MPE the protein-DNA complexes were eluted from the gel and the cleavage products were displayed on the sequencing gels shown. Reaction products

from unbound DNA (-) and complexes A, B and C are shown for each strand. R and Y represent purine and pyrimidine sequence ladders, respectively. To the left of each figure the position relative to the start point of transcription is shown; to the right are the positions of the overlapping HSEs.

FIG. 3

Dimethyl sulfate interference (DMS INT) and protection (DMS FP) analysis of complexes A, B and C. Experiments were performed as described in the text and previously (12). The figure is labeled as in Fig. 2.

FIG. 4

(A) Summary of the hsp83 DNaseI and dimethyl sulfate (DMS) protection data for complexes A, B and C. An unbroken line spans regions of the promoter completely protected from DNase I cleavage. Broken lines span partially protected sequences. Solid circles denote bases that are completely (or nearly so) protected from methylation by bound HSTF. Dashed circles represent partial protection. Single and double carats denote bases whose reactivity toward DMS is slightly or greatly enhanced, respectively, by bound HSTF.

(B) Summary of the hsp83 MPE footprinting and DMS interference data for complexes A, B and C. MPE footprinting data is represented as described above. Circled bases represent residues that when methylated interfere with HSTF binding either weakly (broken circles) or strongly (solid circles). The consensus sequence

matches for the overlapping HSEs are shown above each figure.

FIG 5

(A) Hsp70 protein-DNA complex gel. See figure 1 legend for detailed description of HSTF titrations.

(B) Hsp70 Methylation footprint of promoter proximal HSE's 1 and 2. Arrows denote critical contacts that differentiate between complexes B* and B.

FIG 6

(A) Three-dimensional graphic representation of the HSTF binding surfaces on the hsp 83 promoter. Illustrated are the strong contacts (red spheres) and the weak contacts (yellow spheres) revealed by the dimethyl sulfate protection analysis of complexes A (top), B (middle) and C (bottom).

(B) A proposed model of the HSTF-DNA interactions at the hsp 83 gene promoter. It should be noted that this figure is purely diagrammatic and not intended to imply any knowledge of size, shape, or symmetry of the HSTF.

FIG 7

Cartoon of HSTF-HSE interactions at the hsp70 and hsp83 gene promoters.

Chapter 5

The *Saccharomyces* and *Drosophila* heat shock transcription factors are identical in size and DNA binding properties.

Greg Wiederrecht, David J. Shuey, Warren A. Kibbe and Carl S. Parker

Cell (1987) February 13; volume 48, number 3, pages 507-515.

Cell, Vol. 48, 507–515, February 13, 1987, Copyright © 1987 by Cell Press

The Saccharomyces and Drosophila Heat Shock Transcription Factors Are Identical in Size and DNA Binding Properties

Greg Wiederrecht, David J. Shuey, Warren A. Kibbe, and Carl S. Parker
Division of Chemistry
California Institute of Technology
Pasadena, California 91125

Summary

The heat shock transcription factor (HSTF) has been purified to apparent homogeneity from S. cerevisiae and D. melanogaster by sequence-specific DNA-affinity chromatography. A synthetic oligonucleotide containing an *hsp83*-like heat shock element (HSE) was prepared and ligated into concatamers and covalently coupled to Sepharose. This DNA-affinity resin allowed the rapid isolation of a yeast and a Drosophila protein with the same apparent molecular weight (70 kd). The yeast HSTF will bind to both its own and the Drosophila HSEs. Similarly, the Drosophila HSTF will bind to both its own and the yeast HSEs. The yeast and Drosophila HSTFs were subjected to preparative SDS gel electrophoresis, and the 70 kd polypeptides were eluted, renatured, and observed to generate the identical footprint pattern as the native HSTFs. Affinity-purified Drosophila HSTF was further shown to stimulate specific HSE-dependent transcription from a Drosophila *hsp70* gene in vitro.

Introduction

Transcription of eukaryotic protein-coding genes requires the presence of one or more distinct promoter elements. RNA polymerase II promoters often have a TATA homology segment located 25–30 bp upstream of the transcription start point (Goldberg, 1979). A transcription factor has been identified from both Drosophila and man that specifically binds to these sequences (Parker and Topol, 1984a; Sawadogo and Roeder, 1985). Other identified promoter-proximal elements include the CCAAT sequence and the GC motif (GGGCGG), for which *trans*-acting factors have been identified that specifically bind to each (Jones et al., 1985; Dynan and Tjian, 1984). These sequences are usually located at various positions upstream of the TATA homology or an AT-rich element. Elements of this kind are found associated with a large number of promoters and may represent general promoter elements.

In addition, a few gene-specific transcription factors have been identified by in vitro studies. These include the heat shock transcription factor (HSTF; Parker and Topol, 1984b) and the adenovirus major late gene transcription factor (USF, Sawadogo and Roeder, 1985; and MLTF, Carthew et al., 1985). The HSTF recognizes a specific element, the heat shock element (HSE, C __ __ GAA __ __ TTC __ __ G) in the promoters of a family of heat shock protein genes, whereas the USF or MLTF recognizes a specific sequence found on a few apparently unrelated genes. In vivo studies have identified what seem to be other activators and their target sequences. These include the *GCN4* and *GAL4* gene products in yeast whose target elements are the TGACT motif and UAS_G, respectively (Guarente, 1984; Bram and Kornberg, 1985; Giniger et al., 1985; Keegan et al., 1986; Hope and Struhl, 1986). Additionally, in mammalian cells the glucocorticoid and metal response elements (GRE and MRE, respectively) may be gene-specific targets for possible transcriptional activators (Miesfeld et al., 1986; Richards et al., 1984).

The heat shock response is a particularly useful system for studying the coordinate activation of a specific set of genes. All of the identified major heat shock genes of Drosophila contain multiple HSEs at various locations on their promoters (Pelham, 1985; Parker and Topol, unpublished observations). A comparison of heat shock gene promoters in species ranging from yeast to man also reveals the presence of very similar HSEs in analogous positions. This remarkable similarity of eukaryotic HSTF binding sites, as well as the highly conserved group of proteins induced, is probably a reflection of the importance of the response to the organism.

The HSTF was originally discovered in Drosophila (Parker and Topol, 1984b), but has subsequently been identified in yeast (this paper) and in HeLa cells (D. Shuey, G. Wiederrecht, and C. Parker, unpublished data). An activity similar to the HSTF has also been observed by in vivo chromatin studies (Wu, 1984). The Drosophila HSTF binds cooperatively to the two contiguous HSEs (sites 1 and 2) nearest the TATA homology of a Drosophila *hsp70* gene (Topol et al., 1985). Both contiguous elements were shown to be required for *hsp70* transcriptional activation in vivo (Dudler and Travers, 1985) and in vitro (Topol et al., 1985). Detailed analysis of the HSTF interactions with sites 1 and 2 revealed that a potential conformational change occurs upon cooperative HSTF binding to these sites (Shuey and Parker, 1986a). Further studies also revealed that the HSTF induces DNA bending upon binding to the *hsp70* promoter (Shuey and Parker, 1986b). These conformational changes may be an important component of a molecular switch that turns on the hsp genes.

All of the previous studies have been performed with partially purified Drosophila HSTF. These studies were valuable for demonstrating sequence-specific DNA binding and in vitro transcriptional activation. To understand the precise molecular mechanisms involved in heat shock gene activation and transcription, it is essential that the HSTF be purified to homogeneity. To accomplish this we have utilized the sequence-specific DNA binding properties of the HSTF and performed affinity chromatography using synthetic oligonucleotides of that sequence. We were able to rapidly and efficiently purify a protein from S. cerevisiae and D. melanogaster that bound specifically and tightly to the affinity resin. Interestingly, the Drosophila and yeast proteins are identical in size (70 kd) and possess indistinguishable footprint boundaries. Furthermore, the affinity-purified Drosophila HSTF was shown to acti-

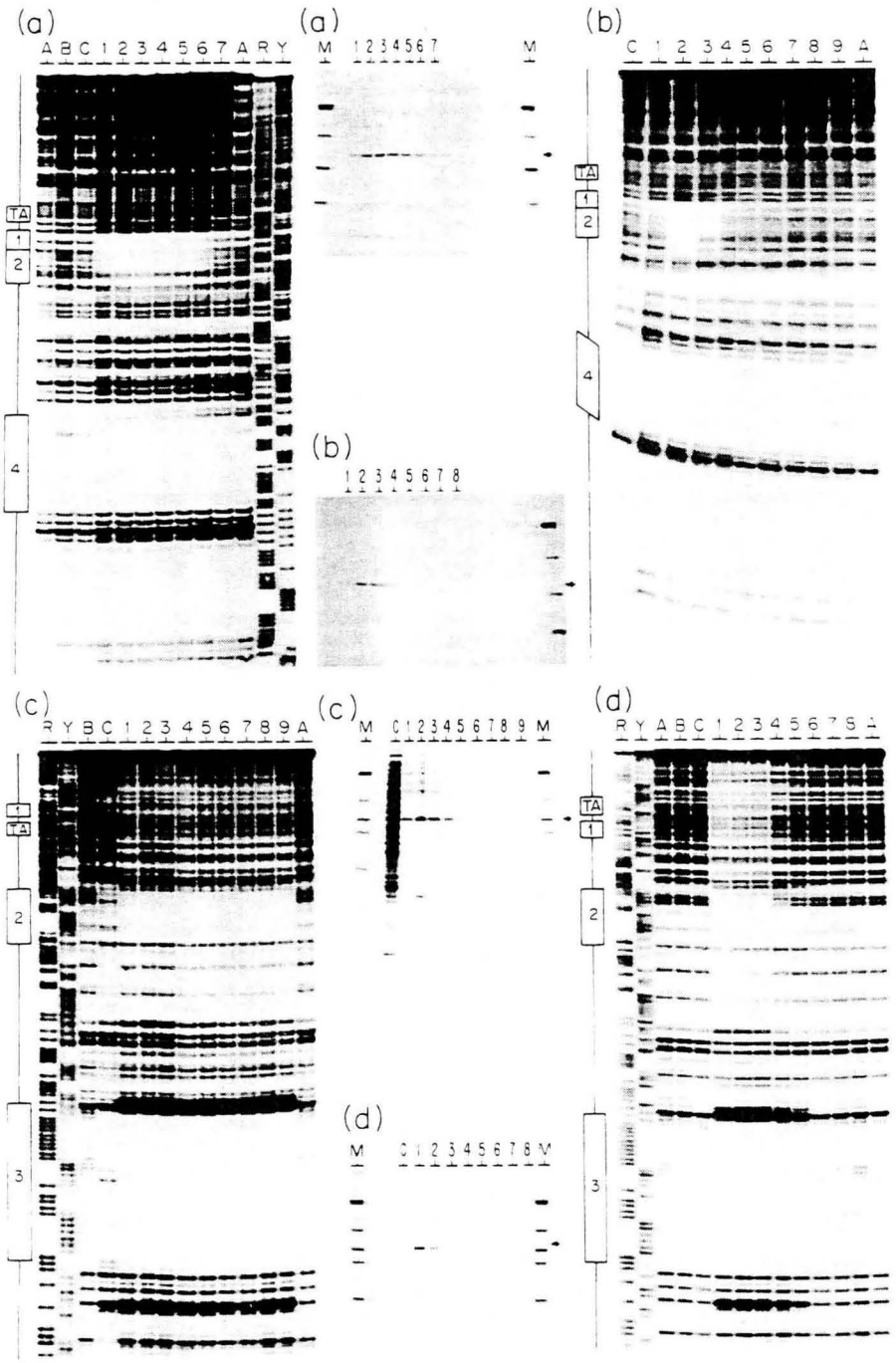


Figure 1. Chromatography of HSTF from Drosophila and Saccharomyces on the Sequence-Specific Oligonucleotide Column

Partially purified Drosophila and Saccharomyces HSTF were cycled twice on the DNA affinity column. Each panel shows the footprint analysis of the fractions eluting from the column and the accompanying protein gel. Lane A: minus protein; lane B: flow-through; lane C: 0.35 M KCl step (0.375

Table 1. Purification Table of the HSTF

Fraction	Volume (ml)	Protein Concentration (mg/ml)	Total Protein (mg)	HSTF Binding Units[a]	Specific Activity U/μg	Yield	Overall Purification[b]
Drosophila HSTF Purification							
Nuclear Extract	35	17	595	ND[c]	ND[c]	ND[c]	-
Phosphocellulose	35	14	490	3,500	0.007	-	-
Heparin Agarose	20	1	20	20,000	1	100%	30
Oligonucleotide	2.5	800 ng/ml	2.0 μg	16,550	8,225	83%	250,000
Saccharomyces HSTF Purification							
Nuclear Extract	90	20	1,800	ND[c]	ND[c]	ND[c]	-
Heparin Agarose	50	3.6	180	29,000	.105	-	-
Affigel Blue	214	0.2	42.8	81,000	1.89	100%	42
Oligonucleotide	50	300 ng/ml	15 μg	58,000	3,866	72%	86,000

[a] HSTF Binding Unit = 1 fmole binding site protected.
[b] Taking the yield into account.
[c] Not determined, footprint analysis in nuclear extracts was not quantitative.

vate *hsp70* transcription in vitro in an HSE-dependent fashion. The striking similarity in molecular weight and DNA binding properties of the HSTF from both sources is very intriguing, yet consistent with the known highly conserved nature of the heat shock response.

Results

HSTF Isolation and Purification by Sequence-Specific DNA Affinity Chromatography

The Drosophila HSTF activity was monitored by DNAase I footprinting on a cloned Drosophila *hsp70* promoter fragment modified to contain binding sites 1, 2, and 4 (Topol et al., 1985). Binding site 4, located between nucleotides −228 and −255, has the highest affinity for the HSTF, thus in footprinting assays this site is occupied first (at lower HSTF concentrations) followed sequentially by sites 1 and 2. A heat-shocked nuclear extract prepared from Drosophila K_c cells (similar results have been obtained for Schneider line 2 cells) was applied to a phosphocellulose (P11) column equilibrated with the standard chromatography buffer (HGE) with 0.1 M KCl. The HSTF flows through the P11 column under these conditions and was directly applied to a heparin-Sepharose column. The column was washed with HGE and 0.1 M KCl followed by step elution of the HSTF with HGE and 0.4 M KCl. The eluted HSTF was dialyzed to 0.1 M KCl in HGE and subjected to sequence-specific DNA-affinity chromatography as described below. At this point in the purification the HSTF had been purified 30-fold (see Table 1 for details).

The activity assay used to follow the elution of the yeast HSTF was also DNAase I footprinting. The promoter used was derived from a yeast *hsp70* gene (YG100; Ingolia et al., 1982) and contains three separate HSEs. The location of these binding sites is shown in Figure 1. Site 1, which has the poorest match to the consensus sequence, also possesses the lowest affinity for the HSTF. (This site is compressed at the top of the DNAase I ladder and is difficult to visualize.) Sites 2 and 3 have approximately equal affinities for the HSTF as expected because both have a good match to the consensus sequence. A yeast extract (prepared as described in Experimental Procedures) was applied to a heparin-Sepharose column equilibrated with 0.1 M KCl in HGE. The column was washed successively with 0.1 M KCl and 0.32 M KCl; the HSTF was eluted with 0.8 M KCl. The HSTF was dialyzed to 0.1 M KCl and applied to an Affigel Blue column. After washing the column with 0.05 and 0.5 M KCl, the yeast HSTF was eluted with 1.5 M KCl. Table 1 shows that the yeast factor was estimated to be purified 42-fold after chromatography on these two resins.

A DNA affinity column was constructed by ligating synthetic oligonucleotides containing the heat shock consensus element. The HSE sequence chosen was a modified form of the *hsp83* element (see Experimental Procedures for details). The *hsp83* HSE contains three overlapping heat shock consensus sequences within a 35 bp region. The HSTF will bind to these elements in a highly cooperative way, resulting in a very stable protein–DNA complex (D. Shuey and C. Parker, unpublished

M step when yeast HSTF was on column); numbered lanes: 1.5 M KCl step fractions; lanes R and Y: chemical cleavage reactions of purines and pyrimidines, respectively. Numbered boxes show locations of HSTF binding sites and arrows show the location of the 70 kd protein. (a) Drosophila HSTF cycled once on the column. Footprint reactions (volume, 25 μl) contained 20 μl of the indicated column fraction and 4.8 ng of the *hsp70* fragment designated d21+ described in Experimental Procedures. Masses of protein gel markers are: 165 kd, 155 kd, 90 kd, 60 kd, and 40 kd.
(b) Drosophila HSTF cycled twice on the column. Footprint reactions and protein gel markers are as in (a).
(c) Saccharomyces HSTF cycled once on the column. Footprint reactions (volume, 25 μl) contained 20 μl of the indicated column fraction and 1 ng of the pYHS-1 HindIII–EcoRI fragment labeled at the HindIII site. Protein gel markers are as in (a) except that a 66 kd marker has been added.
(d) Saccharomyces HSTF cycled twice on the column. Footprint reactions and protein gel markers are as in (c).

