Fine structure analyses of the *Drosophila* and *Saccharomyces* heat shock factor – heat shock element interactions

Mary Fernandes, Hua Xiao and John T. Lis*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received October 28, 1993; Revised and Accepted December 14, 1993

ABSTRACT

Heat shock genes are activated by the binding of the heat shock transcription factor (HSF) to heat shock elements (HSEs), consisting of arrays of the 5-bp unit NGAAN arranged as inverted repeats. Here, we have investigated the interaction of the 5-bp unit with HSFs of *Drosophila* and *Saccharomyces*. Mutations within the conserved, central trinucleotide GAA reduce the relative binding affinity of both HSFs. In addition, the base at position 1 (N1) also influences binding, with a strong preference for an A at this position. Methylation interference initially indicated that HSF contacts A1 in the minor groove, but interacts with the immediately adjacent base G2 in the major groove. Further characterization of this apparently abrupt minor to major groove transition by substitution of A1 with an inosine, shows that HSF contacts A1 in the major groove. We offer an explanation for this apparent contradiction and propose that HSF recognizes the HSE primarily through contacts within the major groove of the DNA helix. Finally, based on these observations and a re-evaluation of the base frequencies and criteria for consensus sequence assignment, we propose that the sequence AGAAN more accurately represents the consensus HSE motif.

INTRODUCTION

Heat shock genes provide an attractive model system for the understanding of gene regulation at the level of transcription. A key step in heat-induced transcriptional activation is the interaction of a specific transcription factor, Heat Shock Factor (HSF), with multiple copies of a sequence called the heat shock element (HSE), found in the promoter regions of heat shock genes. The basic motif of the HSE is the 5-bp sequence NGAAN (1, 2). A typical HSE is composed of multiple, contiguous units of this 5-bp sequence arranged in alternating orientation, i.e., 5'-NGAANNNTTCNNGAAN-3' (3). The number of 5-bp units in a functional HSE can vary but usually ranges from three to six. The number of HSEs in the promoter regions of different heat shock genes can also vary as can the distance between these

HSEs. Another important variable is the degree of homology of the bases in each 5-bp unit to the standard NGAAN motif. All of these factors could influence the affinity with which HSF binds the HSEs of a particular heat shock gene.

The active binding form of both *Drosophila* HSF (dHSF) and *Saccharomyces* HSF (yHSF) is a homotrimer (4–7). Although each 5-bp unit in an HSE is a recognition site for a subunit of dHSF, at least two adjacent 5-bp units are required for stable binding in vitro (4). The sequence NGAANNNTTCNNGAAN containing three alternately-oriented 5-bp units therefore represents a complete binding site for the homotrimeric HSF molecule.

Sequence comparison of genetically defined heat shock regulatory regions in *Drosophila* has shown that within the basic unit NGAAN, G2 is absolutely conserved, with A3 and A4 being highly conserved (3). While the base at position 5 is essentially random, position 1 appears to be non-random, with A seen frequently and T being rarely found. We were interested in determining whether a correlation existed between the statistical observations and the experimentally measured binding affinity of HSF. Analysis of the binding affinities of several *E. coli* proteins has demonstrated a good correlation between the in vitro binding affinity of a regulatory protein to its DNA sequence and a statistical measure expressing the difference of that sequence from the consensus (8). A functional analysis of the HSF-HSE interaction would greatly enhance our understanding of HSE recognition and would help in predicting the relative affinities of complex native HSEs for HSF. We decided therefore, to test the importance of nucleotide identity at each position of the 5-bp unit in the binding of both dHSF and yHSF.

The method of Liu-Johnson *et al.*, (9) was used to determine the relative binding affinities of both dHSF and yHSF for a trimeric HSE and its mutant derivatives that contain a single point mutation in the central 5-bp unit. Our results confirm the assignment of the consensus bases GAA. In addition, we show that position 1 can strongly influence binding, there being a fairly stringent requirement for an A at this position. Statistical analysis also shows that this base is important, indicating that the consensus HSE motif for the binding of dHSF and yHSF may be redefined as AGAAN. Methylation interference experiments

*To whom correspondence should be addressed

designed to probe the contacts made by HSF indicated a minor groove contact at A1 followed by a major groove contact at G2. We further investigated this perplexing major to minor groove switch in adjacent base pairs by substituting A1 with an inosine and determining the relative binding affinity and the methylation interference pattern. The results of these experiments indicate that the DNA binding domains of both dHSF and yHSF contact the HSE in the major groove of the DNA helix.

