Genome-wide Analysis Reveals New Roles for the Activation Domains of the *Saccharomyces cerevisiae* Heat Shock Transcription Factor (Hsf1) during the Transient Heat Shock Response*\(^\text{**}\)

Received for publication, March 15, 2006, and in revised form, August 4, 2006 Published, JBC Papers in Press, August 22, 2006, DOI 10.1074/jbc.M602454200

Dawn L. Eastmond and Hillary C. M. Nelson\(^1\)

*From the Department of Biochemistry and Molecular Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104*

In response to elevated temperatures, cells from many organisms rapidly transcribe a number of mRNAs. In *Saccharomyces cerevisiae*, this protective response involves two regulatory systems: the heat shock transcription factor (Hsf1) and the Msn2 and Msn4 (Msn2/4) transcription factors. Both systems modulate the induction of specific heat shock genes. However, the contribution of Hsf1, independent of Msn2/4, is only beginning to emerge. To address this question, we constructed an *msn2/4* double mutant and used microarrays to elucidate the genome-wide expression program of Hsf1. The data showed that 7.6% of the genome was heat-induced. The up-regulated genes belong to a wide range of functional categories, with a significant increase in the chaperone and metabolism genes. We then focused on the contribution of the activation domains of Hsf1 to the expression profile and extended our analysis to include *msn2/4* strains deleted for the N-terminal or C-terminal activation domain of Hsf1. Cluster analysis of the heat-induced genes revealed activation domain-specific patterns of expression, with each cluster also showing distinct preferences for functional categories. Computational analysis of the promoters of the induced genes affected by the loss of an activation domain showed a distinct preference for positioning and topology of the Hsf1 binding site. This study provides insight into the important role that both activation domains play for the Hsf1 regulatory system to rapidly and effectively transcribe its regulon in response to heat shock.

The ability to respond to a large number of environmental stresses is crucial for the survival of all organisms (1). Heat shock is among a highly diverse group of environmental conditions that alter gene expression in prokaryotic and eukaryotic cells. The response to heat shock is characterized by a rapid induction of a conserved group of heat shock proteins (HSPs).\(^2\) In *Saccharomyces cerevisiae*, two regulatory systems are involved in this response: the heat shock transcription factor (Hsf1) (2, 3) and the Msn2 and Msn4 (Msn2/4) transcription factors (4, 5).

Yeast Hsf1 is an essential protein that binds to inverted repeats of *NGAAAn* called heat shock elements (HSEs) within the promoters of many HSPs and activates their transcription. Hsf1 is composed of several well-defined domains that are important for its function. They include the highly conserved central core, which is made up of the winged-helix-turn-helix DNA-binding domain (6, 7) and the hydrophobic coiled-coil region essential for the regulation of homotrimer formation (8–10). In addition, yeast Hsf1 has two trans-activation domains, one at the N terminus and the other at the C terminus (11, 12).

Although the structure and function of HSF is generally conserved from various organisms, there is variability in the number and importance of HSF genes in any particular organism. The yeasts, *S. cerevisiae* and *Schizosaccharomyces pombe*, have one HSF gene that is essential for cell survival (2, 13, 14), whereas *Drosophila melanogaster* has a single HSF gene that is essential for oogenesis, early larval development, and survival in response to acute stress but is unnecessary for cell growth and viability (15). In contrast, chickens and mammals have 3 HSF genes that vary in their cellular and physiological roles (16), whereas *Arabidopsis thaliana* has 21 HSF genes that are tightly regulated into a network of interacting proteins (17).

The existence of multiple HSF species in higher eukaryotes suggests that HSF isoforms may have specific functions that are triggered by distinct stimuli or activate specific target genes. So how does *S. cerevisiae*, with one Hsf1 isoform, regulate such a complex response to environmental stress? Yeast Hsf1 differs from metazoan HSFs in several important ways (16). First, *S. cerevisiae* utilizes the single essential Hsf1 to activate the expression of a wide variety of genes under normal physiological conditions and in response to environmental stress. Metazoan cells have three heat shock factors that only function under specific developmental or environmental conditions.

---

\(^{*}\) This work was supported by National Institutes of Health Grant GM44086 and by the University of Pennsylvania Research Foundation. 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 on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1.

1 To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biophysics, University of Pennsylvania School of Medicine, Rm. 814, Stellar-Chance Bldg., 422 Curie Blvd., Philadelphia, PA 19104. Tel.: 215-573-7473; Fax: 215-573-2503; E-mail: hnelson@mail.med.upenn.edu.

\(^2\) The abbreviations used are: HSP, heat shock protein; Hsf1, heat shock transcription factor; Msn2/4, Msn2 and Msn4 transcription factor; HSE, heat shock element; NAD, N-terminal activation domain; CAD, C-terminal activation domain; STRE, stress response element; MOPS, 4-morpholinepropanesulfonic acid; ORF, open reading frame; MES, 4-morpholineethanesulfonic acid; ChIP, chromatin immunoprecipitation; SGD, *Saccharomyces* Genome Database; QT, quality threshold; TAF, TATA-binding protein-associated factor.
Role of the Activation Domains of Yeast Hsf1

Second, in the absence of heat shock, *S. cerevisiae* Hsf1 exists in the nucleus as a trimer bound to HSEs. In contrast, mammalian HSF1 exists in the cytoplasm as an inactive monomer. In response to heat shock, mammalian HSF undergoes nuclear localization, trimerizes, and binds to HSEs. Third, *S. cerevisiae* Hsf1 has two trans-activation domains, whereas the metazoan HSF has only one located at the C terminus.