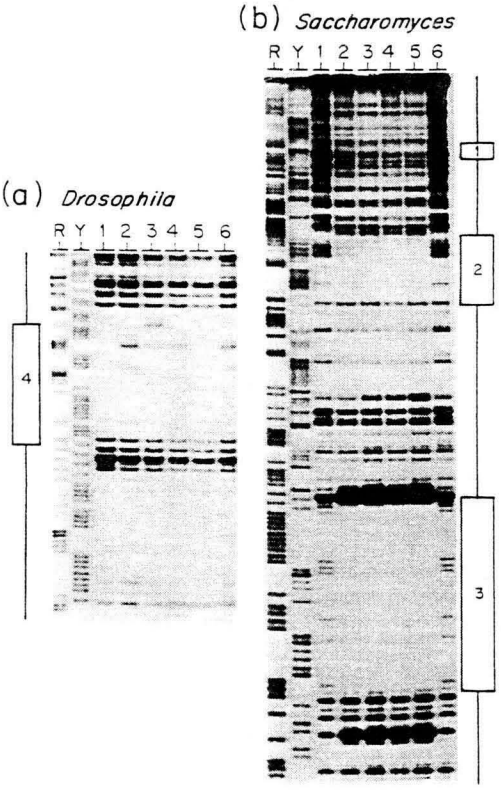


Figure 2. Renaturation of HSTF Binding Activity

(a) Footprint of renatured Drosophila HSTF. Footprint reactions (volume, 50 μl) contained 17 ng of HSTF and 4.8 ng of the d21+ fragment labeled at the EcoRI site. Lane 1: native Drosophila HSTF cycled twice on the DNA affinity column; lanes 2, 4, and 6: no protein; lane 3: renatured 25 kd polypeptide; lane 5: renatured 70 kd polypeptide.

(b) Footprint of renatured Saccharomyces HSTF. Footprint reactions (volume, 25 μl) contained 1 ng of the pYHS HindIII–EcoRI fragment labeled at the HindIII site. Lanes 1 and 6: no protein; lanes 2–5: 8, 16, 32, and 64 ng of renatured Saccharomyces 70 kd HSTF, respectively. Numbered boxes show locations of binding sites.

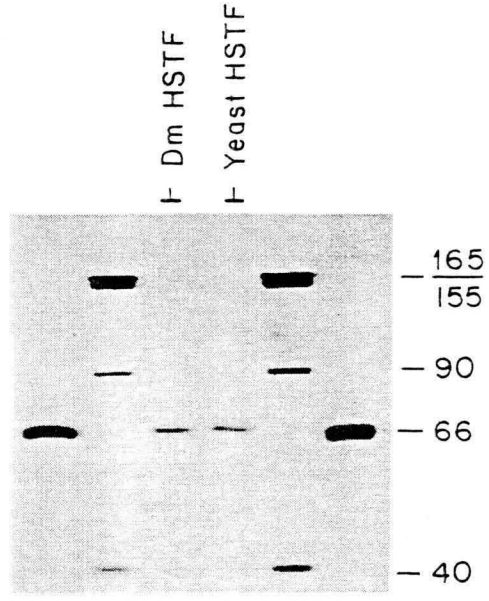


Figure 3. Comparison of Purified Drosophilia and Saccharomyces HSTF

Thirty nanograms of affinity-purified HSTF from both yeast and Drosophila was subjected to electrophoresis on a 7.5% SDS-polyacrylamide gel.

observations). It was determined experimentally that the unligated synthetic oligonucleotide served as a poor HSTF binding site. Ligating this oligonucleotide into concatamers, however, allowed high affinity binding by the HSTF. Thus we coupled ligated concatamers of synthetic *hsp83*-like HSEs to Sepharose following the procedure of Arndt-Jovin et al. (1975).

Drosophila or yeast HSTF, partially purified as described above, was applied to the affinity column and washed extensively with HGE containing 0.1 M KCl. This was followed by a second extensive wash with 0.35 M KCl for the Drosophila HSTF or 0.375 M KCl for the yeast HSTF. Both HSTFs were eluted from the column with 1.5 M KCl. A portion of each fraction was assayed by DNAase I footprinting and the protein composition was determined by SDS-polyacrylamide gel electrophoresis. The results of these analyses are shown in Figure 1a for the Drosophila HSTF and in Figure 1c for the yeast HSTF. Binding activity to both *hsp70* promoters co-chromatographs with an abundant 70 kd protein from both organisms (indicated by the arrows in Figure 1). A minor 25 kd protein is observed on both protein gels, as well as some minor higher molecular weight polypeptides. The purification resulting from chromatography on the affinity column is remarkable; the Drosophila HSTF is purified 8300-fold and the Saccharomyces HSTF is purified 2000-fold in a single chromatographic step. The overall purification of the Drosophila HSTF is 250,000-fold and 86,000-fold for the yeast HSTF (Table 1).

To remove minor impurities, the final step in the purification was to rechromatograph the pooled active fractions on the affinity resin. Footprinting assays and SDS protein gel analysis of the second chromatographic cycle are shown in Figure 1b for Drosophila and in Figure 1d for yeast. In the case of Drosophila a major polypeptide of 70 kd and a minor 25 kd polypeptide (barely visible in the experiment shown) are the only proteins observed. Greater than 90% of the protein in the yeast HSTF preparation is present in the 70 kd polypeptide. The yeast HSTF still contains two very minor contaminants of 130 kd and 100 kd and in a few preparations the 25 kd protein is also observed. (Estimates of HSTF mass after affinity chromatography were made by comparison to known amounts of standard proteins on silver-stained gels.)

SDS Gel Electrophoresis and Renaturation of the HSTF

To rigorously demonstrate that the identified 70 kd proteins were responsible for the specific DNA binding activity observed, we further purified the Drosophila and yeast HSTFs by preparative SDS-polyacrylamide gel electrophoresis. The proteins were visualized by KCl staining. The only two bands observed in the Drosophila preparations were 70 kd and 25 kd. Each of these proteins was excised from the gel and electroeluted. Only the 70 kd protein was clearly identifiable in the yeast HSTF preparations, and it was also cut from the gel and electroeluted. To renature these proteins we followed, fundamentally, the procedure of Hager and Burgess (1980) as completely described in Experimental Procedures. Briefly, we fully denatured the eluted proteins in 6 M guanidine hydrochloride and renatured them by slow dialysis to remove the denaturant.

The ability of each of the renatured proteins to bind with sequence-specificity was assessed by DNAase I footprinting. Figure 2b shows a titration of the renatured yeast 70 kd protein on the yeast *hsp70* promoter. The pattern of protection and dimensions of the footprints observed are indistinguishable from those of native HSTF (compare Figure 2b with Figure 1c or 1d). Similarly, the renatured Drosophila 70 kd protein also generated a footprint on site 4 of its own promoter identical to that of native HSTF (Figure 2a, compare lanes 1 and 5). (Note that sites 1 and 2 are not shown in Figure 2a, because the concentration of renatured Drosophila HSTF was insufficient to protect these sites.) Interestingly, the 25 kd Drosophila protein yielded a weak footprint on site 4 (Figure 2a, lane 3). This suggests that it contains the same DNA binding domain as the 70 kd protein and that is derived by proteolysis of the 70 kd HSTF.

A direct comparison of the sizes of affinity purified Drosophila HSTF and yeast HSTF is shown in Figure 3. Here we observe two apparently homogeneous proteins of identical molecular weights. (In these preparations no contaminating proteins were observed.)

Cross Species Binding of the Drosophila and Yeast HSTFs

When the Drosophila *hsp70* gene is transfected into mouse cells, monkey cells, or sea urchin embryos, it retains transcriptional heat-inducibility (for a review see Pelham, 1985). This implies that components of the transcriptional apparatus, such as the HSTF, are conserved between species. This idea is supported by our observation that the HSTFs from two widely divergent species have nearly identical mass. Further evidence that the two 70 kd HSTFs are homologous comes from a comparison of "cross-species" footprints with "same-species" footprints. Figure 4 reveals that the DNAase I footprints formed by binding of affinity-purified yeast and Drosophila HSTF to the yeast *hsp70* promoter have identical boundaries (compare lanes 2 and 3 to lanes 8 and 9). On site 3, the footprints from both HSTFs extend over a 30 bp region between nucleotides −278 and −309. On site 2, the footprints cover 27 bp between nucleotides −183 and −211. Similarly, the footprints formed by binding of affinity-purified yeast and Drosophila HSTF to the Drosophila *hsp70* promoter also have identical boundaries (compare lanes 5 and 6 with lanes 11 and 12). On site 4, 26 bp between nucleotides −228 and −255 are protected from DNAase I cleavage by both HSTFs. On sites 1 and 2, a 54 bp region between nucleotides −37 and −92 is protected. To summarize, the lengths of the footprints vary slightly in a manner dependent upon their location on the promoter. However, at any one binding site, both the yeast and Drosophila HSTFs always generate footprints with identical boundaries.

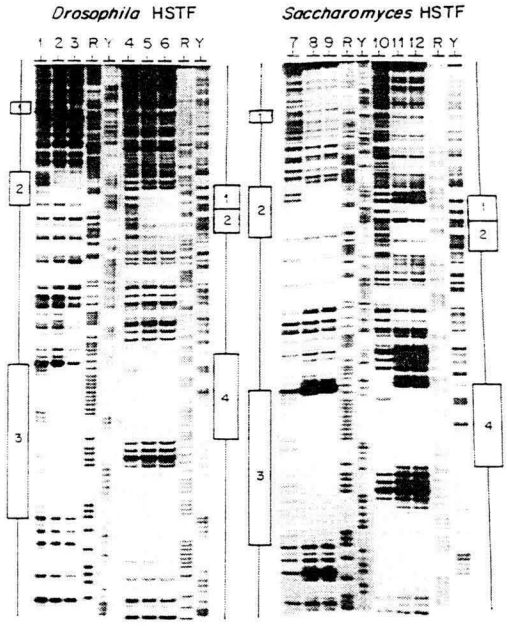


Figure 4. Cross-Species Footprints of Drosophila and Saccharomyces HSTFs

Composition of footprint reactions has been described in the legends to Figures 1 and 2. Lanes 1–3 and lanes 7–9 show footprints on the Saccharomyces *hsp70* promoter and lanes 4–6 and lanes 10–12 on the Drosophila *hsp70* promoter. Lanes 1, 4, 7, and 10: no protein; lanes 2 and 5: 15 ng Drosophila HSTF; lanes 3 and 6: 30 ng Drosophila HSTF; lanes 8 and 11: 16 ng Saccharomyces HSTF; lanes 9 and 12: 32 ng Saccharomyces HSTF; lanes R and Y: chemical cleavage reactions of purines and pyrimidines, respectively. Numbered boxes show locations of binding sites.

The only apparent difference between the footprint patterns is that the yeast HSTF induces a higher degree of DNAase hypersensitivity at the footprint boundaries than does the Drosophila HSTF. For example, the hypersensitivity induced by binding of the yeast HSTF to site 4 of the Drosophila DNA fragment extends 15 bases 3′ from the border of the footprint. Another difference is that binding of the yeast HSTF to the contiguous sites 1 and 2 on the Drosophila *hsp70* promoter leaves two bases accessible to DNAase cleavage that are not accessible when Dro-

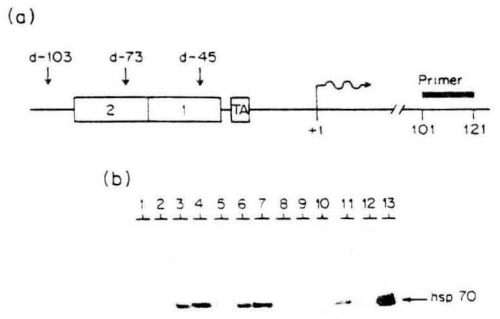


Figure 5. Reconstituted Transcription of the Drosophila *hsp70* Gene with Highly Purified Drosophila HSTF

Transcription reactions (volume, 50 μl) contained 45 μg protein from the heparin-Sepharose flow-through, 30 μg protein from the phosphocellulose 0.8 KCl step, and 250 ng circular template containing the *hsp70* gene, 56H8, deleted to varying degrees at the 5′ end. The deleted templates have been previously described (Topol et al., 1985).
(a) The deletion endpoints, locations of binding sites 1 and 2, the TATA box, the transcription startpoint and the location of the primer.
(b) Lanes 1–7: transcription from the d-103 template; lanes 8–11: transcription from the d-73 template; lane 12: transcription from the d-45 template; lane 13: in vivo *hsp70* RNA. Transcription reactions were reconstituted with HSTF cycled once (lanes 1–4, 8, 9) or twice (lanes 5–7, 10–12) over the sequence-specific DNA column. Increasing amounts of HSTF were added to the transcription reactions as follows. Lane 1: no HSTF; lanes 2 and 5: 1.6 ng HSTF; lanes 3, 6, 8, and 10: 8 ng HSTF; lanes 4, 7, 9, 11, and 12: 16 ng HSTF. Transcription was measured by primer extension assay.

sophila HSTF binds (Figure 4, compare lanes 5 and 6 to lanes 11 and 12). It is not clear what significance should be assigned to the minor differences in the footprints. What is clear, however, is that based upon three criteria, mass equivalence, conserved DNA recognition sequences, and identity of footprint boundaries, the HSTF is a highly conserved transcriptional regulatory protein.

HSTF-Dependent Transcription of *hsp70*

In addition to showing that affinity-purified Drosophila HSTF binds to the correct sequences on the *hsp70* promoter, it was important to determine if affinity-purified HSTF could also stimulate specific transcription of the *hsp70* gene in vitro. Most of the previous HSTF preparations used for binding and in vitro transcription studies were purified less than 1% when compared to the purity of HSTF used in this study. Therefore, it is conceivable that the binding activity and the transcriptional activity assigned to the HSTF in previous preparations were, in fact, the activities of two different proteins. In the present study we used our most highly purified Drosophila HSTF to reconstitute transcription in vitro in an HSTF-dependent Drosophila extract. A 0.8 M KCl step from the initial P11 column and the heparin-Sepharose flow-through contain all factors required, except the HSTF, for transcription of the Drosophila *hsp70* gene. Figure 5 shows that when low levels of Drosophila HSTF are added back to these complementing fractions, there is only a low basal level of *hsp70* RNA transcribed from a template containing binding sites 1 and 2. (Transcription was monitored using circular deletion templates as indicated and primer-extension assays; see Experimental Procedures for complete details.) As increasing amounts of HSTF are added back, there is a 25- to 50-fold stimulation of *hsp70* transcription. As a control, templates lacking binding site 2 and lacking both HSTF binding sites were tested. It is known from previous in vivo (Dudler and Travers, 1984) and in vitro (Topol et al., 1985) analyses that both binding sites must be present for maximal *hsp70* transcription. Figure 5 shows that the same amount of pure HSTF which maximally stimulates transcription from a template containing both binding sites does not reconstitute transcription on templates lacking one or both binding sites (compare lanes 3, 4, 6, and 7 with lanes 8–12).

These results provide compelling evidence that DNA binding and transcriptional stimulation are performed by the same protein. We were not able to test the transcriptional activity of the yeast HSTF because a yeast in vitro RNA polymerase II transcription system does not currently exist. However, preliminary results suggest that the yeast HSTF will not reconstitute the Drosophila HSTF-dependent transcription of *hsp70*. Because the yeast HSTF binds to the HSEs on the Drosophila *hsp70* promoter (this paper), it suggests that binding alone is not sufficient to activate transcription.

Discussion

Sequence-specific DNA-affinity chromatography has been successfully used by Rosenfeld and Kelly (1986) to purify nuclear factor 1 from HeLa cells. In their procedure, multiple copies of the DNA binding site were cloned into a plasmid that was then coupled to cellulose. In the procedure described here, concatamers of a synthetic oligonucleotide containing overlapping HSEs were coupled to Sepharose in order to purify the HSTF 250,000-fold from Drosophila and 86,000-fold from yeast. We have successfully applied this technology to the purification of other sequence-specific DNA binding proteins including an enhancer binding protein (K. Harshman, S. Sogo, and C. S. Parker, unpublished observations). A similar procedure has been developed by Kadonaga and Tjian (1986; also Briggs et al., 1986) to purify the Sp1 transcription factor 30,000-fold. It is probable that DNA-affinity chromatography will aid in the purification of many different sequence-specific DNA binding proteins in the future.

The remarkable similarity in size between the Drosophila and yeast HSTF is suggestive of a great deal of evolutionary conservation. Two caveats to this conclusion, however, must be considered. It is possible that the 70 kd species are actually derived from a larger protein of unknown molecular weight. Although we have made use of strong protease inhibitors and shortened the time of purification to just 48 hr, we cannot rule out the possibility of

limited proteolysis generating identically sized proteins in both Drosophila and yeast. The other issue concerns the yeast HSTF, which has not been formally shown to be a transcription factor. The data presented in this paper show that the yeast protein is identical in size and DNA binding properties to the Drosophila factor. Until it can be rigorously shown in vivo or in vitro that the yeast protein is the HSTF by activity criteria we cannot make any definitive claims at this time. Because the apparent molecular weights of the Drosophila and yeast HSTFs are 70 kd we tested the possibility that they are related to *hsp70* or one of its cognates. Using monoclonal antibodies directed against *hsp70* and its cognates (a gift from Dr. Susan Lindquist), we determined that there was no cross-reactivity with the HSTF.

We have shown that both the Drosophila and yeast HSTFs can be further purified through SDS protein gels, renatured, and observed to footprint identically compared to the native proteins. This method, originally developed by Hager and Burgess (1980), should also be generally applicable to many DNA binding proteins. Indeed, we have also applied this technique to an enhancer binding protein and unambiguously identified the protein responsible for the observed activity (K. Harshman, unpublished data). Should a preparation always contain multiple species, this technique may be very useful to determine which of the proteins can bind with sequence-specificity.

The affinity-purified Drosophila HSTF stimulates specific transcription from an *hsp70* template in an HSE-dependent fashion. This data suggests that a native full-length HSTF has been isolated. One concern with this conclusion is that the number of active templates in the reconstitution reaction is less than 10%. Hence it is formally possible that a minor component in the preparation is actually responsible for the transcriptional stimulation observed. For this to be true, however, the minor species would have to possess a significantly higher binding constant than the 70 kd protein. This is because the 70 kd protein is present in at least a 10-fold excess over the putative minor species (based on silver staining) and the 70 kd protein is present in sufficient quantities to saturate all of the templates in the transcription reaction.