MATERIALS AND METHODS

Oligonucleotide synthesis and labeling

Oligonucleotides were synthesized at the DNA synthesis facility of the Biotechnology Program at Cornell University. The sequences of all the oligonucleotides used in this study are as follows:

Wild type: 5'-GGCGCCTGAT ATTCTAGAATATTCT ATGACGCGC -3'
3'-CCGCGGACTA TAAGATCTTATAAGA TACTGCGCGC -5'

Mutant 1: 5'-CTGAT ATTCTGAATATTCT ATGAC -3'
3'-GACTA TAAGAACTTATAAGA TACTG -5'

Mutant 2: 5'-CCTGAT ATTCTATAATATTCT ATGACC -3'
3'-GGACTA TAAGATATTATAAGA TACTGG -5'

Mutant 3: 5'-CCCTGAT ATTCTAGCATATTCT ATGACGC -3'
3'-GGGACTA TAAGATCGTATAAGA TACTGCG -5'

Mutant 4: 5'-GCGCCTGAT ATTCTAGATTATTCT ATGACGCGG -3'
3'-CGCGGACTA TAAGATCTAATAAGA TACTGCGCC -5'

Mutant 4': 5'-CGCCTGAT ATTCTAGATTATTCT ATGACGCG -3'
3'-GCGGACTA TAAGATCTCATAAGA TACTGCGC -5'

Mutant 5: 5'-GGACGCCTGAT ATTCTAGAAAATTCT ATGACGCGC -3'
3'-CCTGCGGACTA TAAGATCTTTAAGA TACTGCGCGC -5'

IC oligo: 5'-TGGCGCCTGAT ATTCTIGAATATTCT ATGACGCGC -3'
3'-ACCGCGGACTA TAAGACCTTATAAGA TACTGCGCGC -5'

One strand of each oligonucleotide was labeled at the 5' end with T4 polynucleotide kinase and [γ - 32 P]ATP (10) and then annealed to its complementary strand. After annealing, the double-stranded oligonucleotides were purified from the unincorporated [γ - 32 P]ATP by electrophoresis on a 12% non-denaturing polyacrylamide gel. After brief autoradiography of the wet gel, bands corresponding to the double-stranded probes were cut out from the gels, and the probes were eluted from the gel slices by soaking in TE buffer (10 mM Tris-HCl [pH 7.5], 0.1 mM EDTA), at 4°C overnight.

Gel retardation assays

These were performed as described (11).

Determination of relative binding constants

The relative binding constants of the various DNA fragments were determined using the method of Liu-Johnson *et al.*, (9). Typically, the different labeled oligonucleotides each at approximately 20 fmol ($\approx 20,000$ cpm), were mixed together with a limiting amount of purified HSF in a single binding reaction as described above. Subsequent steps were performed as described (11) and are outlined in Figure 2. Binding affinity of oligonucleotide *n* was given by $K_n = C_n/D_n$ where *C* is the concentration of labeled DNA in the bound fraction and *D* is the concentration of labeled DNA in the unbound fraction. Relative binding constants of different oligonucleotides *n* and *m* were calculated by using the equation $K_{\text{relative}} = K_n/K_m = [C_n/D_n]/[C_m/D_m]$ (9).

Plasmid constructions

dHSF and yHSF were produced using a baculovirus expression system. Wild type baculoviral DNA (AcNPV) was a gift from A. Wood (Boyce Thompson Institute, Ithaca, NY). The transfer plasmid pJVP10Z was from C.D. Richardson (Biotechnology Research Institute, Montreal, Canada).

Full length *Drosophila* HSF cDNA was a gift from Carl Wu (12). A ScaI site was introduced two nucleotides upstream of the first ATG and a ScaI-EcoRI fragment was inserted by blunt-end ligation into NheI-cut pJVP10Z DNA. Full length *Saccharomyces* HSF cDNA was a gift from Peter Sorger (13). A PvuII-EcoRI fragment was inserted into NheI-cut pJVP10Z DNA.

Insect cell culture

Propagation and purification of recombinant baculovirus was done in *Spodoptera frugiperda* (Sf 21) insect cells obtained from A. Wood (Boyce Thompson Institute, Ithaca, NY). Cells were cotransfected with wild type AcNPV baculoviral DNA and recombinant plasmids containing either the dHSF or yHSF cDNAs. Recombinant viruses were identified visually by using the β -galactosidase indicator X-gal, as 'blue plaques' that lack the occlusion bodies characteristic of wild type virus. These were then purified by passage through seven successive rounds of infection followed by plaque screening and selection. Western analyses and gel shift assays were performed with lysates from infected cells to confirm the identity of the recombinant virus.