The activation domains of the *S. cerevisiae* Hsf1 appear to mediate different temporal aspects of the heat shock response, which can be divided into two phases: the transient response and the sustained response (11, 12, 18). The transient response is characterized as how cells tolerate an increase in temperature over 35 °C for <1 h, whereas the sustained response is characterized as how cells acclimate to the higher temperatures for long periods of time. The N-terminal activation domain (NAD), found in the first 65 amino acids, is thought to mediate the transient response, whereas the C-terminal activation domain (CAD), found between residues 595 and 783, is thought to be responsible for the sustained response to stress. Although there have been no detailed studies of the genes effected by the NAD, the role of the CAD has been studied for specific genes. The CAD is critical for the heat-induced expression of *CUP1, HSP82,* and *HSP26,* whereas the loss of the CAD has no effect on the heat-induced expression of *SSA1, SSA3,* and *HSP104* (19–21).

The second regulatory system of the heat shock response, Msn2 and Msn4, has not been conserved between yeasts and humans. These proteins are non-essential transcription factors that bind to stress response elements (STREs) found in the promoters of most HSPs (22). Msn2 and Msn4 are 41% identical in amino acid sequence and 100% conserved at three residues that are pertinent for DNA sequence recognition (23). Under constitutive conditions, Msn2/4 proteins are localized to the cytoplasm; however, during heat shock, they are transported to the nucleus where they activate transcription of HSPs (24).

The contributions of Msn2/4 and Hsf1 proteins to the yeast heat shock response have been addressed previously (25–28). These studies, although not exhaustive, establish that each regulatory system contributes differently to the expression of genes. For some genes, such as *HSP26* and *HSP104,* their expression is regulated by cooperation between both Msn2/4 and Hsf1 (19, 29), whereas other genes appear to be regulated by only one system. The Msn2/4 regulatory system controls the expression of genes classified as chaperones, carbon metabolism, and oxidative stress (25, 28), whereas the Hsf1 regulatory system controls the expression of genes classified as chaperones, cell wall maintenance, and energy regeneration (26, 27, 30).

In this study, we used DNA microarrays to quantify the contribution that the Hsf1 regulatory system makes to the transient heat shock response in the absence of the Msn2/4 transcription factors. In particular, we focused on the roles of the activation domains of Hsf1 by analyzing the expression patterns of strains with NAD or CAD deletions. We applied clustering algorithms to our data to illuminate physiological and mechanical relationships among genes. Promoter analysis showed architectural distinctions between the different clustered genes. These data show that Hsf1, its activation domains, and HSEs all play distinctive roles in the reorganizing and orchestrating of cellular events during the heat shock response.

**EXPERIMENTAL PROCEDURES**

*S. cerevisiae* strain W303-1A (MATa ade2–1 trp-1 can-1 leu2-3,112 his3-11,15 ura3-1). We used the KanMX cassette in combination with cre-loxP recombination technology to create the *msn2::loxP msn4::loxP* null alleles (31). Creation of the activation domain deletion strains has been previously described (32).

Briefly, *URA3*-marked integration plasmids that contained Hsf1 genes deleted for either the N- or C-terminal activation domains were constructed. These plasmids were used to replace the wild-type *HSF1* allele with either the *hsf1(66–833)* or *hsf1(1–583)* alleles in both *MSN2/4* and *msn2/4* strains. Strains with the wild-type *HSF1* were designated Hsf, strains with *hsf1(66–833)* were designated HSFANAD, and strains with the *hsf1(1–583)* allele were designated HSFACAD. Cells were grown in yeast extract-peptone-dextrose medium supplemented with adenine (YPAD) at 30 °C to mid-log phase (**A**<sub>600</sub> of 0.5).

**RNA Preparation and Northern Blot Analysis**—The culture was split into two, with one half remaining at standard growing conditions, while the other half was heat shocked at 37 °C for 15 min. Cells were harvested by centrifugation at 258 × G at 4 °C for 5 min, flash frozen in liquid nitrogen, and stored at −80 °C. Total RNA was isolated using the hot acidic phenol method (33). Pellets were resuspended in RNase-free dH<sub>2</sub>O, quantified, and store at −80 °C until needed. RNA samples (7 µg/lane) were incubated in a 1:1 ratio of loading buffer (50% formamide, 19% formaldehyde, 40 mM MOPS, pH 7, 10 mM sodium acetate, 4.2% glycerol, 0.03% xylene cyanol, 0.03% bromphenol blue, 1 mM EDTA, pH 8) at 65 °C for 10 min. The RNA samples were resolved on formaldehyde-agarose gels and then transferred onto ZetaProbe Nylon membrane (Bio-Rad). The membrane was cross-linked in the UV Stratalinker 2400 (Stratagene) and washed in 0.1× SSC at room temperature. The membrane was pre-hybridized in RapidHyb buffer (Amersham Biosciences) at 55 °C for a minimum of 2 h. Probes were generated by incorporating Biotin-16-dUTP nucleotide (Roche Applied Science) into PCR products of the *HSP12, HSP26, HSP82, HSP104, SSA3, SSA4,* and *SSE2* open reading frames (ORFs). Hybridization was carried out with denatured probes in the same buffer at 55 °C overnight. The following day, the filters were washed three times in 300 mM NaCl; 30 mM sodium citrate, 0.1% SDS at 55 °C for 20 min. The protocol from the North2South Chemiluminescent Detection Kit (Pierce) was used for probe detection and substrate development. Signals were then quantified on a Fluorchem 8800 cooled charge-coupled device detection system (Alpha Innotech Corp.).