One of the goals of our studies on the heat shock response has been to clone the gene for the HSTF. Cloning and characterizing the HSTF gene will help to confirm the native molecular weight of the factor. We recently were successful in obtaining an N-terminal sequence of a proteolytic fragment of the yeast HSTF. With this sequence we have generated several sets of oligonucleotides which will allow us to clone the gene from a yeast λgt10 library (G. Wiederrecht and C. S. Parker, unpublished observations). Additionally, monoclonal antibodies that we have raised against the HSTF will be useful for studying the pre- and post-heat shock cellular and chromosomal locations of the factor. Clearly, having the HSTF gene and the antibodies will allow a large number of different and interesting experiments to be performed, leading to a more complete understanding of the molecular mechanism of heat shock gene activation.

Experimental Procedures

Yeast Strain, Growth of Yeast, and Processing

The protease-deficient Saccharomyces cerevisiae strain EJ926

$$\left(MAT\ \frac{\alpha}{\alpha}\ \frac{trp1}{+}\ \frac{+}{his1}\ \frac{prc1-126}{prc1-126}\ \frac{pep4-3}{pep4-3}\ \frac{prb1-1122}{prb1-1122}\ \frac{can1}{can1}\right)$$

obtained from E. Jones (Carnegie-Mellon) was the strain used for the isolation of the HSTF. The yeast were grown in a 350 liter fermenter at 30°C (heat-shocked) in medium containing Bacto-yeast nitrogen base supplemented with 2% dextrose, 50 μg/ml histidine, and 50 μg/ml tryptophan. When the culture reached an $O.D._{600}$ of 5.0, they were harvested with a Sharples continuous flow centrifuge. The yeast paste was resuspended in a buffer solution containing 1 M sorbitol, 50 mM Tris (pH 7.8), 10 mM $MgCl_2$, and 30 mM DTT. This buffer solution minus the DTT is designated Y buffer. The suspension ratio was 1 kg yeast/liter buffer. The suspension was quick-frozen as small droplets in liquid nitrogen and stored at −80°C.

Nuclear Extract Preparation

Typically, the amount of starting material for a yeast extract was 1.1 kg of the frozen suspension described above. The frozen material was thawed and then subjected to centrifugation (2200 × g; 5 min) in order to pellet the yeast. The pellet was resuspended in three volumes of Y buffer plus 3 mM DTT. Zymolyase 100 T (Miles) was added (100 mg/kg yeast) to the suspension and was then incubated for 1 hr at 37°C. After the zymolyase treatment, PMSF (Sigma) was added to a concentration of 1 mM. The zymolyase-treated cells were subjected to centrifugation (2200 × g; 5 min) and the pellet was resuspended in Y buffer plus 1 mM PMSF. This material was subjected to centrifugation (2200 × g; 6 min). The soft pellet was resuspended in A buffer (10 mM Hepes [pH 7.6], 15 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA) plus 1 mM PMSF and then placed in an ice bath for 20 min. The suspension was homogenized (4–5 strokes) in a large Dounce homogenizer. The homogenate was subjected to centrifugation (18,000 × g; 10 min) and the pellet was resuspended in A buffer plus PMSF such that the final volume was about 1200 ml. The solution was divided equally among eighteen 45 Ti polycarbonate tubes. Ammonium sulfate was added to each tube to a final concentration of 0.25 M and diisopropyl-fluorophosphate (DIFP; Sigma) was added to a final concentration of 0.4 mM. The tubes were rotated on a rotation apparatus for 15 min and then subjected to centrifugation in a 45 Ti rotor (35,000 rpm, 2 hr). Ammonium sulfate was added to the supernatant (0.2 g $[NH_4]_2SO_4$ added per ml supernatant) and the precipitated protein was pelleted by centrifugation (18,000 × g; 45 min). The protein pellet was resuspended in 90 ml of a buffer solution containing 25 mM Hepes (pH 7.6), 10% (v/v) glycerol, 50 mM KCl, 0.1 mM EDTA, and 0.2% Triton X-100 (abbreviated as 0.05 HGKET where 0.05 is the molarity of KCl). The solution was dialyzed overnight against 4 liters of the same buffer plus 1 mM PMSF. Insoluble material was removed by centrifugation (10,000 × g, 10 min) and the supernatant was subjected to the chromatographic techniques described below.

Nuclear extracts derived from K_c cells were prepared essentially as described previously (Parker and Topol, 1984a) with the addition of protease inhibitors as described above.

Chromatographic Procedures

The yeast nuclear extract (total volume, 90 ml) was applied to a column (volume, 100 ml) of heparin-Sepharose equilibrated with 0.05 HGKE. The column was washed with three column volumes of 0.05 HGKE, five column volumes of 0.32 HGKE, and then three column volumes of 0.8 HGKE. Individual fractions (volume, 15 ml) were assayed for HSTF binding activity and the active fractions from the 0.8 step were dialyzed against 0.1 HGKET. Following dialysis, insoluble material was pelleted by centrifugation (12,000 × g; 10 min) and the supernatant was applied to a column (volume, 20 ml) of Affigel Blue at a flow rate of 20 ml/hr. The column was washed with 10 column volumes of 0.05 HGKE, 10 column volumes of 0.5 HGKE, and 20 column volumes of 1.5 HGKET. The fractions from the 1.5 HGKET step were combined, dialyzed against 0.1 HGKET, and then applied to the sequence-specific oligodeoxyribonucleotide column as described below.

The dialyzed K_c nuclear extract (total volume, 35 ml) was applied to a column (volume, 50 ml) of phosphocellulose (P11; Whatman) equilibrated in 0.1 HGKET, and then washed with the same buffer. Protein-containing fractions in the flow-through were combined and applied to a column (volume, 10 ml) of heparin-Sepharose equilibrated with 0.1 HGKET. The column was washed with three column volumes of 0.1 HGKET and HSTF was then step-eluted with three column volumes of 0.4 HGKET. The step-eluted material was dialyzed against 0.1 HGKET, subjected to centrifugation (12,000 × g) to remove insoluble material, and the supernatant applied to the sequence-specific DNA column as described below.

The procedures for chromatography of the partially purified yeast and Drosophila HSTF preparations on the sequence-specific DNA column were nearly identical. The partially purified preparations were loaded onto a 1.1 ml column (constructed as described below) at a flow rate of 15 column volumes per hr. The column was washed with 10 volumes of 0.1 HGKET, 100 volumes of 0.35 (0.375 for yeast HSTF) HGKET, and the HSTF was step-eluted with 10 column volumes of 1.5 HGKET. The HSTF was dialyzed overnight against 0.05 HGKET and loaded onto a 0.2 ml sequence-specific DNA column. Chromatography was performed as described for the larger column.

SDS Gel Electrophoresis and Renaturation

HSTF-containing fractions from the sequence-specific DNA column were combined and the protein concentrated by the addition of 3 volumes acetone. The mixture was frozen in a dry ice–ethanol bath and the protein pelleted by centrifugation (12,000 × g; 30 min) at 0°C. The supernatant was discarded and the protein pellet was vacuum-dried. The pellet was resuspended in SDS protein gel loading buffer, heated at 70°C for 5 min, and applied to a 10% SDS polyacrylamide gel (Laemmli, 1970). Electrophoresis was performed at 3°C. The gel was soaked in 250 ml of 0.25 M KCl for 1–2 min. Staining for longer periods only raised the background. Gel slices containing potassium-SDS-protein precipitates were excised and placed into small dialysis bags containing 600 μl of SDS gel running buffer. Protein was electro-eluted for 3 hr at 250 V by placing the dialysis bags on a horizontal gel box filled with SDS gel running buffer. The dialysis bags containing the HSTF were dialyzed for 3 hr against 0.05 HGKET to remove glycine. The protein was precipitated by the addition of 3 volumes of acetone followed by freezing in a dry ice–acetone bath. The precipitate was pelleted by centrifugation (12,000 × g; 30 min). The pellet was washed with ice-cold 80% acetone to remove residual SDS. Typically, all of the HSTF from a yeast preparation was resuspended in 600 μl 0.1 HGKET containing 6 M guanidine hydrochloride. All of the HSTF from a K_c cell preparation was resuspended in 150 μl of the denaturing buffer. The guanidine hydrochloride was dialyzed out overnight against 0.1 HGKET. We estimate the yield of specific binding activity resulting from this procedure to be approximately 15%.

Preparation of the Sequence-Specific Oligodeoxyribonucleotide Column

Two complementary synthetic oligodeoxyribonucleotides modeled after the arrangement of overlapping HSEs in the Drosophila *hsp83* promoter were prepared at Caltech's Core Microchemical Facility. The sequence of strand 1 was:

5′-AGCTTCTAGAACGTTCTAGAAGCTTCGAGA-3′

The sequence of strand 2 was:

5′-AGCTTCTCGAAGCTTCTAGAACGTTCTAGA-3′

The strands were separately suspended in 7 M Urea, 1× TBE, and dyes and purified on a 20% polyacrylamide–7 M Urea gel. The DNA was visualized under shortwave UV light, over a fluorescent silica gel TLC plate. The band containing the DNA was excised from the gel, the gel slice was crushed, and the DNA was eluted into a 5 ml solution containing 0.5 M sodium acetate, 10 mM magnesium acetate, 1 mM EDTA, and 0.1% (w/v) SDS for 6 hr at 37°C. This elution was repeated twice and the pooled eluate was concentrated to 1.5 ml by extraction with sec-butanol. The concentrated DNA was desalted on a Sephadex G-25 column. Equimolar amounts of DNA-containing fractions (determined by absorbance at 260 nm) were combined, frozen, and lyophilized. The dry pellet was resuspended in 0.5 ml of a solution containing 100 mM $NaHCO_3$ (pH 8) and 330 mM NaCl. This solution was heated to 90°C for 10 min, and allowed to cool slowly for several hours. The annealed strands (5 mg) were precipitated with 95% ethanol, washed with 70% ethanol, and resuspended in 0.5 ml of 10 mM Tris (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, and 1 mM ATP. The annealed DNA was phosphorylated with T4 polynucleotide kinase (100 units, BMB) at 37°C for 1 hr. Two subsequent additions of 1 mM ATP were made at 30 min intervals. Following this 2 hr incubation, a final 1 mM ATP addition was made and 20 μg of T4 DNA Ligase (gift from S. Scherer) was added. Ligation was performed at 14°C for 12 hr. The ligated DNA was covalently attached to Sepharose CL-4B as described by Arndt-Jovin et al. (1975). Activated Sepharose was neutralized with 0.1 M ethanolamine. We calculate that 250 μg of sequence-specific DNA was coupled per ml of resin.

Drosophila and Saccharomyces *hsp70* Promoter Constructions

Footprinting studies performed on the Drosophila *hsp70* promoter employed a fragment lacking HSE 3. The fragment, d21$^+$, contains an upstream BglI (EcoRI)–XhoI (HindIII) fragment extending from −358 to −184 on the *hsp70* gene, 56H8 (Artavanis-Tsakonas et al., 1979), ligated to a 5′ Bal31 deletion construct extending from −103 (HindIII) to a BglI site at −384.

The yeast *hsp70* promoter fragment used in the footprinting studies is derived from YG100 (Ingolia et al., 1982) and was a gift from E. Craig. A HincII–RsaI fragment derived from YG100 and containing HSEs 1, 2, and 3 was cloned into the polysite linker of pUC9. This clone was designated pYHS-1.

Preparation of Labeled Fragments and Footprint Reactions

The Drosophila d21$^+$ fragment was labeled on the coding strand with T4 polynucleotide kinase and γ-^{32}P-ATP near the BglI site (where an EcoRI linker was ligated and thus labeled by cutting with EcoRI) 358 bp upstream of the transcription startpoint. The pYHS-1 fragment was labeled on the coding strand by phosphorylation near the HincII site (at the HindIII site of the pUC9 polylinker) 337 bp upstream of the transcription startpoint. Footprinting reactions contained, in a total volume of 25 μl, 1–5 ng of labeled fragment, 50 μg/ml poly dA · dT (Pharmacia), 5 mM $MgCl_2$, and HSTF as indicated. Footprinting reactions were performed at 0°C and DNAase I (Cooper) was added to a final concentration of 50 μg/ml. After 30 sec of DNAase digestion, the reaction was terminated by the addition of 100 μl of transcription reaction-termination buffer which also contained 100 μg/ml proteinase K (EMS Labs) and 25 μg/ml calf thymus DNA (Sigma) and treated as previously described (Parker and Topol, 1984a).

In Vitro Transcription Assays

The composition of the transcription reactions has been described (Parker and Topol, 1984b). The plasmid templates used were 5′ Bal31 deletions of 56H8 and are designated d-103, d-73, and d-45. These templates have been previously described (Topol et al., 1985) and contain binding sites 1 and 2, site 1, and no HSTF binding sites, respectively. The transcription reactions were allowed to proceed for 45 min at 22°C, terminated with transcription reaction-termination buffer, extracted with phenol, and the nucleic acids precipitated with 90% ethanol, 0.1 M NaOAc. The nucleic acids were resuspended in 10 μl TE (10 mM Tris [pH 8], 1 mM EDTA) containing a 7.1 picomoles of 5′ end-labeled synthetic primer homologous to RNA sequences between 100 and 121 nucleotides downstream of the *hsp70* cap site. Two microliters of a buffer solution containing TE and 1.25 M KCl was added, the reaction mixture was heated to 65°C and allowed to anneal by slowly cooling to room temperature. Primer extension was performed by adding 25 μl of a buffer solution containing 20 mM Tris (pH 8.7), 10 mM $MgCl_2$, 5 mM DTT, 300 μM deoxyribonucleotides, 10 μg/ml actinomycin D, and 0.5–2 U AMV reverse transcriptase (Boehringer-Mannheim). The reaction was incubated at 37°C for 30 min. Nucleic acids were precipitated with 110 μl 90% ethanol, 0.1 M NaOAc, and analyzed on an 8% polyacrylamide–6 M Urea sequencing gel.

Acknowledgments

We wish to thank Dr. Suzanna Horvath of the California Institute of Technology Microchemical Facility for her expert preparation of the oligonucleotides used in these experiments. We are also indebted to

Dr. Susan Lindquist for her gift of the anti-hsp 70 and hsp 70 cognate monoclonal antibodies and to Dr. Elizabeth Craig for the YG100 clone and unpublished sequence data. We thank Drs. Norman Davidson, Eric Davidson, Judy Campbell, Joan Kobori, and Susan Manly for their critical readings and Phoebe Ray for her expert preparation of the manuscript.

G. W. is a recipient of a Helen Hay Whitney postdoctoral fellowship. D. J. S. and W. A. K. are National Institutes of Health predoctoral trainees (GM 07616). This research was supported by a grant from the ACS (NP-0482) and NIH (GM 29430) to C. S. P. C. S. P. is a Rita Allen Foundation Scholar.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 14, 1986; revised November 26, 1986.

References

Arndt-Jovin, D. J., Jovin, T. M., Bähr, W., Frischauf, M., and Marguardt, M. (1975). Covalent attachment of DNA to agarose. Eur. J. Biochem. *54*, 411–418.

Artavanis-Tsakonas, S., Schedl, P., Mirault, M.-E., Moran, L., and Lis, J. (1979). Genes for the 70,000 dalton heat shock protein in two cloned D. melanogaster DNA segments. Cell *17*, 9–18.

Bram, R., and Kornberg, R. (1985). Specific protein binding to far upstream activating sequences in polymerase II promoters. Proc. Natl. Acad. Sci. USA *82*, 43–47.

Briggs, M., Kadonaga, J., Bell, S., and Tjian, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor Sp1. Science *234*, 47–52.

Carthew, R. W., Chodosh, L. A., and Sharp, P. A. (1985). An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell *43*, 439–448.

Craig, E. A., and Jacobsen, K. (1984). Mutations of the heat inducible 70 kilodalton genes of yeast confer temperature sensitive growth. Cell *38*, 841–849.

Dudler, R., and Travers, A. A. (1984). Upstream elements necessary for optimal function of the hsp 70 promoter in transformed flies. Cell *38*, 391–398.

Dynan, W. S., and Tjian, R. (1983). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell *32*, 669–680.

Giniger, E., Varnum, S. M., and Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell *40*, 767–774.

Goldberg, M. (1979). Ph.D. Thesis, Stanford University, Stanford, CA.

Guarente, L. (1984). Yeast promoters: positive and negative elements. Cell *36*, 799–800.

Hager, D., and Burgess, K. (1980). Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity. Anal. Biochem. *109*, 76–86.

Hope, I. A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell *46*, 885–894.

Ingolia, T., Slater, M., and Craig, E. (1982). Saccharomyces cerevisiae contains a complex multigene family related to the major heat-shock inducible gene of Drosophila. Mol. Cell. Biol. *2*, 1388–1398.

Jones, K. A., Yamamoto, K. R., and Tjian, R. (1985). Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell *42*, 559–572.

Kadonaga, J., and Tjian, R. (1986). Affinity purification of sequence-specific DNA binding proteins. Proc. Natl. Acad. Sci. USA *83*, 5889–5893.

Keegan, L., Gill, G., and Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science *231*, 699–704.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

Miesfeld, R., Rusconi, S., Godowski, P. J., Maler, B. A., Okret, S., Wikström, A.-C., Gustafsson, J.-Å., and Yamamoto, K. R. (1986). Genetic complementation of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. Cell *46*, 389–399.