Purification of dHSF and yHSF

Protein production was done in *Trichoplusia ni* (Tn 5B1-4) cells gifted by R.R. Granados (Boyce Thompson Institute, Ithaca, NY). Cells were plated at $\approx 25\%$ confluence on 100 mm tissue culture plates ($\approx 3 \times 10^6$ cells/plate) and infected with purified recombinant virus at an MOI of 10. Cells were harvested 48 hrs post-infection and heat shocked at 37°C for 20 min. Nuclear extracts were prepared as described (14) except that 0.25 mM TPCK, 0.025 mM TLCK, 2 μ g/ml pepstatin A, and 1 mM benzamidine were present during cell homogenization and nuclear lysis steps. After dialysis and centrifugation for 5 min at 10,000 $\times$ g, the extract was applied to a 10 ml heparin-agarose column equilibrated with 0.2 M NaCl in buffer A (20mM HEPES [pH 7.9], 0.1 mM EDTA, 0.5 mM DTT, 10% glycerol, 0.5 mM PMSF). The column was washed with the same buffer and eluted with a gradient of NaCl (0.2 to 1.5 M; 2 column volumes) in buffer A. Fractions containing HSF (determined by gel retardation assays using an end-labeled probe containing four 5-bp units) were pooled, diluted with buffer A to 0.05 M NaCl, and applied to a 5 ml Mono Q column (Pharmacia), equilibrated in the same buffer. The column was eluted with a gradient of NaCl (0.05 M to 1 M; six column volumes) in buffer A. Active fractions eluted at 0.2 M NaCl.

Methylation interference assays

Plasmid pI15 containing a single trimeric binding site for HSF (4) was restricted with AccI, singly end-labeled using Klenow and [α - 32 P]dATP at the 3' end and purified from a 1.5% low melt agarose gel. To 0.2 pmoles of this end-labeled DNA in a final volume of 100 μ l DMS reaction buffer (50 mM Na cacodylate, 1 mM EDTA, 10 mM MgCl₂), 1 μ l of DMS was added and the reaction was incubated at RT for 1 min. 50 μ l of DMS stop buffer (1 M β -mercaptoethanol, 1.5 M Na acetate)

and 750 μ l of ethanol were then added and the DNA was precipitated. Precipitation was repeated with the addition of 1 μ g polyIdC, 250 μ l 0.3 M Na acetate (7.0) / 1 mM EDTA and 750 μ l ethanol and the DNA was washed, dried and resuspended in T.E buffer. Binding reactions with methylated DNA were performed as described above using $\approx 40,000$ cpms of end-labeled, methylated DNA per binding reaction and the reaction mixture was loaded on a 2% low melt agarose gel. After overnight exposure to film at 4°C, the film was autoradiographed and DNA from bound and unbound fractions was isolated and purified by phenol extraction and ethanol precipitation. Since adenines methylate less efficiently than guanines, an adenine-enhanced cleavage technique was employed to increase the signal from methylated adenines (15). This involved gentle treatment with dilute acid followed by alkali cleavage. Briefly, the pellet was resuspended in 20 μ l dd and 5 μ l of 0.5 M HCl, mixed and stored on ice for 2 hrs with occasional stirring. The DNA was then precipitated by the addition of 0.3 M Na acetate and ethanol, air dried and resuspended in 90 μ l of 0.1 M NaOH/1mM EDTA. Cleavage was carried out by incubation at 90° for 30 min. Cleaved fragments were precipitated as before, air-dried, resuspended in formamide loading dye and separated on an 8 M urea-8% polyacrylamide sequencing gel. Methylation interference with the inosine-substituted oligonucleotide was performed similarly, except that the relevant strand was labeled at the 5' end with [γ - 32 P]ATP and DMS treatment was done for 2 min.

Quantitation

Sequencing gels from relative binding affinity experiments were quantitated using a Betascope (Betagen). Gels from methylation interference experiments were dried down after exposure to film, exposed on a phosphorimager screen and quantitated using a Molecular Dynamics 400A PhosphorImager. The accuracy of

the quantitation was verified by scanning the corresponding autoradiograms using a Bio Rad Model 620 video densitometer. Figures for the interference experiments were made from the phosphorimager printouts.

RESULTS

A minimal, complete HSE (containing three 5-bp units) was used in all the experiments as the binding site for HSF. The central 5-bp unit of this trimeric HSE was chosen as the target for mutation since it is most critical for HSF binding (11). Although binding occurs if any two of three 5-bp units are present, it is most affected when the central 5-bp unit is disrupted. We chose to examine substitutions of bases that, based on the frequency of occurrence in known regulatory regions, were predicted to have the most dramatic effect on the binding of HSF (2, 3). Each base substitution represents the most severe, transverse mutation.

Six different oligonucleotides were employed (numbered 1 to 5 and wild type, see Fig. 1A), each containing one complete binding site for a trimeric HSF molecule. We chose to designate the 5-bp unit AGAAT 'wild type', for consistency with earlier binding studies (4, 11) and because it contains the most conserved bases at positions 1 to 4, the base at position 5 being essentially random (see Table 1). Oligonucleotides 1 to 5 contain a single nucleotide change at positions 1 to 5 respectively, in the central 5-bp unit of the wild type binding site. Each oligonucleotide

Table 1. Re-evaluation of the base pair frequencies in heat shock elements of (A) *Drosophila* and (B) *Saccharomyces*.

