Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 6911–6915, August 1991
Biochemistry

Isolation of a cDNA for HSF2: Evidence for two heat shock factor genes in humans

(human transcription factor)

THOMAS J. SCHUETZ*†, GREGORY J. GALLO*†, LYNN SHELDON*†, PAUL TEMPST†, AND ROBERT E. KINGSTON*†‡

*Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114; and †Department of Genetics, Harvard Medical School, Boston, MA 02115

Communicated by Phillip A. Sharp, April 30, 1991

ABSTRACT **The heat shock response is transcriptionally regulated by an evolutionarily conserved protein termed heat shock factor (HSF). We report the purification to homogeneity and the partial peptide sequence of HSF from HeLa cells. The peptide sequence was used to isolate a human cDNA with a predicted open reading frame that has homology to the DNA binding domains of both *Saccharomyces cerevisiae* and *Drosophila* HSFs. The cDNA directs the synthesis of a protein that binds to the heat shock element with specificity identical to HeLa HSF and stimulates transcription from a heat shock promoter. The expressed protein cross-reacts with anti-HSF antibodies. Surprisingly, however, this cDNA does not encode all of the peptides obtained from purified HeLa HSF. These peptides are encoded by a distinct human cDNA, *HSF1*, described by Rabindran *et al.* [Rabindran, S. K., Giorgi, G., Clos, J. & Wu, C. (1991) *Proc. Natl. Acad. Sci. USA* 88, 6906–6910.] It therefore appears that there is a human heat shock factor gene family and that at least two separate but related HSF proteins regulate the stress response in humans.**

When any organism is stressed by an increase in temperature, a set of protective proteins is synthesized that presumably functions to maintain intracellular homeostasis. These proteins are referred to as the heat shock or stress-induced proteins, and their synthesis is regulated at many levels including transcription, RNA processing, and translation (1). Transcription of the heat shock genes is rapidly induced after temperature stress; this response is mediated by a promoter element, the heat shock element (HSE), which is constructed of three inverted repeats of the sequence NGAAN (2–4). The HSE binds a multimeric transcriptional stimulatory protein termed heat shock factor (HSF) (5). The gene for HSF has been isolated from *Saccharomyces cerevisiae* (6–10), *Drosophila* (11), and tomato (12). Since the HSE is absolutely conserved among all eukaryotes, it was anticipated that there would be extensive regions of homology between *HSF* genes from separate species. The predicted amino acid sequences of HSF from *S. cerevisiae* and *Drosophila* show surprisingly little homology, however, with the conspicuous exception of regions involved in DNA binding and protein multimerization (9–11).

The heat shock response is induced by numerous and varied stresses, including temperature stress, ethanol, heavy metals, and changes in oxygen tension (1). To understand the mechanism of the stress response, it is critical to determine how many different factors regulate this response. Recent data indicate that there are at least three *HSF*-related genes in tomato (12). We report here partial peptide sequence of HeLa HSF and the isolation of a cDNA for a human HSF-related protein.§ These data, in conjunction with the isolation of a separate human *HSF* cDNA by Rabindran *et al.* (13), indicate that there are at least two related *HSF* genes in humans.

MATERIALS AND METHODS

Purification of HeLa HSF. Approximately 1.8×10^{10} S3 HeLa cells (30 liters of culture) growing in Joklik's MEM (with 5% horse serum) were heat shocked at 43°C for 1 hr and nuclear extracts were prepared (14). HSF was purified essentially as described (15) with the following differences: Nonidet P-40 was 0.1% in all buffers, *n*-octyl glucoside was 5 mM in load buffer B, and affinity columns were washed with load buffer B containing 0.8 M KCl and eluted with load buffer B adjusted to 2.0 M KCl and 0.1 M $MgCl_2$.