Parker, C. S., and Topol, J. (1984a). A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA binding activity. Cell *36*, 357–369.

Parker, C. S., and Topol, J. (1984b). A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. Cell *37*, 273–283.

Pelham, H. (1985). Activation of heat-shock genes in eukaryotes. Trends in Genet. *1*, 31–35.

Richards, R. I., Heguy, A., and Karin, M. (1984). Structural and functional analysis of the human metallothionein-I_A gene: differential induction by metal ions and glucocorticoids. Cell *37*, 263–272.

Rosenfeld, P., and Kelly, T. (1986). Purification of nuclear factor 1 by DNA recognition site affinity chromatography. J. Biol. Chem. *261*, 1398–1408.

Sawadogo, M., and Roeder, R. G. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell *43*, 165–175.

Shuey, D. J., and Parker, C. S. (1986a). Binding of Drosophila heat-shock gene transcription factor to the hsp 70 promoter. J. Biol. Chem. *261*, 7934–7940.

Shuey, D. J., and Parker, C. S. (1986b). Bending of promoter DNA on binding of heat shock transcription factor. Nature *323*, 459–461.

Topol, J., Ruden, D. M., and Parker, C. S. (1985). Sequences required for in vitro transcriptional activation of a Drosophila *hsp70* gene. Cell *42*, 527–537.

Wu, C. (1984). Two protein binding sites in chromatin implicated in the activation of heat shock genes. Nature *309*, 229–234.

Chapter 6

Summary

Summary

Due to the temporal gap between thesis completion and submission, this discussion will be presented in a bipartite fashion.

Part one will focus on a review of the data presented within Chapters 1-4 and the contemporaneous interpretations that were made. Additionally this section will include a presentation of the two significant technological advancements made during these studies, namely the development of the gel-shift contact point methodology and the sequence-specific DNA affinity chromatography procedure.

Part two will incorporate data generated since the completion of these studies and will focus on reinterpreting our results to reflect these scientific advances. This re-evaluation will incorporate the two emerging central theories of HSTF-HSE interactions; the redefinition of the HSE and the apparent trimeric structure of the active HSTF.

Discussion – part 1

Chapter 1 – Gel-shift contact point methodology and HSTF-Hsp70 promoter studies

Chapters 1 through 3 of this thesis provide insights into the underlying molecular mechanisms of a particular aspect of heat shock transcriptional induction; namely the

DNA binding properties of the *Drosophila* Heat Shock Transcription Factor (HSTF). Recall that in this section data are viewed in the context of the historical definition of the cis acting heat shock promoter element; the HSE. This element was first identified by homology and described by Pelham (18,34) as a 14 bp sequence found in the promoters of numerous heat shock genes. This 14 bp consensus sequence (5'-CnnGAAnnTTCnnG-3') is referred to throughput chapters 1-4 as an HSE. Notably this element exhibits hyphenated dyad symmetry and is reminiscent of many well defined procaryotic DNA binding sites that are occupied by rotationally symmetric protein multimers, usually dimers or tetramers (30).

Concurrent *in vitro* and *in vivo* studies demonstrably showed that the existence of a single HSE within a promoter could not provide appreciable heat shock transcriptional activation (28,29). At this time, other groups also provided evidence that both procaryotic and eucaryotic regulatory proteins, by virtue of their sequence-specific DNA binding to upstream cis-acting elements, may also alter the tertiary structure of the DNA molecule itself (35-37). It was in this environment that we proceeded with the experiments described below.

In an attempt to gain insights into the mechanisms of heat shock gene activation, we pursued a series of studies designed to reveal the molecular determinants of the HSTF-HSE interaction. Due to the state of current available technologies these analyses required the invention of a novel contact point technology. Initially, numerous attempts were made to employ existing filter binding methods (38) to study

the contact points of specific HSTF-DNA complexes. These experiments failed to sufficiently resolve specifically bound HSTF-DNA complexes from non-specific ones under saturating equilibrium binding conditions, presumably due to the presence of contaminating DNA binding proteins in our HSTF preparations that could also trap labeled DNA template probes to the nitrocellulose filters. Recall that the studies presented in chapters 1-3 were performed with only partially purified HSTF prior to the development of the sequence-specific affinity isolation method presented in Chapter 4. Also, at this time no anti-HSTF antibody was available to allow utilization of the immunoprecipitation complex analysis technique (39).

In retrospect, these apparent limitations were in fact fortuitous circumstances that demanded the development of a more flexible and adaptive gel-based methodology (40). The data quality generated by this novel technology was generally superior to existing methods, presumably resulting from a more efficient the partitioning of specifically bound protein-DNA complexes from unbound DNA or non-specifically bound complexes. Also, as will be discussed in detail below, this method is the only one currently available that can resolve specific protein-DNA complexes of differing molecular weights and thus facilitates the contact point analysis of such potentially interesting "intermediate" homogeneous protein-DNA complexes or even complexes formed with more than one sequence-specific binding protein.

In developing this technique numerous polymer gel systems were tried including low percentage and low crosslink content polyacrylamide gels (currently used by the

majority of researchers), agarose-polyacrylamide hybrid gels, and high percentage agarose gels. We found that although the latter gel system was significantly more difficult to cast, the resolution properties for high molecular weight protein-DNA complexes observed for the HSTF-HSE studies was superior. Additionally, extended electrophoresis times with these agarose gels at reduced temperatures (4°C) seemed to stabilize HSTF off-rates and facilitated the high resolution separation of intermediate complexes that other groups have consistently failed to identify.

In conclusion, the protein-DNA complex gel separation technique presented offers several advantages over previously existing methods. One can generate high quality reproducible contact point and footprinting data from relatively crude protein samples at sub-saturating concentrations. Also, an antibody to the protein of interest is not required. Perhaps most importantly, it is possible to resolve discrete complexes formed at neighboring and interactive sites and analyze these protein-DNA interactions in detail. This technique is still widely utilized and generally regarded as the method of choice for these studies (41). This technical breakthrough enabled the detailed analysis of the HSTF-HSE interactions at the *Drosophila* hsp70 and hsp83 gene promoters. Chapters 1 and 3 of this thesis describe a previously undocumented protein-DNA binding phenomenon; namely the apparent sequence-specific template-directed HSTF polymerization at two HSE containing promoters.

Enzymatic and chemical probes were used to study the HSTF interactions formed at both the hsp70 and hsp83 gene promoters. When the HSTF-hsp70 interactions were

initially examined it appeared that at low concentrations the HSTF binds to a high affinity site (site 1) with subsequent cooperative binding to site 2 at higher concentrations. These studies exhibited marked contact point symmetry and suggested that the simplest explanation for the data would entail a dimeric HSTF bound to site 1 to generate the lower molecular weight complex A followed by the cooperative binding of a dimeric HSTF to site 2 forming complex B (40). Interestingly, a small subset of the site 1 contacts including the distal consensus site reverse complement guanosine (base-paired to C1 of the 14 bp site 1 HSE), changed upon sequential HSTF binding to the neighboring site 2. As it was known that heat inducubility required both HSE's, we postulated that this apparent HSTF conformational change perhaps facilitated protein-protein interactions with the basal machinery and thus provided the "trigger" for heat-shock gene activation.

Subsequent studies using this same hsp70 template (chapters 2 and 3) revealed a transitory complex of intermediate apparent molecular weight, complex B*, that in addition to site 1 occupancy appeared to have only an HSTF monomer bound to the proximal half of site 2. Curiously, for the hsp70 promoter, occupancy seems to nucleate with a bound dimer, but then appears to polymerize distally with incremental HSTF binding to generate three discrete complexes of increasing apparent molecular weights.

Chapter 2 – HSTF induces an HSE directed bend in hsp70 promoter DNA

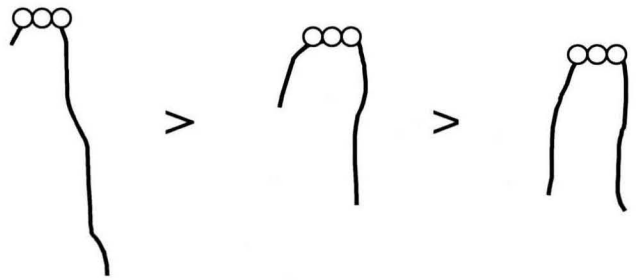
Several studies of procaryotic transcriptional activators strongly suggested that these DNA binding proteins may in fact alter the conformation of their respective target site. In lambda phage, the cI repressor bound to operators separated by five to six turns of the helix appear to involve a DNA bending mechanism (35). Also, it was postulated that repressors bound to DNA operators separated by 150 bp may actually contact each other and promote the "looping out" of a section of promoter DNA at the ara BAD *E. coli* promoter (36). Concurrently, studies of SV40 and the relative spatial distribution of distally positioned elements of this viral promoter provided evidence that such elements may direct factor interactions with the basal transcriptional machinery from a distance (37). This group also proposed a DNA bending or "looping" mechanism as a possible explanation for such seemingly surprising results.

At this time, our observations obtained from the gel shift contact point studies at the *Drosophila* hsp70 gene promoter suggested that a conformational change may accompany HSTF binding to sites1 and 2. This was indicated by a subset of contacts within site 1 changing upon sequential HSTF occupancy at site 2 (see Chapter 1 and below). This observation along with the recently emerging data from others described above made it desirable to investigate these protein-DNA complexes to determine to what extent DNA structural alterations, if any, might contribute to these observed contact point changes.

It has been shown that the gel electrophoretic mobility of identically sized DNA fragments harboring a specific bending locus is dependent upon the relative position

of the bend in relation to the ends of the DNA fragment (42). When a bend is located centrally the DNA will migrate with a larger apparent molecular weight that when it is at the end of the molecule. Please see figure D-1 below.

Figure D-1: Gel migration of "bent" DNA



The application of electrophoretic analayses to DNA bending phenomena was originally demonstrated with "naked" or protein-independent telomeric sequences and this gel-based technology was quickly extended to the analysis of protein-induced DNA bending loci (42). A cloning strategy for the analysis of potential HSTF induced bending was therefore developed and resulted in a plasmid harboring the two hsp70 HSE's flanked by direct repeats of spacer procaryotic DNA from pBR322. The resulting plasmid could then be restricted with numerous enzymes located within the pBR322 sequences to generate DNA fragments of identical size and base composition (43). The relative location of the HSE's is thus transposed within this set of binding templates (Chapter 2; figure 2).

When the five separately restricted HSE containing templates were prepared and labeled (described in Chapter 2) and subjected to gel-shift analysis, it was clear that

those templates with distally located binding elements migrated significantly faster than those with more centrally distributed HSE's. It was therefore postulated that the HSTF may in fact bend DNA upon binding to the HSE's 1 and 2 of the *Drosophila* hsp70 promoter. The Complex A bend center mapped to the center of the site 1 HSE and, as one would expect, this bend center appeared to shift distally towards site 2 when the larger molecular weight complexes (B* and B) were analyzed and compared to complex A. An accurate extrapolation and curve fit of these extremely slowly migrating complexes was difficult and this apparent distal shift was within the statistical margins of our error determinations. All five HSE templates migrated identically on agarose gels in the absence of HSTF, demonstrating that the bend was indeed HSTF-induced. Unfortunately due to a publisher's printing error this critical control data was omitted from our manuscript in Chapter 2 (figure 3b). This data is thus included in an appendix to Chapter 2.

In an attempt to improve resolution of these "bent" HSTF-HSE complexes we dramatically extended the electrophoresis times. This alteration revealed an extremely interesting and previously undetected HSTF-HSE intermediate formed at the hsp70 promoter; namely Complex B*. See figure 3a of Chapter 2. This minor band migrates much more slowly than the previously defined Complex A and at a slightly lower apparent molecular weight than Complex B. Complex B* appears transiently with increasing concentrations of HSTF and may represent a transitional species between site 1 occupancy (Complex A) and sites 1 and 2 occupancy

(Complex B). A detailed exploration of this complex is described in Chapter 3 below.

It should be noted that theory of gel-based determinations of DNA bending is still a poorly understood phenomenon that appears to involve the radius of raptation of DNA molecules in relation to the average pore size of the gel medium (44,45). Our argument is thus almost purely by analogy to previously documented anomalous gel migration patterns of "bent" DNA molecules that were also confirmed by additional biophysical methods (42). Also of note is that subsequent co-crystallization studies failed to detect any HSTF-induced DNA bending (46,47). These apparently incongruent results are discussed in part 2 below.

Chapter 3 – Intermediate HSTF-HSE complexes observed at Hsp 83 and Hsp 70 promoters; implications for HSTF polymerization

The *Drosophila* hsp83 gene promoter was also examined by the same enzymatic and chemical probes used to analyze the hsp70 promoter described above and even included an additional tool. In an attempt to gain a more sequence-neutral perspective on the boundaries of HSTF-HSE interactions, we expanded our repertoire of footprinting reagents to include methidiumpropyl-EDTA-iron(II) (or MPE) (48,49). This reagent intercalates into the minor groove of B-form DNA and is significantly smaller and less sequence specific than DNaseI, enabling higher resolution mapping of the boundaries of the HSTF-HSE interactions. Using these

methods to perform high resolution contact point studies on the hsp83 promoter it was revealed that the central HSE (site 1) appears to be occupied at low concentrations of HSTF, followed by the sequential binding to the flanking HSE's (sites 2 and 3) at higher protein concentrations. See Chapter 2; figure 6 for computer generated graphics display.

The Hsp83 data presented above along with previously described intermediate complexes observed at the hsp70 gene promoter revealed in the bending studies (Chapter 2) motivated a reevaluation of our previous analysis of this interaction. When high-resolution gel shift contact point experiments were performed with the *Drosophila* hsp70 promoter, it was clear that the originally described complex B could in fact be resolved into two distinct protein-DNA complexes (B and B*). DMS methylation footprint analysis of these complexes suggest that complex A consists of HSTF binding to the TATA proximal HSE of the Hsp70 gene promoter (40) as previously postulated. Interestingly, resolved complexes B and B* exhibit marked differences in the methylation patterns seen at the distal end of the second HSE suggesting that the B* complex represents a discrete intermediate between the more prevalent HSTF-HSE complexes seen in A and B. Unlike the Hsp83 promoter, however, the Hsp70 binding seems to nucleate first at the strong TATA-box proximal HSE site 1 and binding sequentially expands distally.

We originally postulated that for both of these HSTF-HSE interactions, a dimeric HSTF binds first at a high affinity site (site 1) with subsequent monomeric HSTF

binding at site 2 and site 3 to form a stable complex consisting of four HSTF monomers bound to HSE elements of the both hsp83 and hsp70 promoters. In each case it appears that the HSTF is actually capable of sequence-specific template-directed polymerization . Please see Chapter 3; figure 7 for a cartoon representation of these previously undescribed DNA binding phenomena. This conclusion will also be revisited below.

Chapter 4 - HSTF sequence-specific DNA affinity purification

Transcription factors, as most regulatory proteins, are generally present in extremely limited quantities in mammalian cells. For the experiments described in Chapters 1-3 above, HSTF was partially purified from 10-30 liter cultures by traditional chromatographic methods involving several ion-exchange steps (DEAE-cellulose or Sephadex and Phospho-cellulose), heparin-Sepharose and non-specific salmon sperm dsDNA-cellulose. The latter two steps can be viewed as pseudo-affinity chromatography. This protocol is labor intensive, takes 2-3 days and results in HSTF preparations that were estimated by silver staining to be only 5-10% pure.

It was clearly desirable to develop a more facile and efficient HSTF purification method. Recently available preparative oligonucleotide synthetic capabilities in addition to our evolving understanding of the sequence directed and cooperative nature of the HSTF-HSE interaction prompted us to investigate sequence-specific oligonucleotide affinity purification regimens. The previously mentioned HSTF binding cooperativity studies (29) revealed that smaller HSE containing DNA

restriction fragments actually competed less well than larger ones harboring identical HSE sequences (D. Ruden unpublished observations). This prompted us to design a small oligonucletide elements that could be ligated into longer concatameric binding templates. The 30-mer oligonucleotide design was modeled after the overlapping HSE arrangement found at the Hsp83 gene promoter and following annealing and ligation would theoretically consist of a repeating -nGAAnnTTCn- HSE element. In the contemporary definition of the HSE, refer to "Discussion - Part 2", this ligated structure could be schematically represented as a repeating $(\rightarrow\leftarrow\rightarrow\leftarrow\rightarrow\leftarrow)_n$ element.

Suzanna Horvath at the Caltech Microchemical Facility generously provided a 1 micromole synthesis of the two complimentary 30-mer oligonucleotides described in Chapter 4 (approximately 4 mg yield for each oligonucleotide following PAGE purification and elution). These were preparatively annealed and kinased under standard conditions. It was clear that an efficient ligation of this magnitude would require an inordinate amount of T4 DNA ligase to generate binding and coupling templates of the desired length; i.e. greater than 5-mers or at least 150 bp in length. A T4 DNA ligase lambda lysogen *E. coli* expression strain was obtained (generous gift of Dr. Roger Kornberg) and we proceeded to purify approximately 10 mg of this enzyme that enabled the high efficiency ligation of the 30-mer oligos described above (D. Shuey unpublished results). Our ligation reactions utilizing the annealed HSE 30-mers and the purified recombinant T4 DNA ligase resulted in an average concatamerization number of 10 or greater (i.e. >300 bp avg. length) with little

evidence of circularization. These were then coupled to Sepharose resins via the CNBr method (50) and the resulting resins were termed HSE83-Sepharose.