(A) *Drosophila*

Nucleotide	Position				
	1	2	3	4	5
A	31 (62%)	0 (0%)	48 (96%)	43 (86%)	23 (46%)
G	4 (8%)	50 (100%)	0 (0%)	5 (10%)	14 (28%)
C	14 (28%)	0 (0%)	2 (4%)	1 (2%)	1 (2%)
T	1 (2%)	0 (0%)	0 (0%)	1 (2%)	12 (24%)

(B) *Saccharomyces*

Nucleotide	Position				
	1	2	3	4	5
A	28 (56%)	0 (0%)	46 (92%)	46 (92%)	12 (24%)
G	12 (24%)	48 (96%)	2 (4%)	2 (4%)	8 (16%)
C	6 (12%)	0 (0%)	0 (0%)	0 (0%)	19 (38%)
T	4 (8%)	2 (4%)	2 (4%)	2 (4%)	11 (22%)

Only those HSEs shown to be functional (by mutational analysis and protein binding) have been included. This analysis does not include motifs with more than one mismatch at positions 2 to 4. A total of 50 motifs were considered in each case [for complete sequences see refs. (3) and (20)]. Nucleotide frequencies at each position of the 5-bp unit are indicated, with the % value shown in parentheses.

Figure 1. Wild type and mutant HSE DNAs. **A.** The central 5-bp unit of the trimeric HSE oligonucleotides. Consensus nucleotides at positions 1 to 5, and nucleotides in the wild type and mutant binding sites are indicated. Circles indicate mutated nucleotides. For complete sequences see Materials and Methods. **B.** Purified, singly end-labeled oligonucleotides denoted 1,2,3,4,5 and WT (wild type), on a native 15% polyacrylamide gel. Although each oligonucleotide has a single trimeric binding site, differences in the lengths of the flanking sequences allow separation according to size.

differs slightly in length from the others, in order to facilitate differentiation on sequencing gels as described below. The oligonucleotides were end-labeled on one strand, annealed to the complementary strand and purified on a native acrylamide gel as shown in Figure 1B.

The method of Liu-Johnson *et al* [see Methods and refs. (9, 11)] was used to determine the relative binding affinities of dHSF and yHSF for the six oligonucleotides. This method allows one to directly compare the relative affinities of different binding sites in a single reaction. A sample experiment is shown in Figure 2. The different sites to be compared are mixed together in a single binding reaction with a limiting amount of HSF (Fig 2A). After a time to achieve equilibrium, the reaction mixture is electrophoresed on a native agarose gel (Fig. 2B). DNA is then isolated from both bound and unbound fractions and separated according to size, on a denaturing, sequencing gel (Fig. 2C). This method is convenient since it does not require prior

Figure 2. Protocol for the measurement of relative binding constants. **A.** The end-labeled oligonucleotides are mixed together with a limiting amount of HSF in a single binding reaction. The asterisk indicates the 5'-end-label and the black bar indicates the trimeric HSE in each oligonucleotide. This particular experiment has an additional mutant (oligonucleotide 4') with an A to G substitution at position 4. As expected, it exhibits a higher affinity for HSF than mutant 4 which has a more severe A to T mutation at this position. Purified *Drosophila* HSF was used in this reaction. **B.** After equilibrium is reached, the reaction is electrophoresed on a native agarose gel to separate bound (B) and unbound (F) DNA. **C.** DNA from the two fractions is purified and examined on a denaturing sequencing gel. Quantitation of each band allows the determination of relative binding constants for each oligonucleotide in the mixture.

determination of the concentration of active HSF and the specific activities of the labeled DNA fragments, both of which are difficult to measure accurately.

We found that differences in fragment length of the different oligonucleotide probes did not affect their recovery from gel slices or the measurements of relative binding affinities (data not shown). Also, dHSF expressed in the baculovirus system behaved essentially like native dHSF from heat-shocked *Drosophila* K_c cells, except that the reduction in relative binding affinities of the mutants obtained with native dHSF was about twice that with the baculovirus-derived dHSF (data not shown).

Role of individual bases of the 5-bp consensus in the binding of dHSF

When the effects of base mutation at each position in the wild type 5-bp unit AGAAT are compared, mutation of G2 is seen to be the most deleterious for the binding of dHSF, with a 16-fold reduction in binding affinity (Fig. 3A). Substitution of a T for A1 of the 5-bp unit is the second most severe, reducing the binding affinity 5-fold. This is followed by mutations of A3 and A4, which display 2-fold reductions in binding affinity. Substitution of an A for T5 does not significantly affect the binding affinity of dHSF.