Isolation of an HSF cDNA. Approximately 8 μg of HSF (150 μl) was concentrated and the buffer was exchanged for water with a Centricon 30 microconcentrater (Amicon). The sample was electrophoresed, transferred to nitrocellulose, stained with Ponceau S, excised from the blot, and digested with trypsin; fragments were isolated and sequenced as described (16–18). The sequenced peptides were ESEPAPASVTALTDAR (2T6), VVHIEQGGLVKPE (2T5), TEFQHPCFLR (2T8), HENEALWR (2T4), and LNMYGFR (2T10). Approximately 3×10^5 members of a cDNA library from the human T-cell line HPB-ALL (19) were screened with a 20-base degenerate oligonucleotide containing all possible codons for the sequence EFQHPCF (except for the last base in the F codon). The filters were washed under conditions predicted to allow for two mismatches (3 M TMAC/50 mM Tris·HCl, pH 8.0/1 mM EDTA at 50°C). Ten colonies that rescreened as positive with the 20-mer were screened with a 17-base degenerate oligonucleotide containing all possible codons for the sequence NMYGFR (the last base in the R codon is excluded). Two colonies hybridized to this probe, each of which contained an insert encoding the cDNA partially depicted in Fig. 2.

***In Vitro* Transcription.** Transcription reactions were performed as described with "G-less" cassette vectors (20). Nuclear extract from non-heat shocked HeLa cells provided general transcription factors (9 μl in Fig. 1; 8 μl in Fig. 5). All reaction mixtures contained two templates: the adenovirus major late promoter as an internal control (400 ng in Fig. 1; 67 ng in Fig. 5), and an HSP70 promoter either deleted to −34 (no HSE) or containing an extended HSE at −34 (400 ng in Fig. 1; 333 ng in Fig. 5) (21, 22). Reaction mixtures included 1 mM $MgCl_2$ for 60 min preceding transcription (1.8 mM $MgCl_2$ for 60 min).

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.

Abbreviations: HSF, heat shock factor; HSE, heat shock element.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M65217).

***In Vitro* Analysis and Overexpression of the cDNA Product.** A 10-base-pair *Cla* I linker was added to the *Hin*cII site of HSF2 and inserted into pBS-ATG (23). RNA was generated from this vector with T3 polymerase (Stratagene) and RNA (40 μg/ml) was translated in rabbit reticulocyte lysate (Promega) according to the manufacturer's instructions. For bacterial overexpression, HSF2 was cloned into pET-3d, expressed in *Escherichia coli* strain BL21, and induced as described (13, 24). Bacteria were lysed in sample buffer and protein was separated by SDS/PAGE. Induced protein was purified by electroelution as described (25), acetone precipitated, denatured in 6 M guanidine hydrochloride, and dialyzed into load buffer B without $MgCl_2$.

Antibody Production. CAF1/J (The Jackson Laboratory) female mice were immunized (four injections at 2 μg each of purified HSF) with an intraperitoneal injection protocol (26). Control antibodies were serum from an unimmunized mouse, ascites A from an unproductive immunization with peptide 2T6 conjugated to keyhole limpet hemocyanin, and ascites B from an immunization with an unrelated protein.

RESULTS AND DISCUSSION

Purification and Cloning of HeLa HSF. Our strategy for isolating a cDNA encoding human HSF depended on the isolation and sequencing of tryptic peptide fragments from the purified HSF protein. We purified human HSF from nuclear extracts of heat shocked HeLa cells. Analysis of the purified fractions on SDS/PAGE followed by silver staining identified an 87-kDa protein (Fig. 1), from which HSF gel shift activity could be eluted and renatured (ref. 27; T.J.S., unpublished data). HSF activity was purified 2000-fold with a 30% overall activity yield. Purified fractions containing this protein stimulated transcription *in vitro* from an HSP70 promoter, and the transcriptional stimulatory activity was dependent on both the protein and an intact HSE (Fig. 1). The DNA binding and transcriptional stimulatory activities copurified (T.J.S., unpublished data).

