Human HSP70 promoter contains at least two distinct regulatory domains

(heat shock induction/cadmium induction/serum stimulation/DNA-mediated gene transfer)

BARBARA J. WU*, ROBERT E. KINGSTON††, AND RICHARD I. MORIMOTO*

*Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, IL 60201; †Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114; and ‡Department of Genetics, Harvard Medical School, Boston, MA 02115

Communicated by Irving M. Klotz, September 26, 1985

ABSTRACT The expression of the human HSP70 gene is induced by a wide range of physiological stresses, including exposure to heat shock and heavy metals, or under nonstress conditions, such as in response to serum stimulation. We have previously demonstrated that in either case the regulated expression is at the primary level of transcription. To determine whether transcription is mediated through a single or multiple genetic elements, we have dissected the sequences upstream of the transcription start site of the human HSP70 gene by constructing chimeric genes retaining variable amounts of 5' flanking regions fused to the bacterial gene encoding chloramphenical acetyltransferase. Transcription from the chimeric genes was determined by S1 nuclease analysis of separate stable transfectants. The sequences required for heat shock and cadmium induction lie between −107 and −68. Within this region is the sequence CTGGAGATAT-TCCCG, which is identical in 12/14 positions with the heat shock element of Drosophila heat shock genes, and a separate sequence, CGNCCCCGG, which is homologous to the core of the human metallothionein II metal-responsive element. The sequences required for serum-stimulated transcription are distinct from the heat shock element. The sequence CCAAT at −68 is required for high levels of correctly initiated transcripts, and a purine-rich sequence, GAAGGAAAAAAT, at −58 is required for serum stimulation. The human HSP70 promoter contains at least two regulatory domains—a distal domain responsive to heat shock or cadmium and a proximal domain responsive to stimulation by serum.

Sequences defining eukaryotic promoters can be separated into three functional classes. The first class is enhancer elements: sequences that function independent of location or orientation and, in some cases, confer tissue specificity (1). The second class is elements that confer the basal activity of promoters and determine the start site of initiation; this class includes the “TATA box” (2–4) and the “CAAT homology” (5–9). These elements generally function in close proximity to the transcriptional start site. It is not clear whether these elements can play a role in regulation of transcription. A third class of elements, regulatory elements, confers inducible regulation in response to a specific stimulus.

Inducible promoters, such as those regulating metallothionein, mouse mammary tumor virus, interferon, and heat shock genes, have been particularly useful in studying rapid changes in transcription. Studies localizing regulatory elements responsive to specific inducers have identified the heat shock elements (HSE) for heat shock genes (10–13), metal-responsive elements (MRE) for metallothionein genes (14, 15), glucocorticoid-responsive elements (GRE) for mouse mammary tumor virus genes (16–18), and interferon-regulatory elements (IRE) for the viral or poly(dI-dC) induction of α- and β-interferon genes (19–21). One class of inducible genes with complex regulation includes the heat shock or stress-induced genes. The expression of heat shock genes is induced by a wide range of stimuli, including heat shock, heavy metals, cessation of anoxia, inhibitors of energy metabolism, amino acid analogues, infection by adenovirus 5 or simian virus 40 (SV40), stimulation with serum, and development (22–25).

We have dissected the sequences upstream of the transcription start site of the human HSP70 gene (which encodes a 70-kDa heat shock protein) by constructing chimeric genes retaining variable amounts of 5' flanking regions. We measured the transcriptional activity of the heterologous gene in response to heat shock, cadmium sulfate, and serum stimulation in stably transformed human cell lines. Sequences responsive to heat shock and cadmium induction are distinct from those responsive to serum stimulated expression, yet both domains reside within 112 nucleotides of the transcription initiation site of the human HSP70 gene.

MATERIALS AND METHODS

Cells and Transfection. Cell line 293 (26) was maintained in Dulbecco's modified Eagle's (DME) medium supplemented with 5% fetal calf serum. Calcium phosphate-mediated DNA cotransfections with pSV2-ecogpt and subsequent selection in DME medium containing mycophenolic acid were performed as described (25).