Preliminary chromatography steps (usually heparin-Sepharose or Affigel-Blue) were often but not always included in an attempt to limit the amount of protein that was loaded onto the HSE83-Sepharose resin and therefore minimize recovery and purity problems associated with the overloading of most affinity columns. It was also determined that adjusting these columns to extremely slow flow rates not exceeding 2 column volumes/hr, often extending for 6-12 hours, was essential for both HSTF recovery and purity. Presumably this was attributable to the fact that non-specific contaminants which were predicted to have greater off-rate constants and thus could more easily be washed away. Purified HSTF from both yeast and *Drosphila* were purified by these methods and analyzed on PAGE gels.

Silver stained gels provided a crude mass estimate of these affinity purified HSTF preparations. Subsequent binding titrations yielded specific activity estimates that seemed to indicate that the 70 Kd species had to be the active DNA binding protein. It was, however, possible that multiple HSTF species arising from proteolysis existed that might negate this inference. Thus it was desirable to attempt to isolate these 70 Kd proteins and assay them for specific HSE binding activity. The presumptive yeast and *Drosphila* 70 Kd HSTF's were fractionated by PAGE, stained with KCl, and recovered by electroelution. After acetone precipitation, these proteins were fully denatured in guanadinium-hydrochloride. Specific HSE binding activity was

recovered following a slow renaturation dialysis procedure. This provided definitive evidence that the major bands on the silver stained gels, the 70 Kd species, were in fact the HSTF.

Employing this technology, HSTF was purified to apparent homogeneity (>95%) from both *Drosophila* and yeast. Preparative isolation of yeast HSTF by Dr. Greg Weiderrhecht and others allowed the generation of anti-HSTF antibodies and valuable confirmatory amino acid sequences of peptide fragments that eventually led to the isolation of the *Saccharomyces cervisiae* HSTF gene (51). This technology also facilitated the homogeneous affinity purification of a number of other yeast transactivators in the Parker lab including ySP1 (K. Harshman, unpublished results) and yAP1 (52). This methodology proved invaluable in reproducibly generating preparative quantities of highly purified HSTF for use in subsequent biochemical studies, including reconstituted transcription driven from exogenous hsp gene promoters.

It should be noted that a similar oligonucleotide based affinity purification approach was developed concurrently at and utilized to isolate the mammalian Sp1 transcription factor (53).

Discussion – part 2

Results described since the completion of the studies presented in Chapters 1-3 and discussed in part one above generally confirm our original observations. The interpretation of our data, however, is subject to reevaluation based on a number of emerging results.

A repeating pentameric Heat-Shock Element (HSE)

Experiments by a number of groups (54-) seem to convincingly redefine the HSE as a series of adjacent pentameric repeats (5'-nGAAn-3') that can be arranged to consist of integral numbers of the HSE almost always found in tandem and in alternating orientations. For convenience of this discussion the minimal HSE unit will be portrayed as either (→) or (←) for the forward 5bp element (5'-nGAAn-3') or the reverse complementary element (5'-nTTCn-3') respectively. Also an apparently "minimal" decameric HSE consisting of an inverted repeat of this sequence element; -nGAAnnTTCn- will be referred to as a head-to-head repeat (→←) and a –nTTCnnGAAn- unit will be referred to as tail-to-tail orientation (←→).

The critical experiments that led to this reassessment of the HSE were conducted in the Lis laboratory and published in 1989 and 1991 (54,55). The former study convincingly demonstrated that under saturating conditions the *Drosophila* HSTF could bind to either head-to-head or tail-to-tail HSE geometries with surprisingly similar affinities. It was also shown that the HSTF could indeed polymerize in a

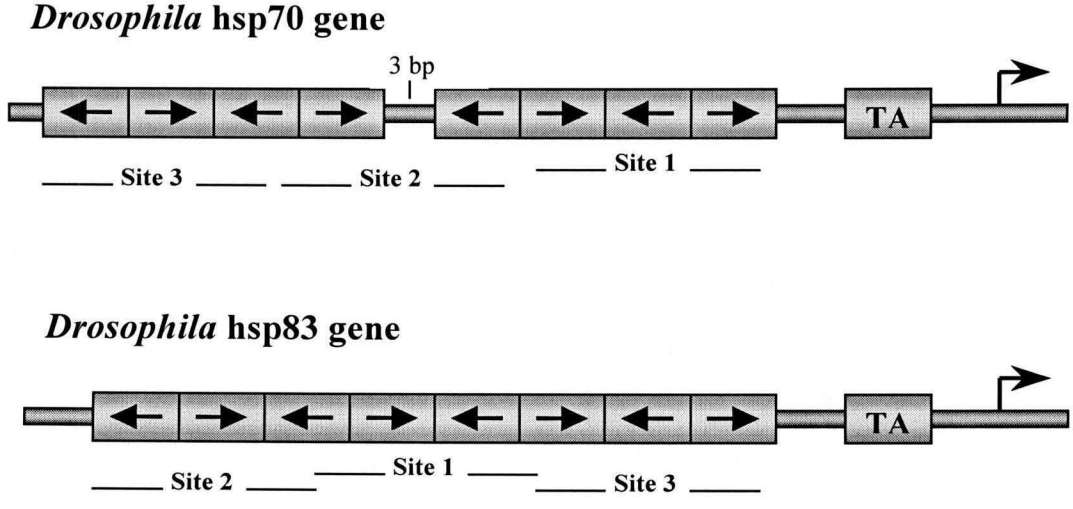
sequence dependent manner and occupy templates harboring repeated pentameric HSE elements spanning from only 10 bp (2 HSE's) up to 45 bp (9 HSE's) exhibiting 5 bp incrementally increasing DNaseI footprint boundaries. Perhaps more interesting, however, was the perceived incongruence of the apparent molecular weights of these HSTF-HSE complexes, where it was observed that a change in gel migration actually seemed to cycle in a 15 bp dependent manner and not the expected gradual 5 bp periodicity of the individual HSE containing test templates. This study went on to discuss a possible revised interpretation of our data employing the notion of the pentameric HSE and a newly postulated trimeric HSTF. This will be explored in detail below. Subsequent studies from the same laboratory confirmed this pentameric unit element of the HSE and also reinforced earlier cooperativity observations (29). Interestingly the thermal dependence of the HSTF-HSE interactions measured at numerous promoters suggest that such interactions may be evolutionarily tailored to exhibit a wide range of transcriptional levels (56).

Apparently, the minimal monomer HSTF binding element can be as simple as a pentameric (nGAAn) or even (nGnnn) when immediately adjacent to a strong HSE. Additionally the minimal HSE binding site appears to consist of two of these inverted pentameric HSE repeats (47,54). This postulate has forced many groups to reevaluate the stoicheometry, orientation and basic nature of the HSTF-HSE interaction. Clearly these putative pentameric HSE units harboring at most three consensus base pairs would not provide the specificity required for the specific gene induction observed at heat shock promoters; a random genomic base pair distribution would result in a

GAA or TTC containing 5bp HSE every 64 bp (4^3). It is thus generally assumed that the spatially controlled, alternately inverted, and repeated nature of these elements lends both binding affinity and specificity to this elaborate protein-DNA interaction.

These convincing arguments for HSE redefinition strongly suggested that the contact point data presented above should be reevaluated. Under this revised HSE concept one would postulate that the *Drosophila* hsp70 site 1 and 2 region could be depicted as a centrally disrupted 43 bp ((←→←→)-3bp-(←→←→)) schematic. Likewise the *Drosophila* hsp83 promoter region could be represented as a more common non-interrupted and adjacent 40 bp HSE arrangement of (←→←→←→←→). See figure D-2 below and note that the approximate boundaries of the three sequentially occupied HSTF binding sites for the three complexes formed on each promoter are drawn below the figures. The TATA-boxes (TA) and start points of transcription (standing arrows) are also shown.

Figure D-2; Redefined Hsp promoter elements



As with most HSE-containing promoters, the hsp70 and hsp83 HSE regions are not always perfectly repeating nGAAn or nTTCn elements but clearly those closer to the optimal sequence are occupied preferentially at low HSTF concentrations. These imperfections are found at most heat-shock gene promoters, and indeed it appears from numerous studies that the only invariant position of this pentameric element is the guanosine-2 of the forward unit (n<u>G</u>AAn or schematically; →) as is found at the distal end of hsp70 site 3.

The active HSTF may be trimeric

The HSTF has been dissected into a number of functional domains (reviewed in 6). The DNA binding domain is located at the N-terminus followed by an extended heptad repeat domain thought to play the critical role in HSTF trimerization described below. The transcriptional activation domain is located at the C-terminus and adjacent to another smaller heptad repeat region shown to be important in repression of the active HSTF. It has been postulated that this domain forms an intramolecular association with the trimerization domain thus precluding HSTF multimerization and the subsequent stable DNA binding of this factor (14,57,58). It has also been shown that a heat-inducible nuclear localization sequence located between these putative interacting heptad repeat regions is revealed and enabled in the activation process (59). Thus it appears that this protein has evolved to remain in a dormant state and

then respond rapidly to cellular stress challenges by converting from an inactive monomer to the active trimeric HSTF.

The biochemical evidence to support a trimeric model of HSTF can generally be classified into two groups; gel-shift protein-DNA complex studies, and solution crosslinking studies. As described above, it was originally reported that increasing the HSE by five bp increments only resulted in significantly larger HSTF-HSE complexes when this expansion reached 15 bp, or 3 HSE unit elements (54) and thus the trimeric HSTF model was postulated. Subsequent studies by Sorger and Nelson (60) utilized engineered short (S) and long (L) forms of the HSTF, both fully competent in DNA binding, and demonstrated that in mixing experiments four discrete HSTF-HSE complexes could be resolved in an approximate 1:3:3:1 ratio. This suggested that the HSTF was trimeric as these complexes were identified as L^3:L^2S:LS^2:S^3. Unfortunately these experiments employed a human 15 bp HSE that contained three HSE unit elements, and therefore this study was biased to detect only trimers. A similar study came later where Bonner and colleagues observed the same "trimeric" distribution with a 20 bp HSE containing four HSE unit elements (61) thus supporting the trimer theory. Unfortunately, these complexes did not reveal the expected ratios. Clearly it would also be desirable to exploit this novel HSTF mixing approach to investigate templates harboring a minimal dimeric 10 bp HSE and HSE regions of greater than 20 bp.

Numerous groups have performed chemical crosslinking studies of free HSTF in solution and also in the presence of HSE containing probes (55,60,62,63) and the resulting reaction products were analyzed on sizing gels or columns. Significantly, even at high crosslinking reagent concentrations, the predominant species observed are consistent with the trimeric model of HSTF binding. Also detected by a number of groups are what appear to be crosslinked HSTF dimers (60,62,63). It is important to note that these studies all employ bifunctional crosslinking agents that would clearly bias the results in this direction at lower protein and reagent concentrations (i.e. higher order HSTF multimers would require more than one crosslinking event).

Clearly a trimeric sequence-specific binding protein that recognizes a rotationally symmetric template in either of two orientations, (→←) or (←→), presents a number of structural questions. It has been postulated that a highly flexible linker or swivel exists between the DNA binding domains and the multimerization domains (64) thus allowing the HSTF to specifically contact the DNA and still present trimerization surfaces in the proper spatial context to facilitate this self-association.

Data reinterpretation

In Chapters 1 through 3 and also in "Discussion-part 1" above we postulated that at each of the *Drosophila* promoters studied, hsp70 and hsp83, a dimeric HSTF binds first at a high affinity traditionally defined HSE site (site 1) with subsequent monomeric HSTF binding to the site 2 and then site 3 "half-sites" to form a stable

complex consisting of four HSTF monomers bound to these HSE elements (see figure 7 of Chapter 3). Using the updated model of the HSE and the trimeric model of the HSTF we have reinterpreted our data for complexes A, B* and B formed at the *Drosophila* hsp70 promoter and complexes A, B and C formed at the *Drosophila* hsp83 promoter (figures D-3 and D-4 below). Arrows denote the presence and orientation of the pentameric HSE unit elements and the TA and standing arrow represent the TATA-box and the start point of transcription, respectively. HSTF trimers are depicted as three DNA binding domains (ovals) nucleated by the remainder of the HSTF drawn as the text box. Open circles denote the conserved G-2 consensus sequence contacts observed in each HSTF-DNA complex. These figures are not intended to mimic any specific topological HSTF arrangements including the orientation of the DNA binding domains within the major groove.

Clearly three trimers would be required to cover this region as the number of HSE unit elements in each promoter is eight, resulting in the three observed complexes. We are also postulating that at one of these binding sites, one portion of the trimeric HSTF would not contact DNA. Clearly, a number of alternate descriptions could also be accurate, possibly including monomeric or dimeric HSTF binding to fill any remaining "partial" sites of a multi-unit HSE element (i.e. those containing only one or two pentameric units).

Figure D-3; HSTF-Hsp70 complexes

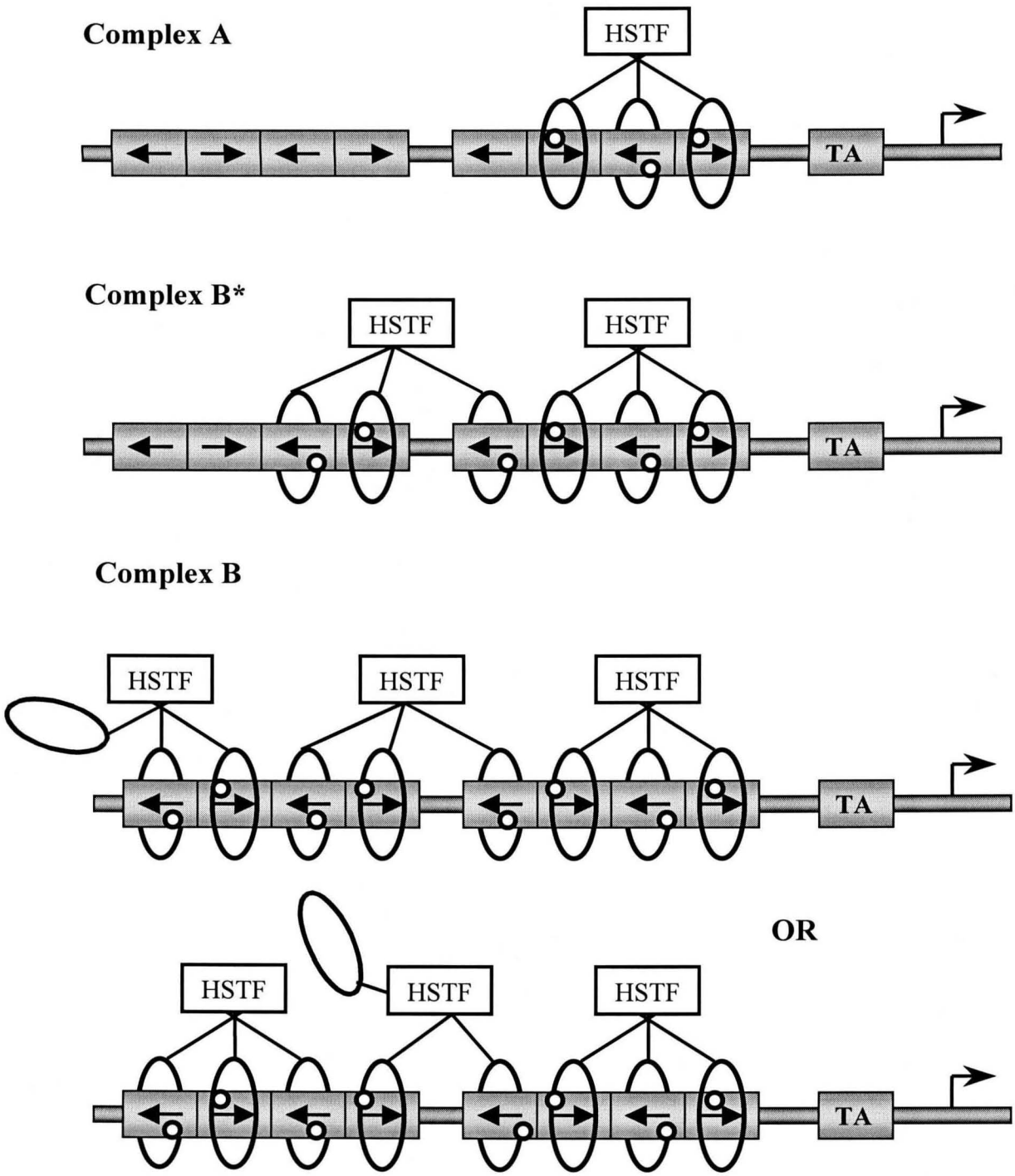
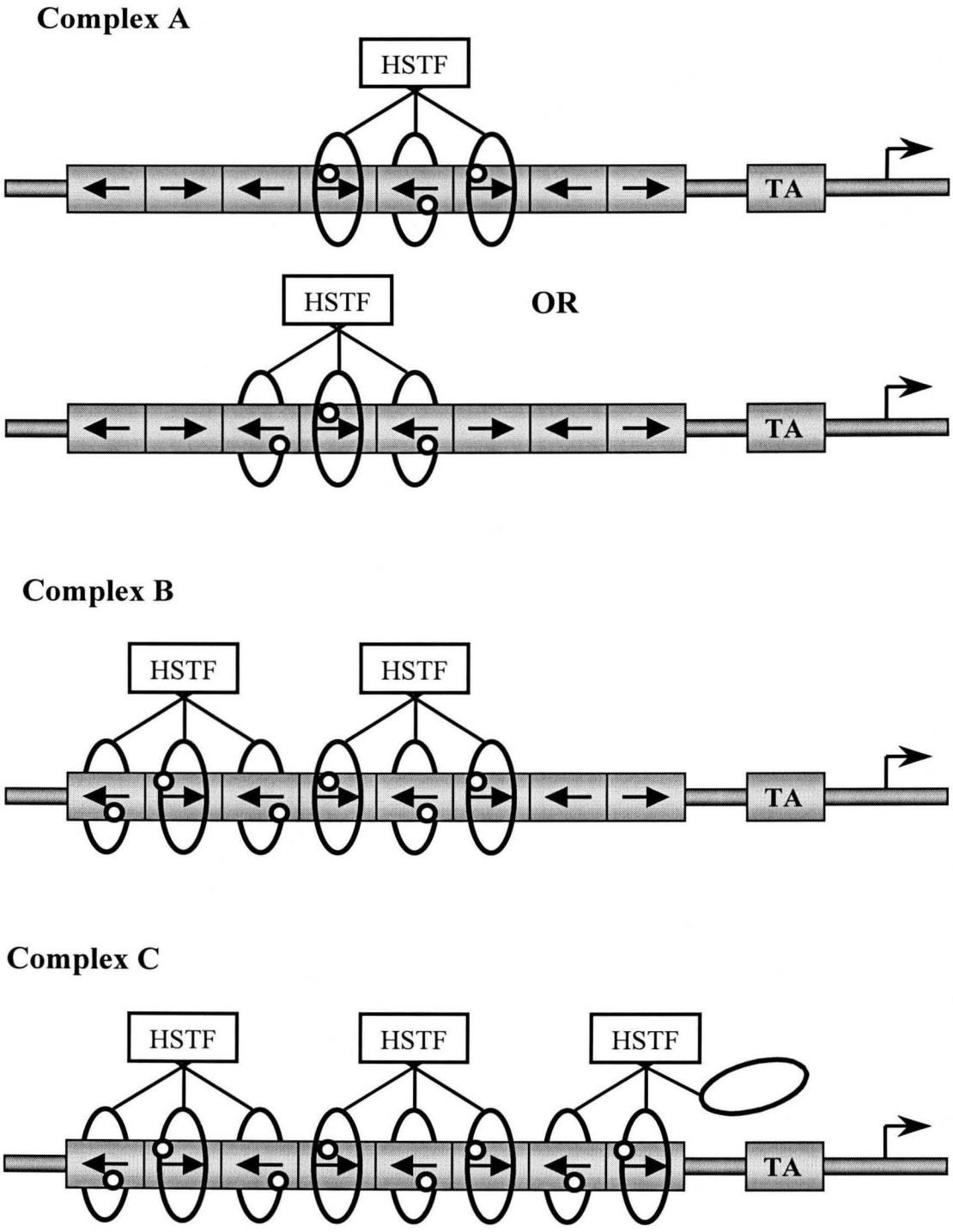