**Microarray Hybridization**—The Affymetrix yeast genome YG_S98 microarrays used in this study contain ~6400 known or predicted open reading frames as described in the *Saccharomyces* Genome Database (SGD) and additional probe sets representing putative ORFs identified by SAGE analysis, mitochondrial proteins, and TY proteins (Affymetrix).
All protocols were conducted as described in the Affymetrix GeneChip Expression Analysis Technical Manual. Briefly, 10 μg of total RNA was incubated with Superscript II reverse transcriptase primed by a poly(T) oligomer that incorporated the T7 promoter to convert the RNA into first-strand cDNA. Second-strand cDNA synthesis was followed by in vitro transcription for linear amplification of each transcript and incorporation of biotinylated CTP and UTP. The cRNA products were fragmented to an average of 200 nucleotides, heated at 99 °C for 5 min, and hybridized for 16 h at 45 °C to the Affymetrix microarrays at the University of Pennsylvania Microarray Core Facility. The microarrays were then washed at low (0.9 m NaCl, 60 mM Na_2HPO_4, pH 7.5, 6 mM EDTA, 0.005% Triton X-100) and high (100 mM MES, 0.1 m NaCl) stringency and stained with streptavidin-phycoerythrin. Fluorescence was amplified by adding biotinylated anti-streptavidin and an additional aliquot of streptavidin-phycoerythrin stain. A confocal scanner was used to collect fluorescence signal at 3-μm resolution after excitation at 570 nm. The average signal from two sequential scans was calculated for each microarray feature. Gene expression values were calculated using Affymetrix Microarray Suite, version 5.0, and exported as “cch” files. The data were imported and analyzed using GeneSpring GX version 7.2 (Agilent Technologies). The cch files were pre-normalized (centered on a trimmed mean of 150). In GeneSpring, the values were further normalized “per chip” by dividing each intensity value by the median intensity of the entire chip, and then “per gene” by dividing the signal levels of a given gene across all 18 samples by the median signal across the same samples for that gene.

Data Acquisition and Analysis—First, an experiment was set up to group the three independent replicates of each condition and strain together. This generated an average signal log ratio value, which reflects the abundance of each Affymetrix probe set. At the end of this summarization process, there were six normalized values for each replicate group, one for each strain and condition tested. Then, a pair-wise comparative analysis was performed between the experimental heat shock and the baseline constitutive microarrays for each of the three strains. The probe sets were designated present or induced if they were enriched by a minimum of 2-fold (normalized fold change (NFC) = 2.0), were used for further analysis. Quality threshold (QT) clustering was performed to uncover patterns of gene expression and the relationships between these patterns (35). The algorithm builds a cluster by starting with a single locus and a cluster diameter of zero. One by one, other loci are added to the cluster and a new cluster diameter is calculated. If a locus cannot be placed into an existing cluster without sacrificing the quality or the diameter threshold, it is placed into a new cluster. As a starting point, we analyzed the loci induced in the HSF strain with the QT clustering algorithm of GeneSpring. When all of the loci had been placed, 18 clusters had been generated. We then manually varied the diameter from 0.3 to 0.7 to reduce the number of clusters and to be able to add more loci to each cluster while maintaining a well matched pattern within the cluster. The end result was a five-cluster pattern.

To identify Hsf1 binding sites in the promoter sequences, we analyzed 1000 bp upstream of the start codon of the loci that represented verifiable ORFs. Sequences were retrieved from the SGD (36, 37) and analyzed using a simple pattern-identification program. We defined three types of HSEs, each having three nGAAn repeats in Perfect (PFT), GAP (GAP), and STEP (STP) arrangements. The perfect HSE (PFT) consists of three contiguous, inverted repeats of the nGAAn sequence, either nGAAnTTCCnGAAn or nTTCnnGAAmTTCc. The GAP HSE consists of an nGAAn repeat, followed by any 5 bp and 2 inverted nGAAn repeats (nGAAn-(5-bp)-nGAAnTTCc) and its complement (nGAAnmTTCc-(5-bp)-nTTCc), as well as the related sequences nTTCn-(5-bp)-nTTCnnGAAAn and nTTCnnGAAAn-(5-bp)-nGAAn. The STP HSE has a 5-bp insert between each of the 3 nGAAn repeats, yielding the sequences nGAAn-(5-bp)-nGAAn-(5-bp)-nGAAn and nTTCn-(5-bp)-nTTCc-(5-bp)-nTTCc (30). As per previous studies (27, 30), we also allowed a single mismatch (nGAR) in one of the three nGAAn repeats for PFT or GAP.

RESULTS

Both Activation Domains of Hsf1 Contribute to the Transient Heat Shock Response—Yeast Hsf1 has two activation domains that play distinct temporal roles during the heat shock response (12, 18). To compare their relative contributions in the initial response to heat stress, we constructed six isogenic strains, three in an MSN2/4 background and three in an MSN2/4Δ background. In both backgrounds, we tested various truncations of Hsf1: HSF, wild-type Hsf1; HSFΔNAD, which eliminated the first 65 residues of Hsf1; and HSFΔCAD, which eliminated the last 250 residues of Hsf1 (Fig. 1). The six strains were subjected to a 15-min heat shock at a moderate (37 °C) temperature. Northern blot analysis was used to measure the constitutive and heat-induced levels of several HSF-controlled mRNAs. In all six strains, the constitutive levels of HSP104, HSP82, and HSP26 mRNAs were low, whereas the heat-induced levels of the same mRNAs increased to varying degrees (Fig. 2). When we compared the heat induction of HSP104 and HSP26 in the MSN2/4 strain versus the msn2/4Δ strain, there was only a slight increase, whereas HSP82 was significantly reduced in the same background comparison. When we analyzed the contribution of the NAD or the CAD in MSN2/4 strains, the results were similar to those obtained in the HSF strain. The effects of
the deleted activation domains were masked and/or compensated for by the Msn2/4 regulatory system.

On the other hand, in the msn2/4Δ strains, each of the HSPs tested was affected differently by the loss of one of the Hsf1 activation domains (Fig. 2B). In the HSF strain, the heat-induced mRNA levels increased moderately for HSP104 (12-fold) and HSP82 (4-fold) and dramatically for HSP26 (37-fold). In the HSFΔNAD strain, the heat-induced mRNA levels increased moderately for HSP104 (7-fold) and HSP26 (7-fold) and dramatically for HSP82 (34-fold). In the HSFΔCAD strain, the heat-induced mRNA levels increased moderately for HSP104 (5-fold) and HSP82 (6-fold) but had no effect on the levels of HSP26 (0.9-fold). From these experiments we were able to confirm that both regulatory systems cooperate in the expression of many genes. More importantly, we could use the msn2/4 null strains to look specifically at the contribution of the Hsf1 regulatory system.