Plasmids and Construction of Deletion Mutants. Plasmid pH2.3 contains the entire coding region of the human HSP70 gene (27). The Escherichia coli xanthine/guanine phosphoribosyltransferase gene fused to the SV40 promoter is contained in pSV2-ecogpt (28). The chimeric pHBCAT (27) contains 2.4 kilobases (kb) of upstream and 150 base pairs (bp) of nontranslated human HSP70 sequences fused to the bacterial chloramphenical acetyltransferase (CAT) gene (29).

The starting material for the deletion series was plasmid pHSm-CAT, which was made from plasmid pHBcat by removal of the 1.3-kb Xba I/Sma I fragment (Fig. 1). Plasmid PHSmCAT was cleaved with HindIII and digested for various amounts of time with BAL-31 (International Biotechnologies, New Haven, CT). HindIII linkers were ligated and the plasmid was digested with HindIII and Xma I. The DNA fragment containing the promoter start site, the CAT gene, and half of the ampicillin gene was isolated and ligated to a Xma I/HindIII fragment of pHBCAT containing the other half of the ampicillin gene and the origin of replication. The deletion endpoint of the resulting plasmids was determined by the dideoxy method of sequencing (30), using a 17-nucleotide phage M13 reverse primer. Plasmids are

Abbreviations: SV40, simian virus 40; CAT, chloramphenicol acetyltransferase; kb, kilobase(s); bp, base pair(s); HSE, heat shock element; MRE, metal-responsive element.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
The position of the 5' deletion endpoints generated with the exonuclease BAL-31 are shown in Fig. 2. In each recombinant gene, the upstream sequences are reconstituted with a HindIII linker to the bacterial vector such that the flanking bacterial sequences remain constant. The deletion mutants differ only in the amount of 5' upstream sequences fused to the CAT gene. Each class of transformants is named according to the amount of upstream sequences fused to the CAT gene. For example, 1A150 refers to a clonal cell line that contains stable copies of the chimera with 150 nucleotides upstream of the transcription start site fused to the CAT gene.

An example of the specificity of RNA analysis in stable transformants is shown in Fig. 3 Left. Cytoplasmic RNA was isolated from cell line 293 or from 1A150 cells and used for S1 nuclease protection. The 32P-labeled templates were selected for their specificity in a solution hybridization reaction, either for the endogenous HSP70 gene or for the heterologous CAT gene (Fig. 3 Right). The levels of endogenous HSP70 mRNA were examined by using a template 3' labeled at the Ava I site of pH2.3, a plasmid that contains the entire HSP70 coding sequences (Fig. 3 Right). RNA isolated from control or heat-shocked cell line 293 or 1A150 cells protects a 350-nucleotide fragment that increases 4-fold after heat shock (Fig. 3 Left). The heterologous CAT mRNA were detected by using a template 5' labeled at the EcoRI site in the CAT coding sequences that results in a 400-nucleotide protected fragment. The CAT template was not protected by RNA isolated from cell line 293. This result is expected because eukaryotic cells do not contain the bacterial CAT gene. However, RNA isolated from 1A150 cells protects a 400-nucleotide fragment that increases 10-fold after heat shock. The template used to detect CAT mRNA extends through the transcription initiation site; therefore it can distinguish between correctly initiated transcripts and aberrant initiation products.

We initially compared the level of CAT mRNA transcribed from the parental pHB-CAT, which contains approximately 2.4 kb of upstream sequence, to the level of CAT mRNA in deletion mutants that retained 292 and 188 nucleotides of upstream sequence. The levels of CAT mRNA in cells maintained at the normal growth temperature (basal), subject to heat shock, or stimulated with serum were similar in Δ292 and Δ188 cell lines to the level observed for chimeras with 2.4 kb of upstream sequence (data not shown). For this reason only the data for more proximal deletions are presented.