Figure D-4; HSTF-Hsp83 Complexes



The studies presented above provide evidence for an active trimeric HSTF and seem to be well accepted. There are, however, a significant number of observations including our own that suggest the possibility that alternative HSTF binding stoicheometries could indeed exist.

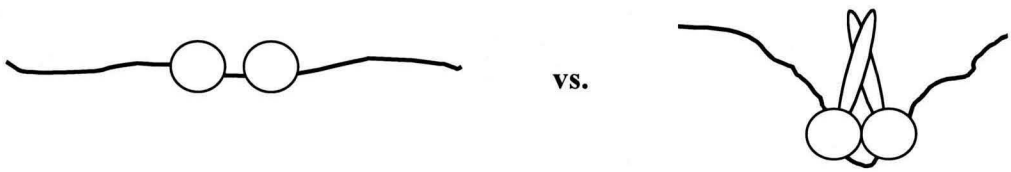
Co-crystallographic studies clearly demonstrate that a that two HSTF monomers are sufficient to occupy a dimeric 10 bp HSE element, albeit at a much lower affinity (47). As described above, numerous groups have also reported gel-shift studies of protein-DNA complexes that seem to involve not only trimers but also dimers of bound HSTF. Therefore the complexes depicted in figure D-3 and D-4 should be viewed as purely hypothetical in nature and simply represent an attempt to incorporate the revised notions of the HSE and HSTF into a simple diagram.

Co-crystallographic studies – Does HSTF induce DNA bending?

Nelson and colleagues (46,47) have presented several crystallographic studies with the DNA binding domains of yeast HSTF and a tail-to-tail nTTCnnGAAn (←→) HSE oligonucleotide. Importantly, and in contrast to our previously described results of Chapter 2, no significant DNA bending was observed in this HSTF-HSE co-crystal structure. This observation led Nelson to conclude that we were in error and that the HSTF does not induce DNA bending as we had postulated in 1986.

It is significant to note that the structures analyzed by these groups did not employ full-length HSTF and thus may have precluded the observation of higher order DNA structures potentially formed only when mediated by the self-association of the HSTF

via it's multimerization domains. For example, one could easily imagine HSTF "clothespins heads" (HSTF DNA binding domains) clipped to a line (DNA) several inches apart without producing significant distortion of this clothesline template. Upon tightly fastening these clothespins together through their "tails" (HSTF multimerization domains) one might however expect to bend this clothesline (see cartoon below).



Notably the 43 bp *Drosophila* hsp70 HSE region utilized in the DNA bending studies consisted of alternating head-to-head and tail-to-tail HSE's (see figure D-2) and it is possible that this arrangement promoted or allowed DNA bending while the smaller 10 bp tail-to-tail oligonucleotide HSE alignment utilized in the crystallographic analyses did not. Clearly the HSTF multimerization interfaces, either (←—→) or (→←) binding orientations, would be expected to be quite different in these two geometries. It is important to note that in a head-to-head orientation the critical consensus guanosine-2 residues would be separated by 6 bp while in the alternate tail-to-tail orientation these residues are separated by only 2 bp. Additionally, the packing of these small oligonucleotide templates into unit cells could favor a more linear DNA arrangement through energetically favorable hydrophobic base stacking interactions of adjoining templates.

Another possible explanation for this apparent discrepency would involve species-specific HSTF-HSE interactions. These co-crystals utilized yeast, *Kluyveromyces lactis*, HSTF DNA binding domains which could in theory interact differently the *Drosophila* HSTF. However, owing to the conservation of HSE binding observed from yeast through man, one might reasonably conclude that this explanation is unlikely.

Clearly a final answer to this question will require co-crystallographic or additional biophysical studies employing the entire HSTF bound to head-to-head and tail-to-tail HSE templates. It is likely that the generation of a higher order full-length HSTF-HSE co-crystal has been attempted without success. Such a complex would be significantly larger and may be correspondingly far more difficult to crystallize and analyze. The fact that such data has not been reported to date suggests that perhaps such an elaborate structure might also be asymmetrical in nature. Possibly a solution NMR structure of a small ^{15}N-labeled HSE template in the presence of either full-length HSTF or the minimal DNA binding domain of the HSTF could settle this issue.

Conclusions

The past 20 years have witnessed a remarkable revolution in our understanding of the mechanisms of gene regulation. Clearly, much remains unknown but some general principles of transcriptional control have emerged: 1) tertiary structure of the protein-DNA complexes at the promoter are clearly important, 2) the multiplicity and

cooperativity of transcription factor interactions are essential, and 3) the localization, phosphorylation, degradation, repression and multimerization of transcriptional regulators all contribute. A detailed structure-based understanding of these and other processes is rapidly emerging.

Aggressive drug development programs are currently underway to take advantage of the knowledge gained by such basic biochemical research. Confirmed molecular interactions can be exploited by numerous biophysical and cell-based screening regimens engineered to recover small molecule disruptors of these interactions or conversely activating "dimerizing ligands" that would promote them. Additionally, recent advancements in peptide and chemical libraries, molecular modeling and structural database searching programs have been exploited in the rational drug design of protein-protein binding inhibitors. There is no theoretical reason that this technology could not extend to specific protein-DNA binding disruptors, and thus generate an entirely new class of gene expression modulating therapeutics.

As described above, we have developed two technologies that significantly aided in the dissection of these protein-DNA and protein-protein interactions and employed these to study specific aspects of the heat-shock system. We hope that in some small way these biological and technical contributions have advanced our basic understanding of eucaryotic gene expression and may someday be fully realized at the clinical level.

Appendix

Characterization of an RNA polymerase activity from HeLa cell mitochondria, which initiates transcription at the heavy strand rRNA promoter and the light strand promoter in human mitochondrial DNA.

David J. Shuey and Giuseppe Attardi

Journal of Biological Chemistry (1985) February 10; volume 260, number 3, pages 1952-1958.

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 260, No. 3, Issue of February 10, pp. 1952-1958, 1985
Printed in U.S.A.

Characterization of an RNA Polymerase Activity from HeLa Cell Mitochondria, Which Initiates Transcription at the Heavy Strand rRNA Promoter and the Light Strand Promoter in Human Mitochondrial DNA*

(Received for publication, June 11, 1984)

David J. Shuey and Giuseppe Attardi

From the Division of Biology, California Institute of Technology, Pasadena, California 91125

An RNA polymerase activity capable of initiating transcription at both the heavy strand rRNA promoter and the light strand promoter of human mitochondrial DNA has been partially purified from HeLa cell mitochondria and characterized in its requirements and products. The ratio of the two transcription initiating activities varied considerably from preparation to preparation. The human mtRNA polymerase partially purified by DEAE-cellulose and heparin-agarose chromatography exhibits a great sensitivity to ionic strength and to Mn^{2+}, characteristics which clearly differentiate this enzyme from bacterial and eukaryotic nuclear RNA polymerases, and in contrast resemble the behavior of the yeast mtRNA polymerase. The human mtRNA polymerase exhibits a requirement for ATP which is 15- to 20-fold higher than that for the other NTPs, a low optimum template DNA concentration, and a marked susceptibility to inhibition by non-mitochondrial DNA.

Recent work has provided an insight into the initiation of transcription in human mitochondrial DNA. Thus, S_1 mapping experiments utilizing *in vitro* capped RNA molecules or nascent chains isolated from purified transcription complexes have led to the identification of two initiation sites for *in vivo* heavy (H)-strand transcription, one located at 20 to 25 base pairs upstream of the $tRNA^{Phe}$ gene, and the other, near the 5′-end of the 12 S rRNA gene (1, 2). Furthermore, a mapping and kinetic analysis of rDNA transcripts labeled *in vivo* or in isolated organelles has established a correlation between these two initiation sites and two distinct transcription events: of these, one, starting at the upstream site, appears to terminate at the 3′-end of the 16 S rRNA gene and to be responsible for the synthesis of the bulk of the two rRNA species; the other transcription event, starting at the downstream initiation site, apparently does not stop at the 3′-end of the 16 S rRNA gene, but extends to cover almost the entire length of the H-strand (3, 4). This polycistronic molecule is destined to yield by processing all the mRNAs and most of the tRNAs encoded in the H-strand (5). Other mapping experiments utilizing *in vivo* synthesized and *in vitro* capped primary transcripts have identified an initiation site for light (L)-strand transcription near the 5′-end of the 7 S RNA coding sequence (1, 6).

* These investigations were supported by National Institutes of Health Grants GM-11726 and T32 GM076165. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The understanding of the factors controlling the initiation of transcription from the two H-strand start sites and from the L-strand start site and the potential termination of H-strand transcription at the 3′-end of the 16 S rRNA gene depends to a great extent on the availability of *in vitro* soluble transcription systems utilizing an exogenous mtDNA template. Two such systems have been recently described, one derived from rat liver mitochondria (7) and the other, from KB cell mitochondria (8). In both cases, L-strand transcription was observed on a mtDNA fragment which contained a portion of the D-loop and the upstream noncoding region including the 5′-end proximal segment of the 12 S rRNA gene. In the KB cell system, an initiation site for L-strand transcription was localized exactly in correspondence of the 5′-end of 7 S RNA, like the *in vivo* initiation site (8).

In the present work, an RNA polymerase activity capable of initiating transcription at both the rRNA-specific H-strand promoter and at the L-strand promoter has been partially purified from HeLa cell mitochondria. This activity has been characterized in its detailed requirements and some of its properties. The available evidence indicates that the same enzyme is probably involved in the reading of the two transcription units. However, the much lower efficiency of *in vitro* initiation at the rRNA-specific H-strand promoter as compared to that at the L-strand promoter and its variability from preparation to preparation suggest that some other factor(s), not operating in the *in vitro* system, modulates the *in vivo* rates of initiation of rRNA synthesis and L-strand transcription. While this manuscript was in preparation, evidence for H-strand transcription initiation by the KB cell polymerase activity (9) and a similar activity isolated from HeLa cells (10) has been reported.

MATERIALS AND METHODS

Isolation of Mitochondria

Generally, 3 to 15 liters of HeLa cell suspension cultures in modified Eagle's medium supplemented with 5% calf serum (11) were grown to late exponential phase. Following homogenization in hypotonic medium, mitochondria were isolated by differential centrifugation (12). The final mitochondrial pellet was resuspended in one-half the volume of the original whole cell pellet of 25 mM Tris, pH 8.1 (25 °C), 5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM dithiothreitol, 0.2 mM phenylmethylsulfonyl fluoride, 20% glycerol (buffer A), and either lysed immediately or frozen in liquid nitrogen and stored at −80 °C.

Isolation of RNA Polymerase Activity

Solubilization of RNA Polymerase Activity—A mitochondrial suspension (about 20 mg/ml protein in buffer A) was homogenized five times with a motor-driven Teflon pestle. Nonidet P-40 and KCl were added to 0.5% and 0.5 M, respectively, and the mixture was vortexed and allowed to stand on ice 5 to 10 min. The homogenization was

repeated 10 times, yielding finally the "mitochondrial lysate." This was either used immediately, frozen in liquid nitrogen and stored at −80 °C, or treated as described below.

DEAE-cellulose Chromatography—The mitochondrial lysate (generally 10 to 20 ml) was spun at 33 krpm (~100,000 × g_{av}) for 60 min in an SW50.1 rotor. The S-100 was carefully removed and dialyzed against 1 liter of buffer A supplemented with 0.2% Nonidet P-40 and 0.1 M KCl until the conductivity of a 1:200 dilution was approximately 60 μmhos × cm^{-1} (~2 h). The dialyzed material was applied, at two column volumes per h, onto a 20-ml column of DEAE-cellulose (DE52, Whatman), prepared as suggested by the manufacturer and pre-equilibrated with buffer A plus 0.2% Nonidet P-40 and 0.1 M KCl. Protein in the flow-through was monitored by the method of Bradford (13) and the fractions containing protein were pooled.

Heparin-Agarose Chromatography—The DE52 flow-through was applied directly, at five column volumes per h, onto a 5-ml heparin-agarose column prepared by the method of Davison *et al.* (14) and preequilibrated with buffer A plus 0.2% Nonidet P-40 and 0.1 M KCl. The column was washed for 1 h and bound components were eluted with 50 ml of a linear KCl gradient (0.1 M to 0.6 M) in buffer A. The fractions were assayed for transcription promoting activity as described below and those containing the peak activity were pooled, frozen in small aliquots in liquid nitrogen, and stored at −80 °C. No loss in activity was detected for at least 4 months. This fraction will be referred to as HA mtRNA polymerase.

Cloned Mitochondrial DNA

The template DNA used in the *in vitro* transcription experiments in all cases consisted of recombinant plasmids harboring human mtDNA inserts or of gel-purified restriction fragments (in the run-off assays). The plasmid pBHK2 (15) (a gift of B. Greenberg, California Biotechnology, Inc., Mountain View, CA) contains the *Kpn*I fragment 2 of human mtDNA (spanning positions 15052–2577 (16)) cloned into pAD23 (pBR322 with *Hin*dIII fragment D of adenovirus-2 DNA inserted into the *Hin*dIII site). The plasmids pmt.H8 and pmt.M9 were constructed in this laboratory by Gaines and King.[1] pmt.H8 contains the *Hpa*II fragment 8 (positions 104–931) cloned into the *Acc*I site of pUC-9 (17). pmt.M9 contains the *Mbo*I fragment 9 (positions 1–739) cloned into the *Bam*HI site of pUC-9.

The M13 vectors mp8.M9 and mp9.M9, carrying the H-strand and L-strand, respectively, of the *Mbo*I-9 fragment, to be used in the hybridization tests, were constructed in this laboratory by M. King by subcloning this fragment directionally into the replicative form of M13 derivatives mp8 and mp9 (18), respectively. Single-stranded phage DNA was prepared as described (18).

Transcription Assays

The standard transcription buffer employed consisted of 10 mM Tris, pH 8.1 (25 °C), 7.5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM dithiothreitol, 10% glycerol, 50 μg/ml bovine serum albumin, 0.5 mM ATP, 0.1 mM CTP and GTP, and 0.01 to 0.02 mM [α-^{32}P]UTP (5 to 10 Ci/mmol (ICN)). In the assays for the measurement of overall RNA polymerase activity, pmt.M9 (closed circular) was used at 3 to 5 μg/ml as template DNA and the reaction mixtures (50 to 100 μl) were incubated at 30 °C for 30 to 45 min. The enzyme preparation was generally added to 5 to 10% of the reaction volume.

The RNA polymerase activity was quantitated by measuring the conversion of [α-^{32}P]UTP into an acid-insoluble form. For this purpose, the reactions were terminated by cooling the mixtures on ice and by adding to each one 1.0 ml of 1 N HCl, 0.1 M sodium pyrophosphate, and 2.5 μg of yeast tRNA. After 30 min, the mixtures were filtered through presoaked 25-mm Whatman GF/C filters; these were then rinsed with 15 ml of cold 1 N HCl, 0.1 M sodium pyrophosphate, washed with 5 ml of ethanol, dried, and counted in a toluene-based scintillation mixture.