Five tryptic peptides obtained from purified HSF were sequenced, degenerate oligonucleotides were synthesized, and a human T-cell cDNA library was screened. We isolated two clones (representing one cDNA), each of which hybridized with two separate degenerate oligonucleotides. A 2411-base-pair cDNA was sequenced on both strands. A 536-amino acid predicted open reading frame was identified (Fig. 2). Northern blot analysis of HeLa mRNA was performed with either the complete cDNA or a small region corresponding to nucleotides 721–1121 as probes. In both cases, one message of ≈2.7 kilobases was identified (T.J.S., unpublished data). While these data indicate that the cloned cDNA is approximately full length, we cannot definitively conclude that we have identified the initiating ATG.

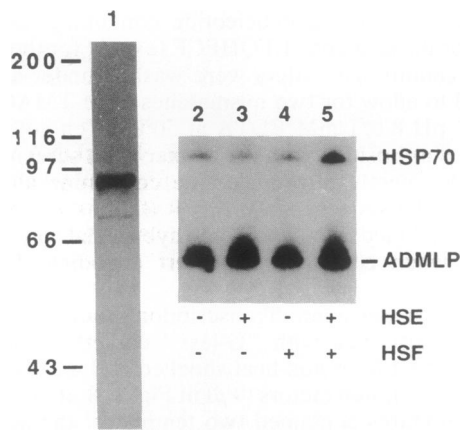

FIG. 1. Purification of and transcriptional stimulation by human HSF. Electrophoresis of 250 ng of a purified HSF fraction followed by silver staining is shown in lane 1. Migration of molecular size standards (kDa) is indicated. *In vitro* transcription data are shown in lanes 2–5. The presence or absence of the HSE in the human HSP70 promoter construct and the inclusion or exclusion of purified HeLa HSF (0.5 μg) are as indicated. HSP70 transcription was stimulated 5- to 10-fold by HSF in this and repeat experiments. Transcripts from the test promoter and from the internal reference promoter are indicated by HSP70 and ADMLP (adenovirus major late promoter), respectively.

Homology with Other HSFs. The predicted amino acid sequence of this cDNA bears strong sequence homology to both the *S. cerevisiae* and *Drosophila* HSFs in the regions of those molecules that are required for DNA binding and multimerization. A comparison of amino acids 2–119 of the cDNA that we have isolated to amino acids 167–284 of the *S. cerevisiae* HSF open reading frame (the region identified as the DNA binding domain; ref. 9) reveals 41% amino acid identity. A 26-amino acid segment within this region (*S. cerevisiae*, amino acids 213–238; human cDNA, amino acids 48–73) contains 85% identity. The DNA binding domain is also 53% identical to the DNA binding domain of *Drosophila* HSF and 43–46% identical to the corresponding regions of the tomato HSF genes (11, 12).

***In Vitro* Analysis of the Protein Encoded by the cDNA.** *In vitro* transcription and translation of the isolated cDNA produced a protein with an apparent mobility of 79 kDa (T.J.S., unpublished data). This protein binds specifically to the HSE as determined by mobility shift gel electrophoresis (see below). To characterize the specificity of this binding, and to compare that specificity with that of purified HeLa HSF, we performed methylation interference analysis. The interference pattern obtained with the protein expressed from the cDNA is virtually identical to that of the protein purified from HeLa cells (Fig. 3*A*; see also refs. 28 and 30). Methylation of G residues that are part of the GAA repeats in the HSE interferes with binding of both proteins, as does methylation of two G residues immediately adjacent to the GAA repeats (see Fig. 3*B*). Methylation of two G residues in the noncoding strand of the imperfect promoter-proximal repeat enhances binding of both proteins (Fig. 3*B*, open arrowheads). These data demonstrate that the isolated cDNA encodes a protein that binds to the HSE with essentially identical specificity to that seen with HeLa HSF.