Global Transcript Profiling of mRNAs Shows a Unique Response to Mutations in Hsf1 during Heat Shock—To understand the broader utilization of Hsf1 and its activation domains during the transient heat shock response, we used microarrays...
to examine genome-wide changes in transcription levels. For each msn2/4Δ strain, we isolated RNA from constitutive and heat-induced conditions in triplicate, resulting in 18 RNA samples. These samples served as templates to generate biotinylated antisense probes for hybridization to Affymetrix yeast genome YG_S98 microarrays. After spot quantitation and normalization, the signal log ratio for each locus on the microarray was determined.

We compared the constitutive and heat shock conditions in each strain and identified loci that responded with significant changes in their transcript level. We then imported the list into the SGD to identify the loci that represented known or putative genes. In the HSF strain, 442 genes were heat-induced. The induction of these genes is significantly higher than would be expected if induction was not correlated with heat shock expression ($p < 1.37 \times 10^{-21}$). Of the 442 genes, 63% were increased by 2- to 3-fold, 28% increased by 3- to 6-fold, and 9% increased by more than 6-fold. The highest expression profile was for the HSP70 homolog SSA4, which increased its levels by 137-fold. In the HSFΔNAD strain, 329 genes were heat-induced ($p = 4.3 \times 10^{-18}$). Of the 329 genes, 63% were increased by 2- to 3-fold, 31% increased by 3- to 6-fold, and 6% increased by more than 6-fold. The highest expression profile was for HSP12, which increased its levels by 95-fold. In the HSFΔCAD strain, 326 genes were heat-induced ($p = 6.1 \times 10^{-18}$). Of the 326 genes, 66% were increased by 2- to 3-fold, 26% increased by 3- to 6-fold, and 8% increased by more than 6-fold. The highest expression profile was not a chaperone, but YFR026C, whose function has not yet been identified. The levels of YRF026C increased by 36.9-fold.

To confirm the reliability of these data, we performed Northern blot hybridizations using selected HSPs that exhibited different levels of dependence of the NAD and CAD in the microarray experiments. In general, the transcript profiles detected by the Northern blot analysis correlated qualitatively with the average values obtained from the microarray experi-
Role of the Activation Domains of Yeast Hsf1

elements (Fig. 3). In addition, the expression changes by the microarray analysis averaged 5- to 8-fold higher than in the Northern blot analysis. This observation, which has been seen by other laboratories (38), may be due to the high sensitivity of the microarray technology. Both experiments show the critical role for the NAD, because the heat-induced levels of HSP42 and SSE2 were dramatically reduced and SSA4 was slightly reduced by the deletion of the NAD, whereas HSP12 was slightly increased. Both experiments also show that the CAD was critical for the heat-induced levels of HSP12 and SSE2 but not critical for SSA4.

As a further test of the validity of our data, we compared our gene list for HSF to previously published data. Other laboratories have compared the heat shock response program between wild-type cells and cells lacking either functional HSF1 or MSN2/4 genes (25–28, 30). In general, genes previously identified as Hsf1-dependent were heat-induced, whereas genes previously identified as Msns2/4-dependent were not heat-induced in our msn2/4Δ strain (25–28, 30). One critical exception was HSP12 (39). To resolve the conflicting data, we used Northern blot analysis to compare the levels of HSP12 in a MSN2/4 wild-type strain (29.5-fold) to an msn2/4Δ strain (20.2-fold) and found that the heat-induced level decreased only 32% (data not shown). These data suggest that, although the transcription level of HSP12 was reduced significantly in the absence of Msn2/4, HSP12 is still heat-induced in an Hsf1-dependent manner (Fig. 3B). HSP12 may be cooperatively controlled by both stress regulatory systems, in a manner similar to HSP26 and HSP104 (19, 29).

To highlight the relationship between the genes in the different strains, we charted the data as a Venn diagram (Fig. 4A). When we compared the HSFΔNAD and the HSF strains, we found that 55% of the genes overlap and seem to be unaffected by the loss of the NAD. The remaining 45% of the genes are not heat-induced in the wild-type Hsf1 strain but gain heat inducibility in the absence of the NAD. When we compared the HSFΔCAD and the HSF strains, we found that only 43% of the genes overlap and seem to be unaffected by the loss of the CAD. The remaining 57% of the genes are not heat-induced in the wild-type Hsf1 strain, but gain heat inducibility in the absence of the CAD.

The majority of the genes that gained heat inducibility in the absence of an activation domain maintained their constitutive levels and increased their heat-induced levels (Fig. 4B). To confirm these data, several genes were tested by Northern blot analysis (Fig. 4C). In all three strains, the constitutive levels of ADH2, IDP2, LSM2, PEP3, and YKR015C were low, whereas the heat-induced levels increased to varying degrees in only the HSFΔNAD and HSFΔCAD strains. In the HSFΔNAD strain, the heat-induced mRNA levels increased moderately for ADH2 (3-fold), SFC1 (4-fold), and IDP2 (3-fold) and dramatically for LSM2 (14-fold). There was very little effect on the heat-induced levels of YKR015C (1.5-fold). In the HSFΔCAD strain, the heat-induced mRNA levels increased moderately for PEP3 (5-fold) and markedly for IDP2 (11-fold) and LSM2 (7-fold). There was no effect on the levels of ADH2 (0.9-fold) and SFC1 (0.8-fold). These results suggest that removal of an activation domain is not globally affecting constitutive activity. Instead, it causes an increase in heat inducibility in select genes. This could be due to the removal of negative-regulating elements that regulate a subset of genes. Alternatively, it could be due to the absence of a repressor normally activated by full-length Hsf1, although the only known heat-induced repressor, XBPI, is induced in all three strains.