Role of individual bases of the 5-bp consensus in the binding of yHSF

With yHSF as with dHSF, mutation of the core consensus bases GAA reduces binding affinity in the order G2 > A3 > A4 (compare Fig. 3B with Fig. 3A). Interestingly however, substitution of A1 with a T is the most detrimental and causes a dramatic, 20-fold reduction in the binding affinity of yHSF. This mutation is more severe than mutation at any position within the trinucleotide consensus including G2 (which exhibits only a 5-fold decrease). Substitution of T5 with an A decreases the binding affinity of yHSF about 1.8-fold.

Base specific contacts detected by methylation interference

Does the reduction in binding affinity caused by substitution of the A at position 1 of the central 5-bp unit reflect a base specific contact by HSF at this position? To address this question, we employed a methylation interference protocol (16). This assay can reveal contacts between HSF and the purines within the HSE.

Figure 3. Effect of mutation at each position of the consensus on the binding of dHSF and yHSF. **A.** Relative binding constants for dHSF. **B.** Relative binding constants for yHSF. Values are the average of three separate experiments, with error bars representing one standard deviation.

Briefly, DNA containing a single, wild type trimeric binding site is labeled on the relevant upper strand and pretreated with dimethyl sulfate (DMS) to produce, on average, one methylation per DNA molecule. This DNA is then incubated with HSF in a binding reaction and electrophoresed on a native agarose gel. Free and bound DNAs are separated, purified, cleaved at the methylated residues and analyzed on a denaturing sequencing gel. Fragments that were methylated so as to weaken the binding of HSF are expected to be underrepresented in the bound complex. Since DMS introduces a methyl group onto the N3 position of adenine, which projects into the minor groove, and the N7 position of guanine which projects into the major groove, these experiments not only identify the purines involved in contacting HSF, but also suggest the locations of the interactions with respect to the grooves of the DNA double helix.

The results of the interference assay indicate that methylation of A1 and G2 (of the central 5-bp unit) strongly interferes with the binding of both yHSF and dHSF (Fig. 4, compare lanes 1 and 2 for yHSF and lanes 3 and 4 for dHSF). A quantitation of the results based on densitometric scanning shows that for yHSF, methylation of A1 and G2 causes a 2.8-fold and a 3.2-fold decrease respectively, in the intensity of the bound probe as compared to the unbound probe. This result implies that HSF makes a minor groove contact at position A1 followed by a major groove contact with the very next base G2. A similar methylation interference pattern has been observed with a dimeric HSE consisting of two 5-bp units (4). Methylation of A1 appears to be more deleterious for binding of yHSF as compared to dHSF.

HSF contacts the A at position 1 in the major groove

The surprising major to minor groove transition implied by the methylation interference experiments provoked us to investigate

more closely the nature of the contact at position A1. To further test whether HSF contacts the A at position 1 in the minor groove, we substituted an I-C base pair for the A-T base pair at position 1 in the central 5-bp unit of the wild type oligonucleotide. In terms of the spatial locations of hydrogen bond donors and acceptors, inosine looks like a G in the major groove and an A in the minor groove. If therefore, HSF contacts the base at position 1 in the minor groove, then this substitution should not affect binding affinity. However, if HSF contacts the major groove, then a decrease in binding affinity would be expected. This decrease would reflect the extent that I departs from A in the major groove. When the relative binding affinities of both HSFs for the substituted oligonucleotide were determined, a 1.5-fold decrease in relative binding affinity was observed with yHSF, and a smaller effect was seen with dHSF (data not shown).

To more decisively discriminate between major and minor groove interactions, we decided to perform a methylation interference experiment with the inosine-substituted oligonucleotide. Inosine, like guanine, is methylated by DMS at the N-7 atom in the major groove (17). If HSF makes a contact in the major groove, methylation of the inosine should strongly interfere with binding, whereas no interference would be indicative of a minor groove interaction. A similar approach has been used to show that TFIID binds in the minor groove of the TATA element (18). As shown in Figure 5, methylation of the I at position 1 severely affects binding of HSF, indicating that HSF makes a major groove contact with the A at position 1 of the consensus sequence AGAAT.

DISCUSSION

This work has assessed the importance of each base in the 5-bp heat shock element motif NGAAN, to the binding of *Drosophila* and *Saccharomyces* HSFs. We have shown that in general, transverse mutations within this 5-bp sequence negatively affect,

Figure 4. Methylation interference pattern of HSF with the wild type HSE. B=bound DNA, F=free DNA, G+A= G+A sequence ladder. The patterns obtained with both yHSF (lanes 1 and 2) and dHSF (lanes 3 and 4) are shown. A densitometric tracing of the methylation interference pattern for the bound (B) and free (F) lanes of yHSF is shown at left.