Immunological Analysis of HeLa HSF and the cDNA Product. The protein encoded by this cDNA also shares antigenic epitopes with HSF purified from heat shocked HeLa cells. We raised polyclonal immune ascites to human HSF by injecting mice with the purified protein displayed in Fig. 1. These antibodies interact with purified HSF to cause a decreased mobility of the HSF–HSE complex on a mobility shift gel (Fig. 4*A*, compare lanes 2 and 4). Also, on a Western blot, the antibodies identify a protein in nuclear extracts from heat shocked HeLa cells that comigrates with pure HSF (lanes 7 and 9). Human HSF has previously been shown to fractionate in the cytoplasmic extract from nonshocked cells and in the nuclear extract from heat shocked cells (28, 30). The factor that is recognized by the antibodies is not present in nuclear extracts from non-heat shocked HeLa cells and displays increased mobility in cytoplasmic extracts from these cells (lanes 6 and 8). The difference in mobility between HSF from nonshocked and shocked cells was previously inferred from UV-crosslinking studies and appears to be due to phosphorylation of HSF following heat shock (30). Hence, the antibodies specifically recognize a protein with the properties of human HSF.

To determine whether these antibodies interact with the protein encoded by the cDNA, mobility gel shift analysis was used. Binding of the *in vitro* transcribed/translated protein to the HSE resulted in a shifted complex in this assay (Fig. 4*B*,