Synthesis, Purification, and Analysis of Run-off Transcripts

Transcription conditions were identical with those described above, except that the template consisted of a gel-purified insert from one of the above described recombinant plasmids. The reactions were carried out at 20 °C for 60 min and terminated by adding an equal volume of 0.5% SDS,[2] 10 mM Tris, pH 8.1 (25 °C), 0.4 M Na^+ acetate, 20 mM EDTA. The reaction mixtures were then phenol-extracted and the nucleic acids were ethanol-precipitated two or three times to remove unincorporated nucleoside triphosphates. The samples were electrophoresed on a 1.4 or 1.8% agarose, 5 mM CH_3HgOH gel or on a 5 or 8% polyacrylamide, 7 M urea gel, and the resolved products were visualized by autoradiography.

For nuclease S_1 mapping, preparative run-off reactions were performed using 1 μg of *Kpn*I-2 and *Hpa*II-8 DNA templates, 10 μl of mitochondrial lysate, and 0.01 mM [α-^{32}P]UTP (25 Ci/mmol) in a volume of 200 μl, and incubating the mixtures at 25 °C for 60 min. The products were visualized in the wet gel by autoradiography, excised, and eluted by crushing the gel slice and incubating it for ~12 h at 37 °C in 0.4 ml of 0.1% SDS, 10 mM Tris, pH 8 (25 °C), 1 mM EDTA. This material was phenol-extracted, ether-extracted, and then ethanol-precipitated twice. Nuclease S_1 analysis was performed essentially as previously described (4). Briefly, *in vitro* synthesized and gel-purified run-off transcripts were hybridized to an excess (0.5 μg) of unlabeled strand-specific M13 clones harboring *Mbo*I-9 inserts (see above) under high formamide conditions favoring RNA-DNA association (19). The samples were digested with 250 units of S_1 nuclease (Sigma) at 41 °C for 30 min and the reactions were terminated by addition of SDS to 1%, followed by two ethanol precipitations. The products were then visualized by autoradiography after electrophoresis under native conditions through an 8% polyacrylamide gel.

RESULTS

Solubilization of mtRNA Polymerase Activity—Mitochondria isolated from HeLa cells grown to late exponential phase yielded very active lysates. The solubilization of the mtRNA polymerase activity was performed essentially as described by Levens *et al.* (20) for the yeast enzyme, utilizing a medium salt/Nonidet P-40 lysis followed by homogenization. Immediately following lysis, the extract was either assayed, frozen at −80 °C, or dialyzed for further purification (see below). At −80 °C, the activity was stable for at least 6 months.

Transcription of Exogenous mtDNA Templates in a Mitochondrial Lysate—Using as templates cloned mtDNA fragments spanning the region upstream of the D-loop (in the direction of H-strand synthesis), which is known to contain the *in vivo* initiation sites for H- and L-strand transcription (1, 2), accurate initiation at both the L-strand and H-strand promoters and efficient elongation were observed in a crude mitochondrial lysate. A major transcript and a minor transcript were obtained with each of three templates used, the *Mbo*I-9, *Hpa*II-8, and *Kpn*I-2 fragments (Fig. 1*A*). These transcripts had the sizes expected for L-strand run-off transcripts initiating at the 5′-end of 7 S RNA (the major ones) and for H-strand run-off transcripts initiating ~25 base pairs upstream of the $tRNA^{Phe}$ gene (the minor ones) (Fig. 1*B*). These sites correspond to the initiation sites previously mapped for RNA synthesized *in vivo* (1, 2) and *in vitro* (8–10). No transcripts shorter by ~95 nucleotides than the H-strand run-off transcripts and, therefore, possibly initiating near the 5′-end of the 12 S rRNA gene, were detected. Surprisingly, very little nuclease activity was present in these preparations, thus allowing the synthesis of RNA greater than 2000 nucleotides in length (Fig. 1*A*). The system was completely dependent on the addition of exogenous template and specific for promoter containing mtDNA, exhibiting negligible nonspecifically initiated transcription.

In the experiment illustrated in Fig. 1, when *Kpn*I-2 was used as a template, besides a 2050-nucleotide run-off transcript spanning the sequences corresponding to the $tRNA^{Phe}$ RNA, the 12 S RNA, the $tRNA^{Val}$ RNA, and the 5′-end proximal half of the 16 S RNA (Fig. 1*B*), two transcripts (indicated by *arrows*) were observed which corresponded perfectly in migration to the RNA species 12 S and 12 S* (Fig. 1*A*): the latter species represents a precursor of 12 S RNA still carrying the $tRNA^{Phe}$ and leader sequences, which accu-

[1] G. Gaines and M. King, unpublished data.

[2] The abbreviation used is: SDS, sodium dodecyl sulfate.

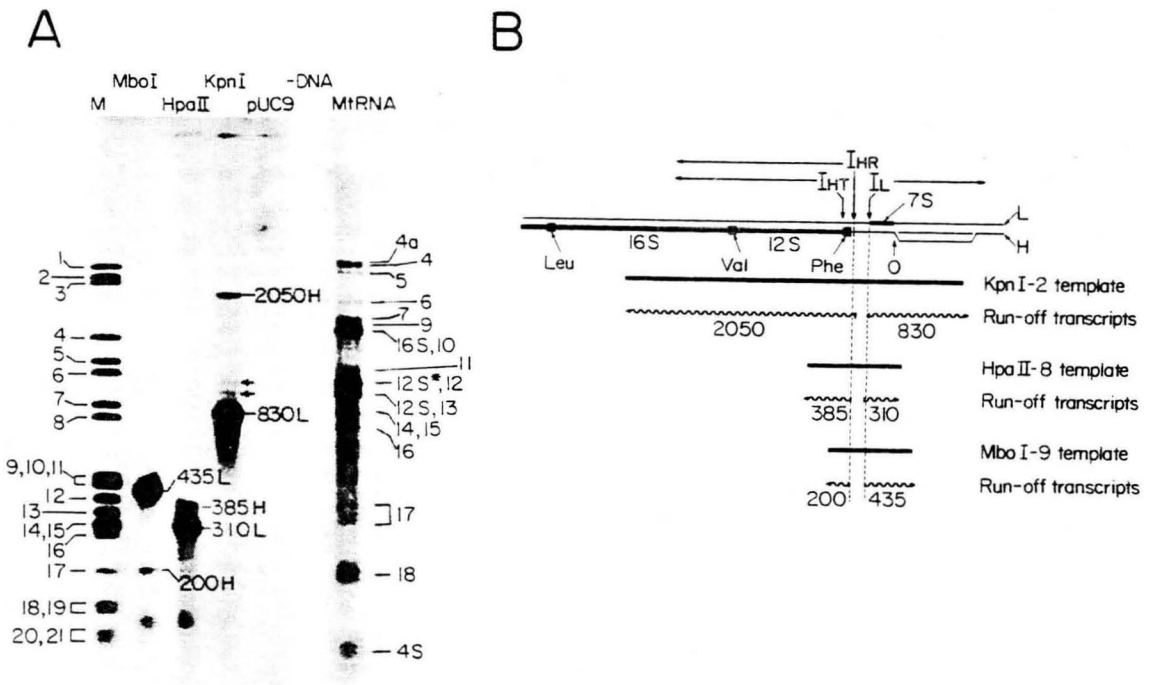


FIG. 1. **Run-off transcripts synthesized by a HeLa cell mitochondrial lysate programmed by added mtDNA templates.** *Panel A*, run-off transcripts were synthesized from 5 μg/ml of the indicated templates in a whole lysate transcription system, phenol-extracted, and analyzed by electrophoresis through a 1.8% agarose-CH_3HgOH slab gel. (The *Kpn*I-2 and *Hpa*II-8 fragments were excised from pBHK2 with the corresponding enzymes, the *Mbo*I-9 fragment was isolated as a slightly larger *Hin*dIII-*Eco*RI fragment from a *Hin*dIII-*Eco*RI digest of pmt.M9.) pUC-9 DNA (closed circular) was used as a nonspecific template. *M*, end-labeled *Hpa*II digest of HeLa mtDNA; *mtRNA*, total RNA synthesized in isolated mitochondria (4) (provided by George Gaines of this laboratory). Presumptive processed RNA species 12 S and 12 S* are indicated by *arrows*. *Panel B*, the restriction fragments used to generate the run-off transcripts are aligned with the map of the origin region of HeLa cell mtDNA. The sizes of the run-off transcripts were estimated from their electrophoretic mobilities on the basis of a size *versus* mobility plot of *in vitro* synthesized mtRNA species.

mulates during incubation of isolated organelles (4) and which has also been detected *in vivo* (2). These observations strongly suggest that some processing of the H-strand transcripts has occurred *in vitro*. The absence of the putative 12 S RNA and 12 S* RNA from the products of the other reactions supports the interpretation that these two labeled species derive indeed from processing of the *Kpn*I-2 programmed transcripts, and that they do not result merely from end-labeling of endogenous 12 S and 12 S* RNAs.

When intact pmt.M9 DNA (closed circular) was used as a template with a crude mitochondrial lysate, a heterogeneous population of large RNA molecules, up to 4 to 5 kilobase pairs (mean size ~1.5 kilobase pairs), was synthesized (not shown). Since pUC9 does not appear to support *in vitro* transcription (Fig. 1*A*), it seems reasonable to interpret the large transcripts synthesized from pmt.M9 as originating within the *Mbo*I-9 insert and terminating or pausing at various points in the plasmid.

S_1 Mapping of Transcripts—To confirm the localization of the initiation sites and the strandedness of the *in vitro* synthesized transcripts, nuclease S_1 mapping was performed on these transcripts. The radiolabeled 830-nt RNA and 385-nt RNA synthesized from the *Kpn*I-2 and *Hpa*II-8 templates, respectively, were gel-purified and hybridized with an excess of M13 clones harboring the H- or L-strand of *Mbo*I-9; after digestion with S_1, the RNA:DNA hybrids were separated by electrophoresis in a polyacrylamide gel (Fig. 2). The sizes of the expected protected fragments were estimated from the length of the region of homology between each probe and the corresponding transcript (assumed to initiate at the site determined in the run-off assays). The observed protected fragments were about 20 nt shorter than predicted, a result which can probably be accounted for by the previous observation that RNA:DNA hybrids migrate in polyacrylamide gels faster than the homologous duplex DNA fragments by 3 to 10% (5).

Partial Purification of mtRNA Polymerase—In order to obtain some information on the properties of HeLa cell mtRNA polymerase, the following rapid purification scheme was devised (Table I). Preparative amounts of mitochondrial lysate (200 to 400 mg of protein) were centrifuged at 100,000 × *g* and the supernatant was dialyzed against 0.1 M KCl as described under "Materials and Methods." Enzyme activity was reproducibly lost during this step (Table I). The dialyzed material was then applied onto a DE52 column. At the above salt concentration, the RNA polymerase activity quantitatively flows through the resin, while the nucleic acids and

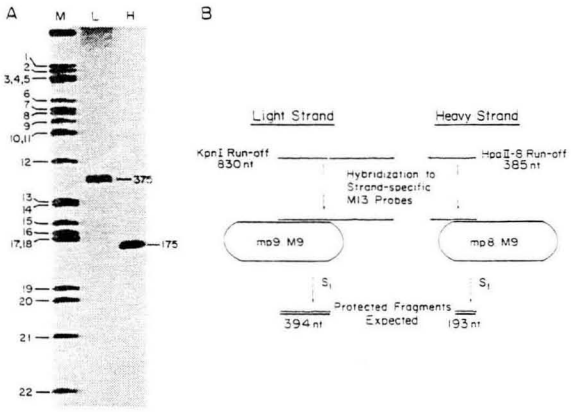


FIG. 2. **Nuclease S_1 mapping of run-off transcripts.** *Panel A*, ^{32}P-labeled run-off transcripts were synthsized in preparative reactions utilizing a mitochondrial lysate and the *Kpn*I-2 or *Hpa*II-8 template. The 830-nucleotide and 385-nucleotide RNAs from the *Kpn*I-2 and *Hpa*II-8 programmed reactions, respectively, were gel-purified and hybridized with 0.5 μg of M13 clones harboring the H- or L-strand of the mtDNA *Mbo*I-9 fragment. After S_1 nuclease digestion of the RNA/DNA hybrids, the S_1-resistant products were analyzed on a 5% polyacrylamide gel. *M*, end-labeled HeLa cell mtDNA cut with *Mbo*I. *Panel B*, S_1 mapping strategy.

~25% of the protein remain bound. An increase in activity was observed after the DEAE-cellulose chromatography (Table I), possibly due to the removal of nucleic acids or nucleases. The flow-through was applied onto a heparin-agarose column and bound components were then eluted with a linear gradient of 0.1 to 0.6 M KCl. Column fractions were assayed for mtRNA polymerase activity and the peak fractions (eluting at 0.25 to 0.35 M KCl) were pooled and frozen at −80 °C. The heparin-agarose chromatography removed most (>90%) of the remaining protein from the 100,000 × *g* supernatant (Table I). This method has the advantage of being rapid and of yielding a 10- to 20-fold purification of the enzyme with 30 to 60% recovery (Table I). The RNA polymerase activity was found to be unstable during dialysis against 0.1 M KCl and, consequently, the enzyme was stored at −80 °C at the salt concentration of elution from the heparin-agarose column: under these conditions, the activity is stable for at least 4 months.

Transcripts synthesized by the partially purified (HA) mtRNA polymerase are indistinguishable from those generated by the lysate (Fig. 3). In particular, the initiation sites of H- and L-strand transcripts do not change during chromatography and the same seems to be true for their ratio.

Properties of Ha mtRNA Polymerase—In order to investigate the requirements of the partially purified enzyme activity, each parameter of the *in vitro* transcription was systematically varied. The activity was quantitated by following the incorporation of [^{32}P]UMP into acid-insoluble material. Since preliminary tests carried out with the mitochondrial lysate failed to show any evidence of nonspecific transcription under different conditions of ionic strength, temperature, divalent cation, and nucleotide concentration, as judged by gel analysis of the products, it seems justifiable to interpret the overall incorporation data as reflecting mainly promoter-specific transcription. All reactions were performed as described under "Materials and Methods," with exceptions noted in the figure legends.

The effect of enzyme concentration on *in vitro* transcription was investigated by testing increasing amounts of HA mtRNA polymerase (from 0.1 to 6 μg) on a constant amount (3 μg/ml) of closed circular pmt.M9. The incorporation of [^{32}P]UMP increased linearly with protein concentrations increasing from 0.12 to 2.4 μg, with a tendency to plateau at higher concentrations (data not shown). In all quantitative assays to be described below, the RNA polymerase activity was monitored in the linear range of the activity curve.

Fig. 4 shows the effects on *in vitro* transcription of varying the ionic strength, the pH, and the concentration of divalent cations and nucleotides. Increasing concentrations of KCl beyond 10 mM progressively inhibit the reaction, with almost complete inhibition being observed at concentrations exceeding 100 mM (Fig. 4*A*). This effect appears to be due to changes in ionic strength, since Na^+ acetate, K^+ acetate, NaCl, and KCl all produce similar curves (data not shown).

$MgCl_2$ is a required cofactor for *in vitro* transcription in the range of 5 to 10 mM (Fig. 4*C*). In contrast, increasing concentrations of $MnCl_2$ above 0.5 mM, in the presence of 5 mM $MgCl_2$, progressively inhibit the reaction (Fig. 4*D*). Also $CaCl_2$ has inhibitory effects on the reaction, although somewhat less pronounced than those of $MnCl_2$ (Fig. 4*D*).

Varying the nucleotide concentration in the reaction mixture revealed substantial differences in the requirements for ATP and the other NTPs for half-maximal synthesis. Approximately 85 μM ATP and, in contrast, only 5 μM UTP were found to be required to reach half-maximal synthesis (Fig. 4, *E* and *F*). CTP and GTP gave values for this parameter similar to those obtained for UTP (<10 μM; curves not shown). A control experiment, involving analysis of the added [γ-^{32}P]ATP by polyethyleneimine thin layer chromatography at different times of the reaction, showed that the ATP remained intact throughout the assay, with no detectable liberation of phosphate or pyrophosphate (data not shown).

A study of the temperature effects revealed that the reaction proceeds at a higher initial rate, but remains linear for a shorter time at the higher temperatures (Fig. 5). To be noticed, in particular, is the rapid plateauing of the incorporation curve at 37 °C. Whether this reflects an intrinsic instability of the enzyme activity at 37 °C, existing also *in vivo*, or rather results from an activation of nucleases or proteases or from other biological events occurring following organelle disruption, has not been determined. An analysis of the transcrip-

TABLE I

Purification of HeLa cell mtRNA polymerase

Fraction	Protein	Activity	Specific activity	Purification	Yield
	mg	*units*[a]	*units*[a]*/mg*	*-fold*	%
Mitochondrial lysate	274	12,100	44	1	100
S-100	231	12,900	56	1.3	107
Dialyzed S-100	218	8,100	37	0.84	67
DEAE-cellulose flow-through	174	12,000	69	1.6	100
Heparin-agarose eluate	10	7,000	700	16	58

[a] 1 unit = 1 pmol of UMP incorporated in 30 min at 30 °C with 5 μg/ml pmt.M9 DNA as a template.