Figure 5. Methylation interference pattern of yHSF with the inosine-substituted HSE. Symbols are as in Figure 4.

albeit to different extents, the binding of both *Drosophila* HSF and *Saccharomyces* HSF. We find that our experimentally derived results are in good agreement with earlier, statistically derived predictions of base importance at each position of the 5-bp motif (3). This study (3) had designated the three central nucleotides, GAA, of the 5-bp unit as consensus nucleotides, based on their frequency of occurrence: 100% for G2, 80% for A3 and 68% for A4. Measurements of relative binding affinities shown here reveal a similar trend in the contribution of these bases to binding. For both dHSF and yHSF, we find that mutations within the conserved trinucleotide core follow the pattern $G2 > A3 > A4$ for reduction in binding affinity.

Interestingly, the A at position 1 of the 5-bp unit seems to play an important role in binding. Substitution of a T for the A reduces the binding affinity of dHSF 5-fold and that of yHSF 20-fold. This base had originally not been included among the consensus nucleotides, based on a frequency of occurrence of 57% (3). However, we observe that mutation of A1 results in a greater reduction in binding affinity than mutations of the two consensus As, A3 and A4. A close comparison of our results and the earlier statistical analysis reveals that the difference between the two studies is one of interpretation rather than data, since A1 qualifies as a consensus base even with a frequency of 57%, based on criteria put forward by Cavener (19). According to these criteria, referred to as the 50/75 Consensus Rule, if a nucleotide at a specific position occurs in more than 50% of the sequences and is greater than twice the frequency of the next most frequent nucleotide, it is assigned as the sole consensus nucleotide.

We have re-evaluated the base pair frequency in functional *Drosophila* heat shock elements, applying more stringent criteria so that motifs with more than one mismatch in the core GAA were not considered. This analysis reveals that A occurs at position 1 with a frequency of 62% (see Table 1). Given all of the above considerations, it appears that A is both the optimal and the consensus base at position 1. A comparison of the known heat shock regulatory regions in *Saccharomyces* (20) indicates that here too, A occurs at position 1 at a high enough frequency to be considered a consensus base (see Table 1). Also in agreement is an earlier, more global comparison of HSEs from heat shock genes of organisms as diverse as *C.elegans*, soybean, chicken and *X.laevis* (2). This work also showed that substitution of a T for A1 dramatically reduces heat-induced gene expression. Furthermore, a recent analysis of the binding of human HSF1 to the HSE of the native hsp70 promoter, has revealed that the effect on binding affinity of mutation from an A to a T at this location is almost as severe as elimination of any consensus base from the central GAA sequence (21).

Although we observe in the case of yHSF, a greater reduction in binding affinity for an A1 to T mutation as opposed to a G2 to T mutation, we do not intend to emphasize the importance of A1 over that of G2, since we have not examined every possible mutation at each position of the 5-bp unit. In this context, it should also be noted that even though A is preferred at position 1 in both *Saccharomyces* and *Drosophila*, G and C do occur at this position at a reasonable frequency, as opposed to position 2, where G is almost exclusively observed. For the binding of human HSF, mutation of A1 to G or C, while suboptimal, are nevertheless tolerated, in contrast with G2 where a change to any other base is detrimental (21). This suggests that base identity at position 1, while important, may be more flexible than that at position 2. Nevertheless, the fact that we observe only a 5-fold reduction in binding of yHSF to a site with a G2 to T mutation (in contrast to a 20-fold reduction for A1 to T), is intriguing given

that yeast HSEs may contain at fairly high frequency, imperfect core sequences with a G2 to T substitution (20).

Intrigued by the apparent significance of A1, we performed methylation interference experiments to probe the contacts made by yHSF and dHSF at this position. Earlier methylation interference experiments using a dimeric binding site (two 5-bp units), for dHSF implicated G2 of the 5-bp consensus in making a major groove contact and also gave some indication that methylation of A1 may interfere with binding of dHSF (4). However, there is no documentation in the literature of the methylation interference pattern produced by *Saccharomyces* HSF. Our results show that methylation of A1 and G2 strongly interferes with the binding of yHSF. We observe similar results with dHSF as well, except that the interference at A1 appears to be somewhat weaker as compared to that observed with yHSF.

Since adenine methylation changes the surface of the minor groove of the DNA helix, the straightforward interpretation of these results would be that HSF makes a minor groove contact at A1. Although it is difficult to imagine how a major and a minor groove contact may be made at successive base pairs, such an observation is not without precedent. Crystal structure of the engrailed homeodomain-DNA complex has revealed that conserved residues from the N-terminal arm contact base pairs within the minor groove that are adjacent to the ones contacted by helix 3 in the major groove (22). Similar adjacent minor and major groove contacts have been observed with the *hin* recombinase and the oncoprotein Ets-1 (23, 24). If indeed HSF makes a minor groove contact, this finding has to be reconciled with the observation of Seeman *et al.*, that recognition in the minor groove is likely to be relatively insensitive to base pair reversal of the kind A-T to T-A (25). We find on the contrary, that mutation of the A-T base pair at position 1 to a T-A base pair drastically affects the binding affinity of both yHSF and dHSF, arguing against interaction in the minor groove.