For a few genes (4%), the constitutive expression level is increased in the strains with a deleted activation domain. One such case is SFC1, where the constitutive level increased dramatically in the HSFΔCAD strain (Fig. 4C). The heat shock level increased as well, thus negating its heat induction (0.7-fold). We also examined the constitutive levels of all ~5800 putative genes on the microarray and found increases in 41 genes (0.007%) in the HSFΔNAD strain and in 37 genes (0.006%) in the HSFΔCAD strain. Many of these genes from the HSFΔNAD strain are involved in intracellular trafficking (29.3%), whereas many of the genes from the HSFΔCAD strain are involved in metabolism (35.1%). Some of these genes can be seen in Fig. 4B by scanning the constitutive conditions in the HSFΔNAD and HSFΔCAD strains and identifying red outliers. Previous studies have shown that deletion or mutation of Hsf1 can remove intramolecular negative regulation and increase constitutive levels to nearly the same as the induced levels (11, 12, 40–43). Our results confirm that the activation domains are involved in regulating Hsf1 activity for a specific subset of genes.

Identification and Classification of Induced mRNAs in Response to Wild-type HSF during the Heat Shock—As a next step in interpreting the data, each gene that was heat-induced in the HS F strain was annotated by a SGD search and organized into a classification system based on its documented or putative molecular function (36, 37). The functional assignments should be considered dynamic, because many genes have more than one function and could be assigned to alternative groups. We then compared our list to the catalogues of the entire genome as described by the Munich Information Center for Protein Sequence (MIPS) Comparative Yeast Genome Database (44). The number of genes in each category was expressed as a percentage of the total number of genes and plotted on a bar graph (Fig. 5). Only 65% of the 442 genes induced in the wild-type HSF strain have known functions. One of the functional categories that we expected to be highly represented is the chaperone category. It was the most overrepresented of the ten categories described. Seventeen induced genes were categorized as chaperones, including the classic heat shock proteins (HSP12, HSP26, HSP42, HSP78, HSP82, HSP104, SSA3, SSA4, and SSE2). Statistical analysis confirmed that induction of 17 of the 45 chaperones in the genome occurred more frequently than expected by chance (p < 7.9 × 10−7).

The functional analysis also emphasized the vast number of genes that were affected by heat shock and the major reprogramming that the cell undergoes to protect itself from environmental stress. Fig. 5B shows a select few of the 20 genes that were verified using Northern blot analysis. Genes involved in metabolism and energy generation were overrepresented in the HSF gene list compared with the total genome (p < 0.000017 and p < 0.00028, respectively) (Fig. 5). The high percentage of genes involved in energy generation was not surprising, because the increase in their functional activity could provide the
energy needed for the functions of ATP-dependent molecular chaperones. Although other genome-wide studies have shown the importance of energy generation during various environmental stresses (26, 45), our studies emphasize the important role of Hsf1 in regulating their expression during heat shock.

On the other hand, genes involved in transcription or protein synthesis were underrepresented in the HSF gene list \( (p < 6.2 \times 10^{-10} \text{ and } p < 2.1 \times 10^{-8}, \text{ respectively}) \), emphasizing that the cell is in a protective mode. Interestingly, one of the genes involved in transcription and induced in this study was \( XBP1 \), a transcription repressor that binds to the promoters of all three \( G_1 \) cyclins and causes a transient cell cycle delay under various stress conditions (46, 47). Another induced gene was \( HAC1 \), a transcription factor that functions to activate the unfolded protein response in response to stress in the endoplasmic reticulum (48, 49). Although Hac1 requires alternative splicing by the serine/threonine kinase, Ire1, to gain activator properties, many of the newly identified genes up-regulated by the unfolded protein response are also induced in this study (48). In particular, genes involved in protein folding (\( JEM1, SCJ1, \))...
ERO1, EUG1, and MPD1) and protein degradation (DER1, HRD1, and HRD3) were found in both gene lists. The presence of these two proteins, XBP1 and HAC1, suggests that Hsf1 plays an indirect role in ceasing translation of new proteins and helping the cell to tolerate unfolded proteins during the heat shock response.

We compared the percentage of genes in the functional categories in the HSF strain to HSF/H9004NAD and HSF/H9004CAD strains and found there were small changes in the number of genes in specific functional categories. One of the most interesting groups was the transcription and RNA processing category. There were 13 genes (2.9%) in the HSF strain, 14 genes (4.3%) in the HSF/H9004NAD strain, and 20 genes (5.9%) in the HSF/H9004CAD strain. Although there were comparable percentages of genes in the three strains, different genes were heat-induced in each strain. The genes ACA1, HAC1, and PUT3 were heat-induced only in the HSF strain, whereas the genes MIG3 and ZAP1 are heat-induced only in the HSF/H9004CAD strain. The differences in transcription factor expression may account for some of the differences in the three gene lists (Fig. 4A). Hsf1 may act in tandem with these transcription factors to indirectly influence expression in an activation domain-specific manner.

Clustering of Loci Induced by Wild-type Hsf1 Identifies Loci That Are Dependent upon an Activation Domain for Heat Inducibility—As shown in Fig. 4A, deletion of an activation domain caused a complex change in gene expression patterns. To understand the role of the activation domains within the context of full-length Hsf1, we used the loci induced in the wild-type HSF strain as starting material for cluster analysis (Fig. 4B, gray circle). The QT clustering algorithm used in this study grouped together loci with similar patterns of expression over the six conditions (three strains and two temperatures). The end product of the clustering was a five-cluster pattern. The five clusters were named by visually examining the expression profiles revealed by the clustering approach. The first cluster is called INDEPENDENT, because the expression patterns are unaffected by the loss of either of the activation domains (Fig. 6A, green graph). The second cluster is called BOTH, because both activation domains are required for heat induction (Fig. 6B, yellow graph). The third cluster is called NAD, because these loci require the NAD for heat inducibility (Fig. 6C, blue graph). The fourth cluster is called CAD, because these loci require the CAD for heat inducibility (Fig. 6D, pink graph). The fifth cluster consists of loci that were unable to be grouped without compromising cluster diameter and quality of the other clusters (data not shown).