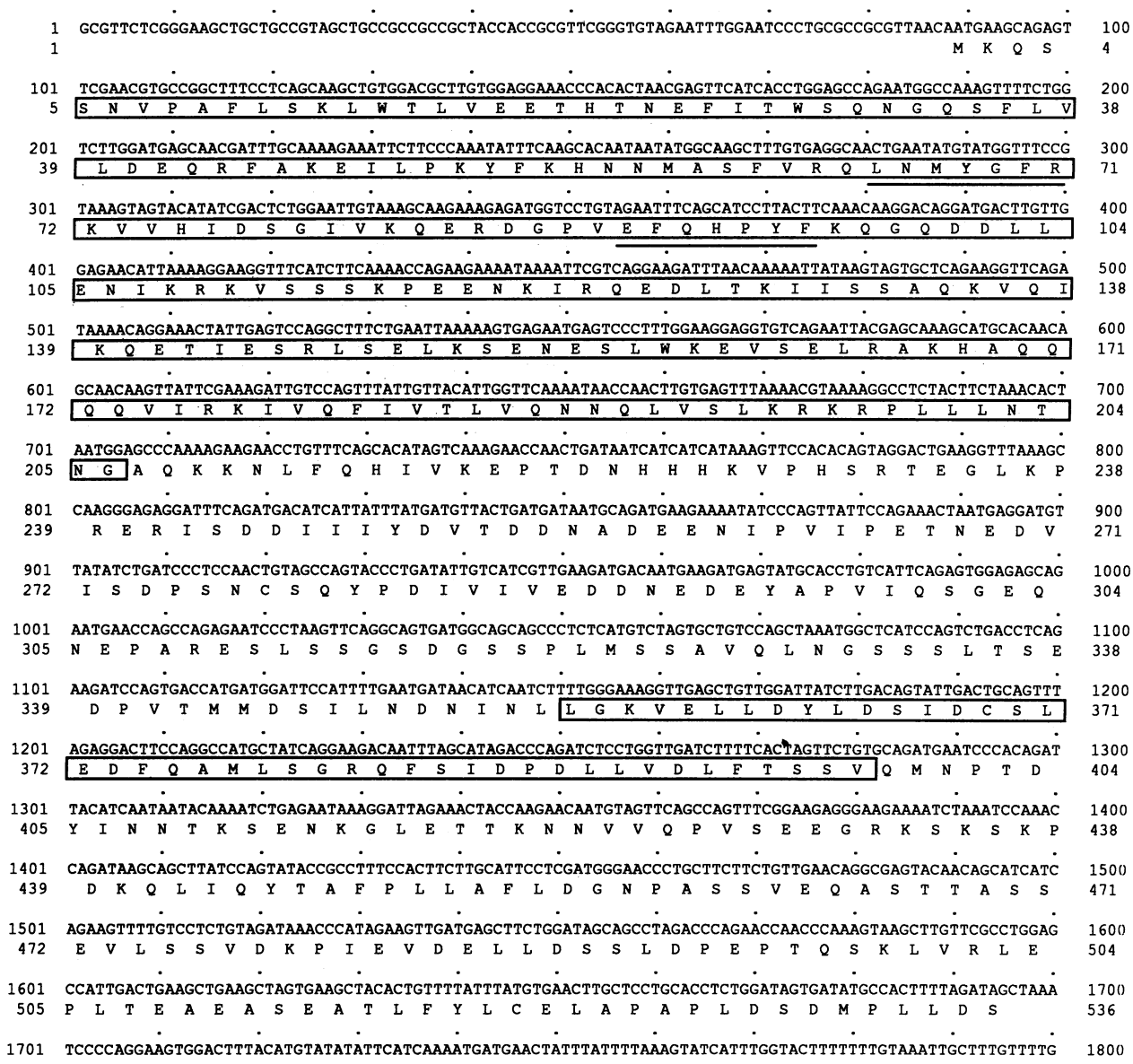

FIG. 2. Partial nucleotide sequence of the 2411-base-pair cDNA in the region of the predicted 536-amino acid open reading frame. Boxed sequences denote regions of homology to the *HSF1* cDNA (ref. 13; see text). The underlined sequences correspond to the degenerate oligonucleotide probes used for screening the cDNA library.

lanes 2 and 3). This complex is the result of specific binding as judged by competition with unlabeled oligonucleotides (lanes 3–5) and by methylation interference analysis (see above). Incubation with antibodies specific to HSF caused a significant decrease in the amount of HSE-containing complex as well as a decrease in the relative mobility of this complex (lane 10). Several control antibodies did not affect binding (lanes 6–9). Western blot analysis indicated that these antibodies interacted equally well with the protein expressed from the cDNA and with purified HeLa HSF (T.J.S., unpublished data). Therefore, the protein encoded by the cDNA either shares epitopes with HeLa HSF or is a component of HeLa HSF.

The HSF cDNA Product Stimulates Transcription *in Vitro*. The HSF cDNA was overexpressed in bacteria and purified by electroelution of the induced protein from an SDS/polyacrylamide gel. This protein stimulates transcription *in vitro* and optimal stimulation requires the presence of a heat shock element (Fig. 5). The level of stimulation is somewhat variable (≈3-fold in lane 4, but 16-fold in lane 6). The average stimulation of the HSE-containing template used in Fig. 5 is ≈10-fold (averaged over 28 pairs of reactions).

Humans Have More Than One HSF. The cDNA we have isolated has sequence homologies with *Drosophila* and *S. cerevisiae* HSF and encodes a protein that binds specifically to the HSE, interacts with antibodies raised against bona fide HeLa HSF, and stimulates transcription *in vitro* in a HSE-dependent manner. Given all of these properties, we were surprised that only one of the five peptides that we had sequenced from HeLa HSF was precisely encoded by the cDNA. Two other peptides had significant identity (5/8 and 6/10 amino acids, respectively), while the final two peptides had either limited or no identity. Peptide 2T10 begins at amino acid 65 of HSF2, and the portion of peptide 2T8 from which the oligonucleotide probe was made (EFQHPCF) matches at 6/7 amino acids the sequence EFQHPYF, which begins at amino acid 90. In retrospect, we were able to isolate our cDNA because one of our oligonucleotide probes was based on the peptide precisely encoded by our cDNA, while the other probe was based on a portion of a second peptide with a 6/7 amino acid identity to our cDNA. Comparison of our five tryptic peptide sequences to the predicted amino acid sequence of a human *HSF* cDNA (*HSF1*) isolated through different means by Rabindran *et al.* (13) revealed that all five