An alternative explanation would be that HSF makes a major groove contact at the position 1 adenine and that methylation at N3 may prevent HSF from making base-specific hydrogen bonds in the major groove. It is known that modification of the purine ring with DMS results in the addition of a positive charge to the ring which then renders the modified residue labile (26). A potential recognition site in the major groove is the imidazole N7 atom of the adenine ring which functions as a hydrogen bond acceptor. Introduction of a positive charge into the ring would result in the withdrawal of electrons from this nitrogen, reducing its ability to act as a hydrogen bond acceptor. The resulting interference with the binding of HSF would be indistinguishable in the methylation interference assay from that caused by the introduction of a methyl group at the N3 position in the minor groove.

Methylation at N3 of the adenine ring could also alter the ability of DNA to adopt a conformation necessary for major groove interaction at A1. A good example is the binding of Eco RI endonuclease to its cognate sequence (27, 28). In this case, although methylation interference experiments indicated a major groove contact at the G and minor groove contacts at the two As in the recognition sequence 5'-GAATTC-3', crystal structure of the protein-DNA complex clearly showed that the G as well as the As were contacted in the major groove.

In the absence of a co-crystal structure for HSF, we decided to resolve this dilemma by examining the binding of HSF to an oligonucleotide identical to the wild type, but with the adenine at position 1 replaced with an inosine (the T on the complementary strand was replaced with a C). Using two different criteria,

namely, determination of the methylation interference pattern and measurement of relative binding affinity, we conclude that HSF makes a major groove contact with the heat shock element. While this is seen with both yHSF and dHSF, we suspect that the involvement of A1 of the HSE in the putative major groove interaction is probably more pronounced in the case of yHSF. This would account for its greater sensitivity to mutation of the type A1 to T1 and also for its decreased binding to the inosine-substituted oligonucleotide relative to the wild type sequence.

Additional support for our hypothesis comes from studies on the binding of human HSF, where data from DMS and DEPC interference and hydroxyl radical protection experiments indicate that the NGAAN element is contacted in the major groove (29, 30, 31). These studies have focused on the human hsp70 gene which contains five 5-bp units, of which the two centered at position -100 mainly contribute to stress-induced transcription. Both 5-bp units have the sequence 5'-GGAAT-3' and DMS methylation and DEPC interference experiments reveal strong HSF contacts at G1 and G2 of each motif. Since DEPC is known to carboxyethylate adenines and guanines at N7 in the major groove (32), these results suggest that human HSF makes significant major groove contacts at positions 1 and 2 of the NGAAN sequence. Furthermore, hydroxyl radical protection experiments show very weak protection of the GG-AAT-containing strand of the 5-bp motif, suggestive of almost exclusive major groove contacts (33). Recent DMS protection experiments performed with the two mouse HSFs, mHSF1 and mHSF2 on the human hsp70 promoter also reveal strong contacts at G1 and G2 (34). Interestingly, protection at A1 of one of the flanking repeats was also observed in this study and interpreted as a minor groove contact. HSF from the fission yeast *Schizosaccharomyces pombe* has also been shown by methylation interference to contact position G1 of the GGAAT units in the human hsp 70 gene within the major groove (35).

It should be noted that our results do not rule out a simultaneous major and minor groove interaction at position 1. It is possible that a flexible arm or loop in HSF could wrap around the DNA making additional contacts in the minor groove. While we favor a major groove interaction based on our data, the test of this prediction rests in the solving of the X-ray structure of the HSF-HSE co-crystal. Nevertheless, we would like to strike a note of caution against the conventional interpretation of interference at DMS-modified adenines being necessarily due to disruption of protein contacts in the minor groove.

This work has highlighted the significance of the nucleotide at position 1 of the 5-bp unit and its participation in the major groove interaction made by both *Drosophila* and *Saccharomyces* HSFs. We have confirmed the importance of the core consensus bases to HSF binding, as well as the relative unimportance of base identity at position 5. In conclusion, we are struck by the overall conservation of the HSE motif in organisms as diverse as yeasts, fruit flies and humans. This observation is consistent with the discovery that HSF-encoding genes from these and other eukaryotic species examined so far, show a high degree of homology only in their DNA binding and oligomerization domains (12, 31 and 36-40).

ACKNOWLEDGEMENTS

We thank Tom O'Brien and Charles Giardina for critically reading the manuscript and Janet Ingraffea for invaluable help with the baculovirus expression system. This work was supported by National Institutes of Health Grant GM 25232 to J.T.L.