**Identification of Sequences Required for Heat Shock and Cadmium Induction.** The kinetics of induction after incubation at elevated temperature or in the presence of cadmium sulfate was examined by using gene-specific templates and S1 nuclease analysis. The conditions of heat shock (60 min at 43°C) or metal induction (25 μM cadmium sulfate for 2 hr)
resulted in a 2- to 4-fold induction of the HSP70 transcript in cell line 293 (data not shown; ref. 27). Using these conditions, we analyzed the level of CAT mRNA transcribed in two to six independent stable transformants for each deletion mutant (Table 1). A typical example of the results obtained from each class of deletion mutants is shown in Fig. 4. The level of CAT mRNA after heat shock is 4- to 10-fold induced in cell lines containing at least 107 nucleotides upstream of the 5′ terminus (Δ292, Δ188, Δ150, Δ132, Δ112, and Δ107). The heat shock response observed in Δ107 cells is absent from Δ68 cells, thus localizing the sequences required for heat shock activation. Within these boundaries is the sequence CTG-GAATATTCCCG, which is identical in 12/14 positions with the consensus HSE located upstream of Drosophila heat shock genes (12). CAT mRNA is not induced by heat shock in Δ68, Δ58, and Δ47 cell lines. With these more extensive deletions, we have consistently observed that the level of CAT transcripts in cells induced by heat shock or cadmium is lower than control levels, suggesting that the RNA synthesized before induction is unstable after the stress. In these proximal deletions, we detect CAT transcripts that are initiated from within the plasmid vector sequence that are not induced after cadmium treatment or heat shock.

Table 1. Transcriptional regulation by deletion mutants

<table>
<thead>
<tr>
<th>Deletion endpoint,* bp</th>
<th>No. cell lines examined</th>
<th>Basal activity†</th>
<th>Heat shock</th>
<th>Cd2*</th>
<th>Serum</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2400</td>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-292</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-188</td>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-150</td>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-131</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-112</td>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-107</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-68</td>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-58</td>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-47</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Deletion endpoints correspond to nucleotides upstream of the HSP70 initiation site.
†Presence (+) or absence (−) of correctly initiated CAT transcripts. Induction (+) or absence (−) of correctly initiated CAT transcripts. Induction by heat shock, cadmium sulfate, or serum stimulation relative to nontreated cells is indicated by +. Lack of induction is indicated by −.

Fig. 4. Heat shock and cadmium-induced expression of the chimeric CAT gene in line 293 cells transfected with each deletion mutant. Cytoplasmic RNAs were prepared from cells maintained at 37°C (C), heat shocked for 1 hr at 43°C (HS), or incubated in 25 μM cadmium sulfate for 120 min (M). S1 nuclease protection of Tcat was performed as described. A and B are 12-hr exposures; C is a 72-hr exposure.
analyzed the level of CAT mRNA in cells serum starved for 48 hr and stimulated by the addition of serum to a concentration of 20% for 12 hr. CAT mRNA is induced by serum stimulation in cell lines containing at least 58 nucleotides of upstream sequences but is not in cell lines retaining 47 nucleotides (Fig. 6). Within this segment is the purine-rich sequence GAAGGGAAAG. We note that the induction of CAT transcripts in Δ150, Δ132, and Δ112 cells is lower than in cells containing chimeric genes with more (pHBCAT, Δ292, Δ188) or less (Δ107, Δ68, Δ58) upstream sequences. We do not have an explanation for the weak stimulation of Δ150, Δ132, and Δ112 cells.

We had previously suggested that serum-stimulated expression of the 12 HSP70 gene corresponds to the basal activity in asynchronous populations of cells. Upon examination of the basal levels of CAT mRNA in the deletion mutants, we find that correctly initiated transcripts are detected in Δ58 cells but not in Δ47 cells (Fig. 7). Thus, we conclude that sequences required for basal activity and serum-stimulated transcription are located within the boundaries of −58 to −47, corresponding to the purine-rich sequence GAAGGGAAAG. However, wild-type levels of basal and serum-stimulated CAT mRNAs also require sequences between −68 and −58, because the level of CAT mRNA in serum-stimulated Δ68 cells is typically 10- to 20-fold greater than in serum-stimulated Δ58 cells. This segment contains the sequence GATTGG, which corresponds to the CAAT homology in inverted orientation. The sequences between −68 to −47 may correspond to transcription elements that function in concert to give wild-type expression levels.

**DISCUSSION**

We have localized the nucleotide sequences required for heat shock-, cadmium-, and serum-stimulated transcription of the human HSP70 gene. The major conclusion from our study is that the human HSP70 gene can be regulated by at least two distinct domains, one that contains sequences responsive to heat shock and cadmium and a second domain necessary for transcription stimulated by serum. This offers a means of regulating HSP70 expression independent from environmental stresses. We base our conclusion on the following: (i) the consistency of the results obtained in independent stable transformants containing integrated copies of the chimeric CAT gene and (ii) the ability to detect low levels of mRNAs afforded by the sensitivity of SI nuclease analysis. These considerations also circumvent potential anomalies that can arise from analysis of gene expression by using transient transfection assays (33) and allow us to examine correctly initiated transcripts. One potential concern is that the recipient of the chimeric genes, cell line 293, contains integrated copies of the adenovirus pre-early region and expresses the adenovirus EIA gene. It is well documented that EIA affects transcription of viral and cellular genes (34-38).