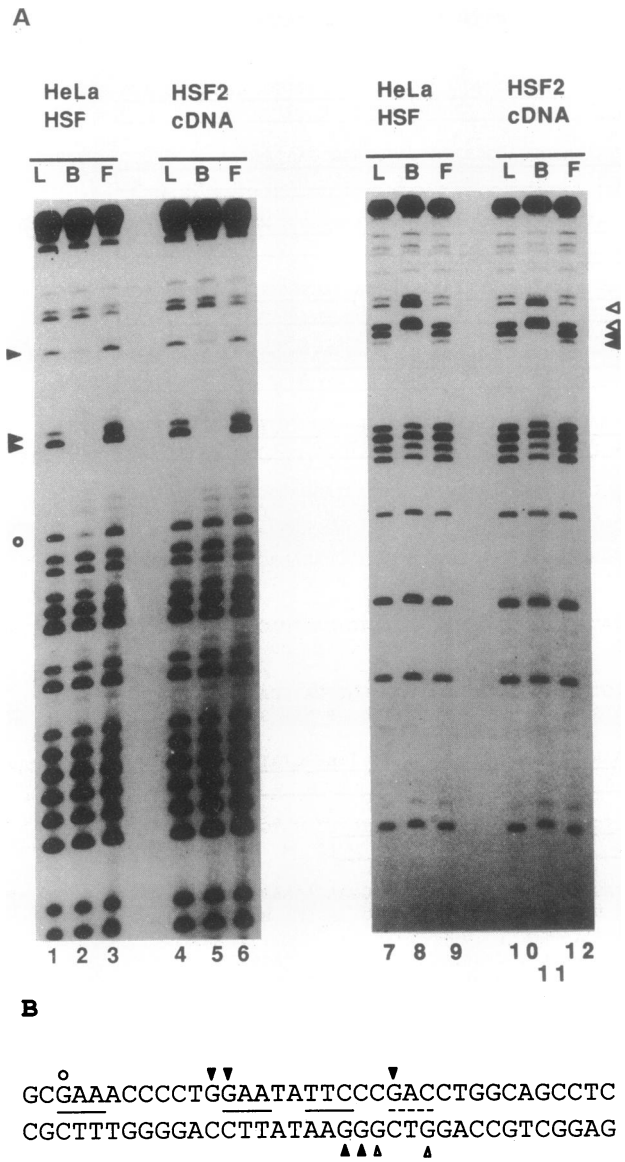

FIG. 3. Methylation interference analysis of the specificity of binding to the HSE by purified HeLa HSF and by the *in vitro* translated product of the isolated cDNA. (*A*) Methylation interference analysis of binding to the HSE promoter fragment with label on the top strand (lanes 1–6) and bottom strand (lanes 7–12) with purified HeLa HSF (60 ng) (lanes 1–3 and 7–9) and the *in vitro* translated protein (15 μl of lysate) (lanes 4–6 and 10–12) was performed (28, 29). L, a G ladder; B, DNA eluted from the DNA–protein complex; F, free DNA from the preparative gel. Solid arrowheads indicate underrepresented G residues in the bound DNA fraction; the open circle represents a G residue partially underrepresented in lane 2; and the open arrowheads indicate G residues overrepresented in lanes 8 and 11. (*B*) Sequence of both strands of the human HSP70A promoter from −116 to −81 relative to the transcription start site (21). Solid underlines indicate matches to the HSE consensus of GAA or TTC. The dashed underline indicates a 2/3-base match. Other symbols are as described for *A*.