Next we analyzed each of the clusters to determine if loci with similar patterns of expression by QT clustering also had
distinct preferences for functional categories (Fig. 7). Only genes with known functions were analyzed. The number of genes in each functional category was expressed as a percentage of the total number of genes in the cluster. The data revealed that the cluster BOTH was dominated by genes that are involved in cell division and DNA processing (47.6%), and the cluster INDEPENDENT was dominated by genes that are involved in metabolism (32.5%) and energy generation (13.2%) (Fig. 7B). The cluster CAD was dominated by genes that are classified as chaperones (12.2%), metabolism (14.9%), and cell rescue, defense, and degradation (20.3%) (Fig. 7C). The cluster NAD was dominated by genes that are involved in metabolism (28.1%) and intracellular trafficking (15.6%) (Fig. 7D).

On the other hand, the cluster BOTH is devoid of loci classified as chaperones, protein synthesis, and cell rescue, defense, and degradation (Fig. 7A). The cluster CAD contained very few loci that are involved in energy generation (2.7%) (Fig. 7D), whereas the cluster NAD contained very few loci that are involved in transcription and RNA processing (3.1%) or biogenesis and organization of cellular components (3.1%) (Fig. 7C). All of these data strongly indicate that not only do genes with similar patterns of dependence on a particular activation domain have distinct preferences for functional categories, but each activation domain of Hsf1 contributes to the transient response to heat shock.

**Promoter Analysis of Loci Identifies Heat Shock Element Architecture That Is Cluster-specific**—The role of highly conserved HSEs in HSF binding to specific loci has been well established in yeasts and metazoans (30, 50–52). HSEs are composed of inverted, alternating repeats of the 5-bp sequence nGAA, where n is any nucleotide (53–56). The number of pentameric units in an HSE varies, but three to six units are thought to be required for heat regulation in vivo (30, 57). Deviations from the consensus in both sequence and/or the distance between the modules can be tolerated, but to what extent is unknown. To determine if there is a correlation between the different HSEs and the affinity of Hsf1 for the motif and, thereby, the level of transcription activation, we searched in the promoters of the genes for three types of HSEs. Each type contains three nGAA core motifs, but the variants are distinguishable from each other by the location of the core motifs within the HSE (“Experimental Procedures”).

Of the heat-induced genes, only 20.4% of them contained HSEs in their promoters, which is statistically significant in comparison to the entire genome, where only 16.7% of the promoters contained HSEs (p < 0.03). There are two possibilities why all of the promoters of heat-induced genes do not contain HSEs. First, the heat induction measured by microarray analysis may be due to changes in the activity of other transcription factors that are activated by heat shock. For example, 23 genes that do not...
contain HSEs do have binding sites for Hac1, a transcription factor that activates the unfolded protein response. Second, there may be other types of HSEs that are yet to be identified or considered. This possibility has been strengthened by the chromatin immunoprecipitation (ChIP) experiment that showed Hsf1 could bind to promoters lacking a known HSE (27). Furthermore, recent studies on the MDJ1 promoter identified a novel non-consensus HSE that consists of three separated nGAA\textsubscript{n} motifs, nTTC\textsubscript{n}-(11-bp)-nGAA\textsubscript{n}-(5-bp)-nGAA\textsubscript{n} (58). When we analyzed promoters of the genes induced by HSF for this non-consensus HSE, we found 2.9% of the genes contained this novel HSE. More detailed analyses will be required to identify all non-consensus HSEs.

When we limit our focus to those genes that were heat-induced by more than 10-fold, over 50% of them contained HSEs in their promoters. The analysis highlighted a significant correlation between the presence of an HSE of a heat-induced gene and its Hsf1-mediated transcriptional activity ($p < 0.004$). This finding is consistent with the notion that Hsf1 is bound to canonical HSEs, constitutively, ready for the signal to transcribe genes that are important in helping the cell cope with changes in its environment.

We then determined the distribution of HSEs along the promoter within the four clusters and found that HSE binding sites are not uniformly distributed over the promoter region (Table 1). In comparison to the randomized distribution of HSEs seen in the promoters of all the genes in the genome, clusters NAD and CAD had a higher concentration of HSEs distributed close to the ORF ($p < 0.02$). On the other hand, HSEs in the promoters of genes in cluster INDEPENDENT were evenly distributed,
whereas cluster BOTH were slightly further away from the ORF. There is no correlation between clusters and the distribution of HSE types along the promoters. However, we found a correlation between the Hsf1-mediated transcriptional activity and its underlying preference for HSE orientation. Of the 90 heat-induced genes that contain HSEs in their promoters, there is an equal number of HSEs that start with nGAA\textsubscript{n} and nTTC\textsubscript{n}, but genes with a -fold change >5-fold have HSEs with the sequences nTTC\textsubscript{n}nnnGAA\textsubscript{n} or nTTC\textsubscript{n}nnnNNNNnTTC than any combination starting with nGAA\textsubscript{n} (66.7%). These findings suggest that the type of Hsf1 binding site is not as important as the topology of the HSE. Previous studies have revealed that all three DNA-binding domains of the trimeric Hsf1 bind to one face of the DNA (59), and our studies confirm that the orientation of the HSE with respect to the transcriptional start site can affect its transcriptional activity.

**DISCUSSION**

This genome-wide study highlights the importance of the Hsf1 transcription factor in reprogramming cellular events to accommodate changes in temperature. Although yeast cells have two regulatory systems to deal with heat shock, each system has a very specific gene expression program. We used DNA microarrays on strains deleted for the Msn2/4 transcription factors to explore changes in gene expression that are independent of these factors. We then compared data from the msn2/4\textDelta strain with wild-type Hsf1 (HSF) to msn2/4\textDelta isogenic strains that were deleted for one of the activation domains (HSF\DeltaNAD and HSF\DeltaCAD) to understand their role during this critical time.