Data from numerous laboratories have shown that adenovirus EIA does not alter promoter sequence requirements (38). The promoter sequences required for optimal basal transcription are also required for optimal EIA-stimulated expression. However, these studies do not reveal whether EIA alters sequence requirements of inducible promoters.

The sequences required for heat shock activation of Drosophila heat shock genes have been characterized by similar analyses. The HSE functions efficiently upstream of heterologous genes when reintroduced into Drosophila, Xenopus, murine, and human cells (12, 39-41). Sequences with homology to HSEs are also located upstream of Xenopus and soybean heat shock genes (42, 43). Located 100 nucleotides upstream of the human HSP70 gene is the 14-nucleotide sequence CTGGAATATTCGCC, which shares identity in 12/14 positions with the consensus sequence of the Drosophila heat shock elements, CTGGAATATTCGCC. Deletion of this sequence from the human HSP70 promoter eliminates responsiveness to heat shock. Heat shock activation of the CAT gene in human cells requires a single HSE. In contrast, the Drosophila heat shock genes have closely linked HSEs between −90 to −60 that are necessary for inducible transcription (44, 45). It is not known why the Drosophila heat shock genes have multiple closely linked HSEs. Even before heat shock, the regions of chromatin corresponding to HSEs are hypersensitive to DNase I (46). Changes in chromatin configuration correlate with the appearance of proteins such as the heat shock-activated
protein (HAP) that binds to the HSE after heat shock. It is suggested that HAP is a positive activator that induces transcription of the HSP70 gene (47).

Our deletion analysis does not separate the sequences required for cadmium induction from those required for heat shock induction. We have compared the sequences within the boundaries required for cadmium induction of the HSP70 gene to the MRE of human metallothionein II gene and have noted a 7-nucleotide identity between HSP70 and metallothionein MRE. Because HSEs are not found in the upstream regions of metallothionein genes, we suggest that heat shock activation and cadmium activation may act through distinct sequences.

Sequences necessary for basal activity and serum-stimulated transcription of the human HSP70 gene in cell line 293 are positioned between −68 and −47. Within these boundaries are two sequence elements, a CCAAT box on the minus strand, corresponding to GATTGG at −68, and a purine-rich sequence, AGAAGGGAAAAGG. The CAAT homology has been localized in the −70 region for a large number of genes, including β-globin, thymidine kinase of herpes simplex virus (HSV-TK), and dihydrofolate reductase. Removal of the CCAAT element reduces the transcriptional efficiency of the human HSP70 gene by at least one order of magnitude. A similar effect of the CCAAT homology has been observed for the HSV-TK (9) and β-globin (34) promoters. It is still possible to detect correctly initiated HSP70 transcripts when the CCAAT homology is deleted. In contrast, no correctly initiated transcripts were detected when the deletions removed the purine-rich sequence. Thus the purine-rich sequence appears to be a necessary element for function of the HSP70 promoter. We note that the purine-rich sequence required for serum stimulation of the human HSP70 gene bears a striking resemblance to the interferon-regulatory element identified for viral or poly(dI-dC) induction of α- and β-interferon genes (19–21). If this homology is relevant, it suggests either that the purine-rich region is shared among a subset of inducible genes or that the inducers of the interferon gene are also inducers of the human HSP70 gene.

In addition to heat shock, cadmium, and serum, many other stimuli also induce the HSP70 gene. The compactness of the genetic information in the HSP70 upstream region prevents at least two interesting possibilities: either each stimulus activates distinct transcription factors that recognize separate promoter elements or the various stimuli will sort into classes and activate the HSP70 promoter through limited pathways.

We thank M. Bienz and C. Wu for their comments on the manuscript. These studies were supported by grants from the Leukemia Foundation, Inc. and the National Institutes of Health. B.J.W. was supported by a fellowship from the Leukemia Society of America.