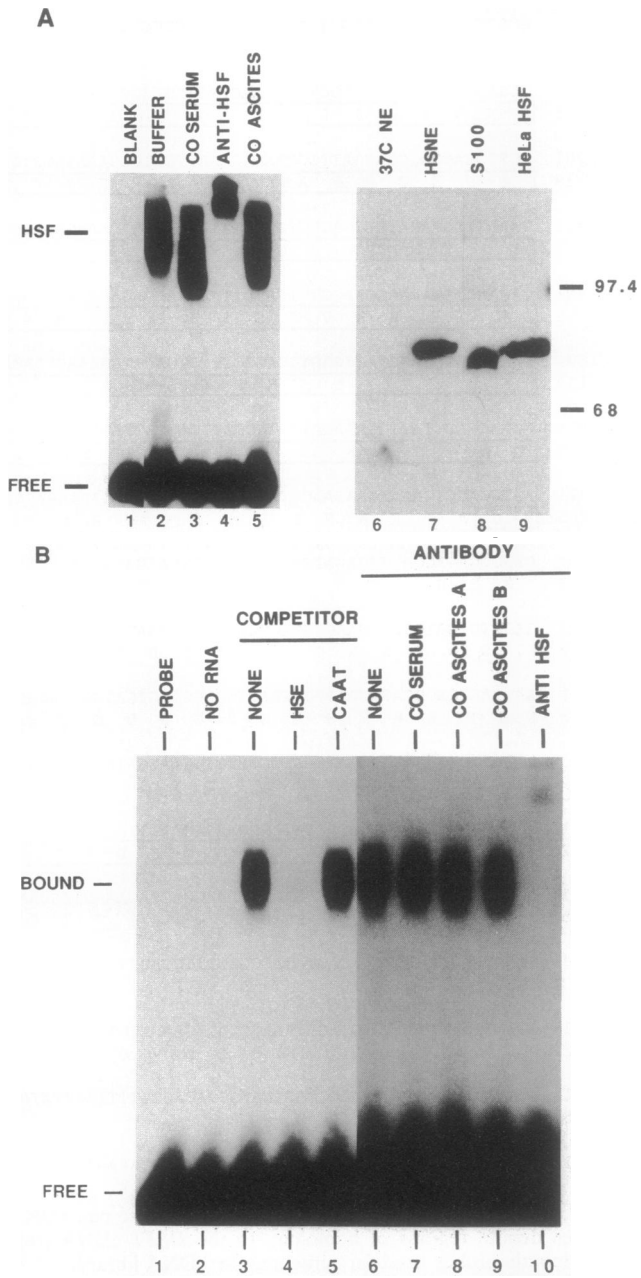

FIG. 4. Purified HeLa HSF and the *in vitro* translated protein are antigenically related. (*A*) Characterization of antibodies raised to human HSF. Lane 1, no protein is included in the mobility shift gel (28). Purified HSF (5 ng) is incubated without (lane 2) or with the following antibodies (1 μl): control serum, lane 3; immune HSF ascites, lane 4; control ascites A, lane 5. Western blot analysis is shown in lanes 6–9: control nuclear extract (lane 6, 420 μg), nuclear extract from heat shocked cells (lane 7, 330 μg), cytoplasmic extract from nonshocked cells (lane 8, 750 μg), and purified HeLa HSF (lane 9, 0.5 μg) are shown. Migrations of molecular size standards (kDa) are as indicated. (*B*) Characterization of the *in vitro* translated protein. Reticulocyte lysate was primed with no RNA in lane 2. Gel shift reactions (4 μl of lysate) were incubated with unlabeled competitor DNA (50-fold molar excess) containing either HSE sequences (lane 4) or CCAAT sequences (lane 5) (28). Lanes 6–10, gel shift incubations of the *in vitro* translated protein with various antibodies (2 μl): no antibodies, lane 6; three different control antibody preparations, lanes 7–9; anti-HSF ascites, lane 10.

of these peptides are encoded by *HSF1*. In the open reading frame of the *HSF1* cDNA, the peptides are found at the following positions: 2T10, amino acid 73; 2T5, amino acid 81; 2T8, amino acid 97; 2T4, amino acid 163; 2T6, amino acid 337.

The *HSF1* cDNA and the cDNA that we have isolated both encode proteins that bind specifically to the HSE and have homology to HSFs of other species. We will refer to our cDNA as *HSF2*. The sequences of the predicted open reading frames of these two human cDNAs have more extensive identity to each other than they do to either *S. cerevisiae* or *Drosophila HSF*. They are 72% identical in the presumed DNA binding domain (ref. 9; amino acids 5–113 of HSF2 and 13–121 of HSF1), 50% identical in the potential multimerization domain (ref. 10; amino acids 115–206 of HSF2 and 126–217 of HSF1), and 50% identical in a C-terminal domain