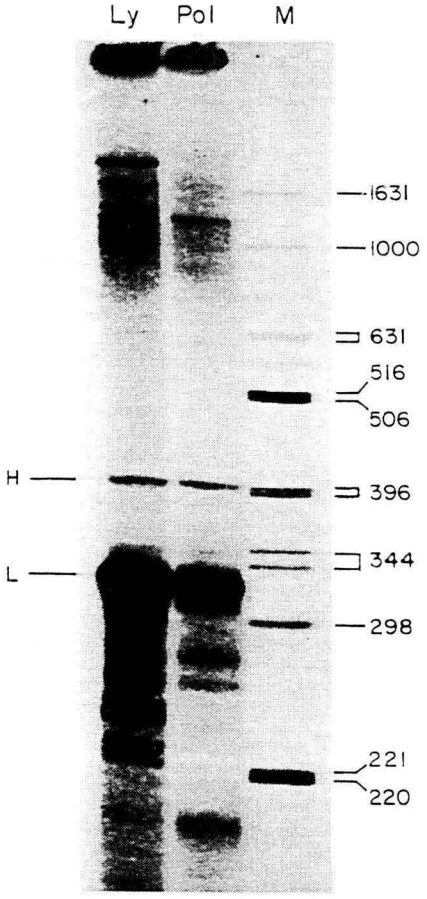


FIG. 3. **Comparison of run-off transcripts synthesized under the direction of whole mitochondrial lysate (*Ly*) and partially purified mtRNA polymerase (*Pol*).** Run-off transcripts were synthesized in 100-μl reaction mixtures containing 3 μl and 6 μl of mitochondrial lysate and HA mtRNA polymerase, respectively, and 0.3 μg of *Hpa*II-8 excised from pmt.H8 with *Eco*RI and *Pst*I (this mode of excision resulted in slightly longer run-off transcripts than seen in the experiment of Fig. 1). *H* and *L* indicate, respectively, the H- and L-strand run-off transcripts. *M*, mixture of end-labeled *Eco*RI, *Pst*I, and *Hin*fI digests of pBR322 DNA.

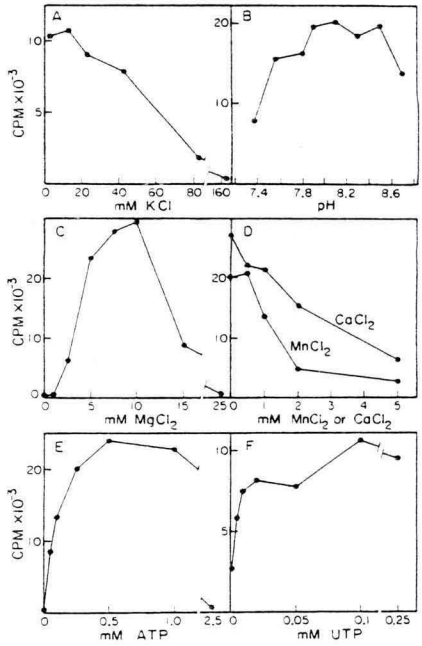


FIG. 4. **Effects of ionic strength, pH, divalent cations, and nucleotides on transcription efficiency with exogenous template by HA mtRNA polymerase.** Transcription reactions were carried out under standard conditions (except as specified below), with the independent variation of the indicated component. *Panel A*, KCl curve: the HA mtRNA polymerase was dialyzed against 40 mM KCl, so that the final salt contribution from this fraction was 2.5 mM. *Panel B*, pH curve: the Tris concentrations were raised from 10 mM to 50 mM; the pH values shown were measured at 30 °C. *Panel C*, $MgCl_2$ curve: EDTA was omitted from the reaction mixture. *Panel D*, $MnCl_2$ and $CaCl_2$ curves: $MgCl_2$ was present at 5 mM in each reaction mixture and EDTA was omitted. *Panel E*, ATP curve. A constant 10 μM [α-^{32}P]UTP (5 Ci/mmol) was used as a labeled precursor. *Panel F*, UTP curve: the UTP specific activity was held constant at 5 Ci/mmol.

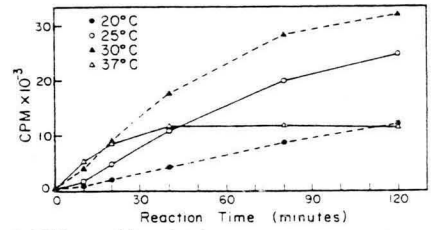


FIG. 5. **Effects of incubation temperature on transcription efficiency with exogenous template by HA mtRNA polymerase.** Transcription reactions were carried out under standard conditions. Four 250-μl reaction mixtures were incubated at the various temperatures, and, at the indicated times, 40-μl portions were removed and precipitated with 1 N HCl, 0.1 M Na^+ pyrophosphate.

tion products revealed a similar temperature dependence for the H-strand and the L-strand transcription. In the present work, all quantitation assays were performed at 30 °C for 30 to 45 min.

In an attempt to study the effect of superhelicity on promoter strength, transcription reactions were carried out in parallel with supercoiled pmt.M9 and with pmt.M9 which had been relaxed by treatment with mouse L cells topoisomerase I (21). No difference was observed in the level of transcription from the two templates. However, ethidium staining of the reaction products obtained with the crude lysate or the HA mtRNA polymerase, after fractionation on an agarose gel, showed that the supercoiled plasmid had been relaxed, giving a pattern very similar to that of the topoisomerase-treated plasmid; in both patterns, several topoisomers up to the fully relaxed form were observed (data not shown). These results indicated that a topoisomerase co-purified with the mtRNA polymerase. Furthermore, there was no evidence for any endonuclease activity. The topoisomerase activity has not been further characterized.

The role of the amount of template in the standard transcription assay was investigated both for the *Kpn*I-2 fragment and the closed circular pmt.M9 plasmid. As shown in Fig. 6*A*, maximum transcription efficiency was reached with 2 μg/ml DNA for the *Kpn*I-2 fragment and 3 μg/ml DNA for the

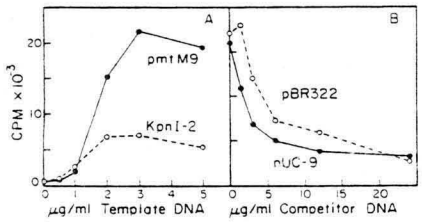


FIG. 6. **Effects of template and nonspecific competitor DNA concentration on transcription efficiency with exogenous template by HA mtRNA polymerase.** Transcription reactions were performed under standard conditions. *Panel A*, the template pmt.M9 was added as supercoiled DNA, but became relaxed during incubation (see "Results"). Given the sizes of pmt.M9 (about 3450 base pairs) and of the *Kpn*I-2 restriction fragment (3013 base pairs), the concentrations of the two templates used were nearly equimolar. *Panel B*, the template DNA (pmt.M9) was held at a constant 3 μg/ml concentration.

pmt.M9 plasmid. Very probably, these maxima reflect the exhaustion of mtRNA polymerase or of some other essential factor(s) in the enzyme preparation added, possibly due to nonspecific binding to DNA (see below). Furthermore, it seems reasonable to assume that the 3-fold higher incorporation obtained with the circular pmt.M9 template as compared to that obtained with an approximately equimolar amount of the *Kpn*I-2 fragment is due to the larger average size of the transcripts in the former case (see above). In view of the failure of pUC-9 DNA to support transcription in this system (Fig. 1A), it seems that one can exclude initiation within the pUC-9 portion of pmt.M9 as a factor contributing to the higher incorporation values obtained with pmt.M9. Thus, initiation appears to be, at least partially, rate-limiting in this system. An interesting observation was that non-promoter-containing DNA, like pUC-9 or pBR322 DNA, can compete, at surprisingly low concentrations, with pmt.M9 for some component essential for transcription, possibly the mtRNA polymerase itself (Fig. 6*B*).

DISCUSSION

The mtRNA polymerase activity analyzed in this work appears to be similar to those isolated by Walberg and Clayton (8) from KB cell mitochondria and by Bogenhagen *et al.* (10) from HeLa cell mitochondria in its capacity of starting RNA synthesis at the *in vivo* initiation site for L-strand transcription (1) and at the upstream initiation site for H-strand transcription identified both *in vivo* (1, 2) and *in vitro* (9, 10). However, the HA mtRNA polymerase isolated here differs from the previously described KB cell and HeLa cell enzyme preparations (8, 10) for exhibiting a much lower level of nonspecific transcription.

In the present work, the H-strand and L-strand transcripts synthesized by the whole mitochondrial lysate on an exogenous mtDNA template were identical in their initiation sites and in their ratio with those synthesized on the same template by the enzyme partially purified through two different chromatographic steps. Although one cannot exclude the possibility that promoter-specific transcription factors or different RNA polymerases co-chromatograph, the above observation suggests the existence of a single RNA polymerase for both transcription units. The similar temperature dependence and *in vitro* stability of the two activities is consistent with this possibility. This is also supported by the identification of very similar putative promoter sequences at the 5′-end of the two transcription units (9, 10). However, a considerable variability from preparation to preparation was observed in the efficiency of H-strand transcription, from an undetectable level to a level one-third to one-fifth of the efficiency of L-strand transcription. On the contrary, *in vivo*, the transcription of the rDNA region and that of the L-strand occur at comparable rates (22), suggesting that some other factor(s) not fully operative in the present *in vitro* system play an essential role *in vivo* in modulating the relative rates of transcription.

The great sensitivity to ionic strength exhibited by the HeLa cell enzyme, and previously reported for the KB cell enzyme (8), clearly differentiates this polymerase from the bacterial or eukaryotic nuclear RNA polymerases (23, 24); this behavior of the human enzyme strongly resembles that of the mtRNA polymerases previously purified from yeast (20) and *Xenopus laevis* (25). Also the strong inhibition by Mn^{2+} of the human mtRNA polymerase distinguishes this enzyme from the nuclear enzymes (24). The yeast mitochondrial enzyme has also been shown to be strongly inhibited by Mn^{2+} (20), whereas the *X. laevis* enzyme has been reported not to be affected by this cation (25).

Another notable feature of the human mtRNA polymerase is the requirement for ATP at a much higher concentration (15–20-fold) than that for the other NTPs. Control experiments have shown that this high ATP requirement is not due to its rapid degradation in the *in vitro* system. One cannot distinguish with the *in vitro* assay utilized here between the ATP requirement for initiation and the requirement for elongation. However, since the 5′-nucleotide of the L-strand transcripts and, probably, that of most of the H-strand transcripts is A (9, 10), it seems possible that the high ATP requirement for *in vitro* transcription of human mtDNA reflects a high K_m for the process of initiation. Evidence that the apparent K_m for initiation is about 10-fold that for polymerization has been previously presented for *in vitro* transcription of various DNAs by *Escherichia coli* RNA polymerase (26). It is interesting that, on the basis of the intracellular distribution of ATP (27) and of the mitochondrial volume (28), the mitochondrial ATP concentration of HeLa cells can be estimated to be ~0.8 mM, *i.e.* close to the optimum ATP concentration for *in vitro* transcription observed here. An appealing possibility is that the mtRNA polymerase has evolved so as to function in a high ATP environment.

An estimation of the rate of *in vitro* transcription of pmt.M9 by the HA mtRNA polymerase gives ~0.15 pmol of UMP incorporated per min per μg of pmt.M9. Furthermore, assuming a size of ~830 nucleotides for the L-strand run-off transcripts synthesized from the *Kpn*I-2 fragment (Fig. 1A) and a U content of ~31% in these transcripts (16), one can calculate that ~0.02 transcript is synthesized per template molecule, under optimum conditions, during the 30-min *in vitro* incubation. This efficiency compares favorably with that of the eukaryotic RNA polymerase II systems (29).

The mtRNA polymerase activity investigated here exhibited an unusually high sensitivity to the presence of nonspecific DNAs (pUC-9 and pBR322). This inhibition may reflect a competition for essential components, possibly the RNA polymerase itself. Since the promoter regions for mtRNA polymerases appear to be relatively small (9, 10, 30–32), it is conceivable that the nonspecific DNA may contain short sequences sufficient for binding the polymerase or transcription factor(s). Binding of polymerase would not necessarily lead to initiation of transcription; the binding and isomerization steps leading to initiation can be kinetically separated for the *E. coli* RNA polymerase (33).

A considerable amount of evidence derived from mapping and kinetic studies on *in vivo* RNA and from analysis of RNA

synthesized in isolated organelles suggests that the H-strand promoter recognized by the HA mtRNA polymerase, which corresponds to the upstream *in vivo* initiation site (Fig. 1*B*), is responsible for the synthesis of the bulk of the rRNAs (1–4). It is interesting that the crude mitochondrial lysate, as well as the partially purified mtRNA polymerase preparations described in the present and previous work, failed to carry out detectable initiation of transcription at the downstream H-strand start site near the 5′-end of the 12 S rRNA gene. The latter initiation site (Fig. 1*B*) has been identified by S_1 mapping of *in vitro* capped RNA (1, 2) and is presumed to be used for the synthesis of the mRNAs and most of the tRNAs encoded in the H-strand (3, 4). A homology of the rDNA and L-strand transcription initiation regions with the region upstream of the 12 S rRNA gene has been noticed, suggesting that the same polymerase may be involved in starting transcription at the three sites (10). It is possible that the assay conditions in the present and previous work did not allow initiation at an observable level at the downstream H-strand site. On the other hand, it is also conceivable that some essential factor is lost or inactivated, or that a supercoiled template is required for the appropriate recognition by the polymerase of the downstream H-strand promoter.

Acknowledgments—We are greatly indebted to Michael King and George Gaines for supplying us with pUC-9 and M13 clones of human mtDNA, to B. Greenberg for the plasmid pBHK2, and to J. Campbell for a gift of mouse topoisomerase I. We also thank Carl Parker, George Gaines, and Anne Chomyn for their critical reading of the manuscript. The technical assistance of Arger Drew is gratefully acknowledged.

REFERENCES

1. Montoya, J., Christianson, T., Levens, D., Rabinowitz, M., and Attardi, G. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 7195–7199
2. Yoza, B. K., and Bogenhagen, D. F. (1984) *J. Biol. Chem.* **259,** 3909–3915
3. Montoya, J., Gaines, G. L., and Attardi, G. (1983) *Cell* **34,** 151–159
4. Gaines, G., and Attardi, G. (1984) *J. Mol. Biol.* **172,** 451–466
5. Ojala, D., Merkel, C., Gelfand, R., and Attardi, G. (1980) *Cell* **22,** 393–403
6. Ojala, D., Crews, S., Montoya, J., Gelfand, R., and Attardi, G. (1981) *J. Mol. Biol.* **150,** 303–314
7. Yaginuma, K., Kobayashi, M., Taira, M., and Koike, K. (1982) *Nucleic Acids Res.* **10,** 7531–7542
8. Walberg, M. W., and Clayton, D. A. (1983) *J. Biol. Chem.* **258,** 1268–1275
9. Chang, D. D., and Clayton, D. A. (1984) *Cell* **36,** 635–643
10. Bogenhagen, D. F., Applegate, E. F., and Yoza, B. K. (1984) *Cell* **36,** 1105–1113
11. Hare, J. F., Ching, E., and Attardi, G. (1980) *Biochemistry* **19,** 2023–2030
12. Attardi, B., Cravioto, B., and Attardi, G. (1969) *J. Mol. Biol.* **44,** 47–70
13. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254
14. Davison, B. L., Leighton, T., and Rabinowitz, J. C. (1979) *J. Biol. Chem.* **254,** 9220–9226
15. Greenberg, B. D., Newbold, J. E., and Sugino, A. (1983) *Gene* **21,** 33–49
16. Anderson, S., Bankier, A. T., Barrell, B. G., deBruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. E., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Stader, R., and Young, I. G. (1981) *Nature (Lond.)* **290,** 457–465
17. Vieira, J., and Messing, J. (1982) *Gene* **19,** 259–268
18. Messing, J., and Vieira, J. (1982) *Gene* **19,** 269–276
19. Casey, J., and Davison, N. (1977) *Nucleic Acids Res.* **4,** 1539–1552
20. Levens, D., Lustig, A., and Rabinowitz, M. (1981) *J. Biol. Chem.* **256,** 1474–1481
21. Tang, D. (1978) *Nucleic Acids Res.* **5,** 2861–2875
22. Attardi, G., Cantatore, P., Chomyn, A., Crews, S., Gelfand, R., Merkel, C., Montoya, J., and Ojala, D. (1982) in *Mitochondrial Genes* (Slonimski, P., Borst, P., and Attardi, G., eds) pp. 51–71, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
23. Chamberlin, M. (1976) in *RNA Polymerase* (Losick, R., and Chamberlin, M., eds) pp. 17–67, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
24. Roeder, R. G. (1976) in *RNA Polymerase* (Losick, R., and Chamberlin, M., eds) pp. 285–329, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
25. Wu, G.-J., and Dawid, I. B. (1972) *Biochemistry* **11,** 3589–3595
26. Anthony, D. D., Wu, C. W., and Goldthwait, D. A. (1969) *Biochemistry* **8,** 246–256
27. Bestwick, R. K., Moffett, G. L., and Mathews, C. K. (1982) *J. Biol. Chem.* **257,** 9300–9304
28. Posakony, J. W., England, J. M., and Attardi, G. (1977) *J. Cell Biol.* **74,** 468–491
29. Heintz, N., and Roeder, R. G. (1983) in *Genetic Engineering* (Setlow, J. K., and Hollaender, A., eds) Vol. 4, pp. 57–89, Academic Press, New York
30. Christianson, T., Edwards, J., Levens, D., Locker, J., and Rabinowitz, M. (1982) *J. Biol. Chem.* **257,** 6494–6500
31. Osinga, K. A., and Tabak, H. F. (1982) *Nucleic Acids Res.* **10,** 3617–3626
32. Christianson, T., and Rabinowitz, M. (1983) *J. Biol. Chem.* **258,** 14025–14033
33. McClure, W. R. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 5634–5638