Our analysis led to three important results. First, we identified a large number of genes that are heat-induced in the absence of the Msn2/4 transcription factors (Fig. 2). The 442 induced genes are involved, to varying degrees, in almost every cellular process (Fig. 6), emphasizing the massive reprogramming that must occur during heat shock. We attribute the identification of such a large number of loci, in part, to improvements in microarray design and reproducibility over the last 5 years. In addition, the number of replicates and the high quality of the data gave us a low false discovery rate, thus broadening the group of genes that are regulated directly or indirectly by Hsf1 during the heat shock response. Although the Northern and microarray data were clean and reproducible, we encountered minor differences caused by cross contamination of the microarray probe sets. One critical example was SSA3, for which our microarray and Northern results did not agree for the HSF\DeltaCAD strain (data not shown). This example emphasizes the importance of using alternative approaches to confirm microarray analysis data for individual genes of interest.

The number of genes identified by our microarray is much larger than the 165 loci identified in the recent Hsf1 ChIP-DNA microarray experiment presented in Hahn et al. (27). When we compared the data, 61 of 165 of genes identified by Hahn et al. were induced by at least 2-fold in our microarray experiments. Conversely, 13.8% of the genes that we identified as induced by at least 2-fold in our microarray experiments were also identified by Hahn et al. The discrepancies between our two experiments may be explained by the different yeast strains used in our respective studies, msn2/4\textDelta for our studies and MSN2/4 for the ChIP experiments. Many of the genes isolated in the ChIP experiment may require Msn2/4 to stabilize or enhance Hsf1 binding and heat inducibility. In addition, the discrepancies may be explained by the different techniques used in our respective studies. ChIP experiments depend on the ability to cross-link a protein to DNA and the ability of the antibody to detect an enrichment of that protein-bound DNA interaction over background (60). This is a useful technique for determining which genes have their promoters tightly bound by Hsf1. As such, the ChIP approach may underestimate the total number of targets if the antibody is not specific or if the interaction between Hsf1 and its target is transient or weak. On the other hand, microarray experiments allowed us to capture the expression profile of all the genes on the array regardless of the transient nature the protein-DNA interaction. Therefore, we were able to identify both direct and indirect targets of Hsf1.

Secondly, we determined that the N- and C-terminal activation domains both play important roles during the transient heat shock response (11, 12). We identified genes that are specifically dependent on the presence of the N- or C-terminal activation domains for heat induction using the QT clustering algorithm (Fig. 6). Several key features in the global expression can be attributed and recognized from the results of the QT clustering. The entire HSF gene list, including the genes with no known or recognized function was the starting material for the clustering. This led to the assigning of hypothetical functions to uncharacterized genes by analyzing well characterized genes in the same cluster (data not shown). Their responses to heat shock provided a clue to their possible function. All of the data can be taken together to propose a physiological picture of the cellular response to each condition tested.

More specifically, we were able to attribute the induction of specific chaperones to one of the two activation domains. 4 of the 17 chaperones (HSP42, JEM1, SCJ1, and SIS1) induced in the wild-type Hsf1 strain depend on the NAD for heat inducibility, whereas 8 of the 17 chaperones (ACA1, HCH1, HSP10, HSP26, HSP82, MDJ1, SSE2, and STI1) depend on the CAD for heat inducibility (Fig. 7). The remaining chaperones have no activation domain specificity but require at least one activation domain for heat inducibility.

The chaperones in cluster NAD include two HSP40 family members and two chaperones localized in the endoplasmic reticulum. The chaperones in the cluster CAD include HSP70 family members, three HSP90 co-chaperones, and two chaperones localized in the mitochondria. The chaperones in cluster INDEPENDENT are also diverse; they include two HSP70 family members, one chaperone that is localized in the mitochondria, and one that is localized in the plasma membrane.

In addition to chaperones, the cluster CAD is dominated by genes that are involved in cell rescue, defense and degradation; both of these functional categories are critical for maintaining protein homeostasis during exposure to stress. Although HSPs suppress the aggregation of many proteins, other proteins are targeted for degradation by the proteasome. Regulation of genes that depend on a specific activation domain is likely to occur through specific interactions with the transcriptional machinery. This possibility is strengthened by the fact that the
Role of the Activation Domains of Yeast Hsf1

TAF-recruiting activity has been mapped to the C-terminal activation domain (61). It is possible that genes that do not depend on the C-terminal activation domain may not require Hsf1 to interact with the TAF-containing form of TATA-binding protein (TBP). As an example, heat-induced expression of SSA4 does not require the C-terminal activation domain (Fig. 3) and does not require TAFs (18, 62).

Until now, the transient heat shock response appeared to be mediated by the N-terminal activation domain, but the targets of the NAD remained a mystery (12, 18). In this study, we identified genes that depend of the NAD for expression. We demonstrated that the cluster NAD is critical for specific chaperones, genes involved in metabolism (28.1%) and intracellular trafficking (15.6%) (Fig. 7C).

Thirdly, we revealed that co-regulated genes have HSEs at a specific distance upstream of the ORF and in a distinct orientation along the promoter (Table 1). The vast majority (82%) of the HSEs found in the promoters of the genes in the cluster NAD lie between 0 and 500.bp upstream of the ORF, far more than would be expected at random (50%) or what we found for all verifiable ORFs (51%). Similar results were seen for the cluster CAD. Many (61.3%) of the HSEs in the promoters of genes in cluster CAD lie between 0 and 500 bp upstream of the ORF (Table 1, column 5).

The promoter analysis also revealed that HSEs found in highly induced genes have an nTTCmC as the first repeat. This fits nicely with the current model of Hsf1-HSE interaction. In order for Hsf1 to make contact with nTTCmGAAm, the DNA-binding domain of two neighboring Hsf1 monomers must come very close together in the major groove of the DNA, whereas, to contact nGAAmnnTTCm, the DNA-binding domain of two monomers must reach across the minor groove into the adjacent major groove (54, 56, 63). Our data suggest that distinct HSE topology may influence the activity of Hsf1 by dictating the protein surfaces available for interactions with other transcription factors. More detailed experiments are needed to test these hypotheses.