REFERENCES

1. Xiao, H. and Lis, J. T. (1988) *Science*, **239**, 1139-1142.
2. Amin, J., Ananthan, J. and Voellmy, R. (1988) *Mol. Cell. Biol.*, **8**, 3761-3769.
3. Lis, J. T., Xiao, H. and Perisic, O. (1990) In Morimoto, R., Tissieres, A. and Georgopoulos, C., (Eds.), *Stress Proteins in Biology and Medicine*. Cold Spring Harbor Press, Cold Spring Harbor, pp. 411-428.
4. Perisic, O., Xiao, H. and Lis, J. T. (1989) *Cell*, **59**, 797-806.
5. Westwood, J. T. and Wu, C. (1993) *Mol. Cell. Biol.*, **13**, 3481-3486.
6. Sorger, P. K. and Nelson, H. C. M. (1989) *Cell*, **59**, 807-813.
7. Peteranderl, R. and Nelson, H. C. M. (1993) *Biochemistry*, **31**, 12272-12276.
8. Berg, O. G. and von Hippel, P. H. (1988) *Trends Biochem. Sci.*, **13**, 207-211.
9. Liu-Johnson, H.-N., Gartenberg, M. R. and Crothers, D. M. (1986) *Cell*, **47**, 995-1005.
10. Sambrook, J., Fritsch, E. F. and Maniatis, T., (1989) *Molecular Cloning—A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
11. Xiao, H., Perisic, O. and Lis, J. T. (1991) *Cell*, **64**, 585-593.
12. Clos, J., Westwood, T., Becker, P. B., Wilson, S., Lambert, K. and Wu, C. (1990) *Cell*, **63**, 1085-1097.
13. Sorger, P. K. and Pelham, H. R. B. (1988) *Cell*, **54**, 855-864.
14. Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V. and Ueda, H. (1987) *Science*, **238**, 1247-1253.
15. Maxam, A. M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
16. Siebenlist, U. and Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA*, **20**, 122-126.
17. Scheit, K.-H. and Holy, A. (1967) *Biochim. Biophys. Acta*, **149**, 344-354.
18. Starr, D. B. and Hawley, D. K. (1991) *Cell*, **67**, 1231-1240.
19. Cavener, D. R. and Ray, S. C. (1991) *Nucleic Acids Res.*, **19**, 3185-3192.
20. Boorstein, W. R. and Craig, E. A. (1990) *J. Biol. Chem.*, **265**, 18912-18921.
21. Cunliff, N. F. A. and Morgan, W. D. (1993) *J. Biol. Chem.*, **268**, 8317-8324.
22. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. and Pabo, C. O. (1990) *Cell*, **63**, 579-590.
23. Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I. and Dervan, P. B. (1990) *Biochemistry*, **29**, 6551-6561.
24. Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D. and Graves, B. J. (1992) *Genes Dev.*, **6**, 975-990.
25. Seeman, N. C., Rosenberg, J. M. and Rich, A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 804-808.
26. Lawley, P. D. and Brookes, P. (1963) *Biochem. J.*, **89**, 127-138.
27. Lu, A.-L., Jack, W. E. and Modrich, P. (1981) *J. Biol. Chem.*, **256**, 13200-13206.
28. McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J. and Rosenberg, J. M. (1986) *Science*, **234**, 1526-1541.
29. Mosser, D. D., Theodorakis, N. G. and Morimoto, R. I. (1988) *Mol. Cell. Biol.*, **8**, 4736-4744.
30. Cunliff, N. F. A., Wagner, J. and Morgan, W. D. (1991) *Mol. Cell Biol.*, **11**, 3504-3514.
31. Schuetz, T. J., Gallo, G. J., Sheldon, L., Tempst, P. and Kingston, R. E. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 6911-6915.
32. Herr, W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8009-8013.
33. Vinson, C. R., Sigler, P. B. and McKnight, S. L. (1989) *Science*, **246**, 911-916.
34. Kroeger, P. E., Sarge, K. D. and Morimoto, R. I. (1993) *Mol. Cell. Biol.*, **13**, 3370-3383.
35. Gallo, G., Schuetz, T. and Kingston, R. (1991) *Mol. Cell. Biol.*, **11**, 281-288.
36. Sorger, P. (1991) *Cell*, **65**, 363-366.
37. Sarge, K. D., Zimarino, V., Holm, K., Wu, C. and Morimoto, R. I. (1991) *Genes Dev.*, **5**, 1902-1911.
38. Scharf, K.-D., Rose, S., Zott, W., Schoff, F. and Nover, L. (1990) *EMBO J.*, **9**, 4495-4501.
39. Wiederrecht, G., Seto, D. and Parker, C. S. (1988) *Cell*, **54**, 841-853.
40. Rabindran, S. K., Giorgi, G., Clos, J. and Wu, C. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 6906-6910.