These different patterns of HSEs might be related to how Hsf1 uses its two activation domains to accomplish what multiple HSF genes achieve in other organisms. Hsf1 may interact with the transcriptional machinery or other transcription factors to regulate distinct classes of genes. Each activation domain may contribute to the transient heat shock response by recognizing specific promoter architecture, recruiting other proteins, and invoking regulatory mechanisms that contribute to the global gene expression pattern of Hsf1 during the transient heat shock response.

Acknowledgments—We thank Dr. Scott Ferguson for yeast strains, Dr. Don Baldwin and the Penn Microarray Facility for technical support, Dr. John Tobias for advice on analysis of microarray data, and Dr. Brian Brunk for assistance with the pattern identification program. We also thank Sara Achenbach, Laura Conlin, Scott Ferguson, and John Tobias for helpful discussions and critical readings of the manuscript.

REFERENCES

1. Morimoto, R. I., Kline, M. P., Bimston, D. N., and Cotto, J. J. (1997) Essays Biochem. 32, 17–29
2. Sorger, P. K., and Pelham, H. R. (1988) Cell 54, 855–864
3. Wu, C. (1992) Annu. Rev. Cell Dev. Biol. 11, 441–469
4. Estruch, F., and Carlsson, M. (1993) Mol. Cell. Biol. 13, 3872–3881
5. Kobayashi, N., and McEntee, K. (1993) Mol. Cell. Biol. 13, 248–256
6. Harrison, C. J., Bohm, A. A., and Nelson, H. C. M. (1994) Science 263, 224–227
7. Littlefield, O., and Nelson, H. C. M. (1999) Nat. Struct. Biol. 6, 464–470
8. Peteranderl, R., and Nelson, H. C. M. (1992) Biochemistry 31, 12272–12276
9. Peteranderl, R., Rabenstein, M., Shin, Y. K., Liu, C. W., Wemmer, D. E., King, D. S., and Nelson, H. C. M. (1999) Biochemistry 38, 3559–3569
10. Sorger, P. K., and Nelson, H. C. M. (1989) Cell 59, 807–813
11. Nieto-Sotelo, J., Wiederrecht, G., Okuda, A., and Parker, C. S. (1990) Cell 62, 807–817
12. Sorger, P. K. (1990) Cell 62, 793–805
13. Gallo, G. J., Prentice, H., and Kingston, R. E. (1993) Mol. Cell. Biol. 13, 749–761
14. Wiederrecht, G., Seto, D., and Parker, C. S. (1988) Cell 54, 841–853
15. Jedlicka, P., Mortin, M. A., and Wu, C. (1997) EMBO J. 16, 2452–2462
16. Pirkkala, L., Nykanen, P., and Sistonen, L. (2001) FASEB J. 15, 1118–1131
17. Baniwal, S. K., Bharti, K., Chan, K. Y., Fauth, M., Ganguli, A., Kotak, S., Mishra, S. K., Nover, L., Port, M., Scharf, K. D., Tripp, J., Weber, C., Zielinski, D., and von Koskull-Doring, P. (2004) J. Biosci. 29, 471–487
18. Young, M. R., and Craig, E. A. (1993) Mol. Cell. Biol. 13, 5637–5646
19. Amoros, M., and Estruch, F. (2001) Mol. Microbiol. 39, 1523–1532
20. Santoro, N., Johansson, N., and Thiele, D. J. (1998) Mol. Cell. Biol. 18, 6340–6352
21. Tamai, K. T., Liu, X., Silar, P., Sosinowski, T., and Thiele, D. J. (1994) Mol. Cell. Biol. 14, 8155–8165
22. Estruch, F. (2000) FEMS Microbiol. Rev. 24, 469–486
23. Martinez-Pastor, M. T., Marchler, G., Schuller, C., Marchler-Bauer, A., Rius, H., and Estruch, F. (1996) EMBO J. 15, 2227–2235
24. Gorner, W., Durchschlag, E., Martinez-Pastor, M. T., Estruch, F., Ammerger, G., Hamilton, B., Rius, H., and Schuller, C. (1998) Genes Dev. 12, 586–597
25. Boy-Marcotte, E., Lagniel, G., Perrot, M., Bussereau, F., Boudsocq, A., Jacquet, M., and Labarre, J. (1999) Mol. Microbiol. 33, 274–283
26. Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., and Brown, P. O. (2000) Mol. Cell. Biol. 21, 4241–4257
27. Hahn, J. S., Hu, Z., Thiele, D. J., and Iyer, V. R. (2004) Mol. Cell. Biol. 24, 5249–5256
28. Treger, J. M., Schmitt, A. P., Simon, J. R., and McEntee, K. (1998) J. Biol. Chem. 273, 26875–26879
29. Grably, M. R., Stanhill, A., Tell, O., and Engelberg, D. (2002) Mol. Microbiol. 44, 21–35
30. Yamamoto, A., Mizukami, Y., and Sakurai, H. (2005) J. Biol. Chem. 280, 11911–11919
31. Ferguson, S. B., Anderson, E. S., Harshaw, R. B., Thate, T., Craig, N. L., and Nelson, H. C. (2005) Genes Dev. 19, 1203–1214
32. Ferguson, S. B. (2005) Negative Regulation of the Heat Shock Transcription Factor by Protein Kinase A in Saccharomyces cerevisiae, Ph.D. dissertation, Cell and Molecular Biology, University of Pennsylvania, Philadelphia, PA
33. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (2004) Current Protocols in Molecular Biology, pp. 13.12.1–13.12.5, Greene Publishing Associates, Inc., and John Wiley & Sons, Inc.
34. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 5116–5121
35. Heyer, L. J., Kruglyak, S., and Yooseph, S. (1999) Genome Res. 9, 1106–1115
36. Ball, C. A., Dolinski, K., Dwight, S. S., Harris, M. A., Issel-Tarver, L., Kasarskis, A., Scafe, C. R., Sherlock, G., Binkley, G., Jin, H., Kaloper, M., Orr, S. D., Schroeder, M., Weng, S., Zhu, Y., Botstein, D., and Cherry, J. M.