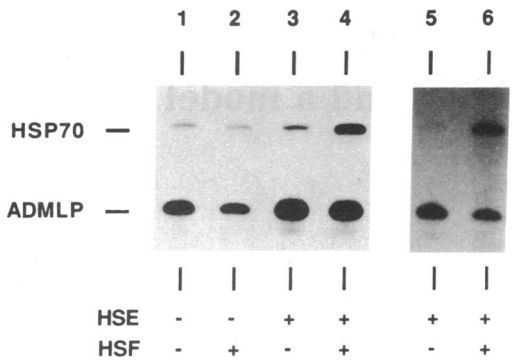

FIG. 5. The cDNA product, overexpressed and purified from bacteria, stimulates transcription *in vitro*. Symbols and abbreviations are the same as in Fig. 1. HSP70 transcription is specifically stimulated 1.4-fold without the HSE (lanes 1 and 2) and 3.1-fold (lanes 3 and 4) or 16-fold (lanes 5 and 6) with the HSE. Lanes 2, 4, and 6 contained 1 μl of overexpressed HSF2. Lanes 1, 3, and 5 contained 1 μl of dialysis buffer.

(amino acids 355–398 of HSF2 and 379–422 of HSF1). The conservation of amino acids in this latter domain implies that it has an important function. It is interesting to note that this C-terminal domain could form an acidic amphipathic helix containing a leucine repeat. This domain might function as a transcriptional activation domain. Comparison of the nucleotide sequence between *HSF1* and *HSF2* shows no region of extensive identity, and low stringency Southern analysis identifies two separate restriction fragments when *Eco*RI- and *Stu* I-digested human genomic DNA is probed with a 100-base-pair fragment encoding the DNA binding domain homology (T.J.S., unpublished data). Therefore, the two *HSF* genes are not allelic, and it appears that there are only two *HSF* genes in humans (although a third, less conserved, gene cannot be ruled out).

What is the function of each of these factors in regulating the heat shock response in humans? While the peptide sequencing data presented above argue that HSF1 is the primary factor expressed in HeLa cells, we do not at present have any information on what the relative levels of HSF1 and HSF2 are in any human cell. The protein encoded by the *HSF2* gene might allow response of human cells to different elicitors of stress than *HSF1* or might help modulate the response to less severe forms of stress. There are differences in the way different cell types respond to stress in intact organisms (31), and *HSF2* might function only in specific cell types to create this diversity. Perhaps humans have evolved two HSF molecules to perform the transient and sustained response functions of the single yeast HSF molecule (7). Finally, both HSF1 and HSF2 retain homology to the presumed multimerization domain of *S. cerevisiae* HSF (10). It is therefore interesting to speculate that HSF1 and HSF2 multimerize to form heteromers that regulate gene expression.

We are grateful to A. Wild for expert technical assistance and C. Wu for communicating unpublished results. We thank A. Aruffo, B. Seed, H. Singh, F. Winston, R. Treisman, A. Sluder, G. Ruvkun, D. Straus, and D. Abalos for gifts of reagents and for help with screening and antibody production. This work was supported by a grant from Hoechst (to R.E.K.), by the Harvard Medical School M.D.-Ph.D. program (T.J.S.), and by National Institutes of Health Fellowship 5 F32 GM12865 (G.J.G.).

1. Lindquist, S. & Craig, E. A. (1988) *Annu. Rev. Genet.* **22,** 631–677.
2. Pelham, H. R. B. (1985) *Trends Genet.* **1,** 31–35.
3. Amin, J., Ananthan, J. & Voellmy, R. (1988) *Mol. Cell. Biol.* **8,** 3761–3769.
4. Xiao, H. & Lis, J. T. (1988) *Science* **239,** 1139–1142.
5. Sorger, P. K. (1991) *Cell* **65,** 363–366.
6. Sorger, P. K. & Pelham, H. R. B. (1988) *Cell* **54,** 855–864.
7. Sorger, P. K. (1990) *Cell* **62,** 793–805.
8. Nieto-Sotelo, J., Wiederrecht, G., Okudo, A. & Parker, C. S. (1990) *Cell* **62,** 807–817.
9. Weiderrecht, G., Seto, D. & Parker, C. S. (1988) *Cell* **54,** 841–853.
10. Sorger, P. K. & Nelson, H. C. M. (1989) *Cell* **59,** 807–813.
11. Clos, J., Westwood, J. T., Becker, P. B., Wilson, S., Lambert, K. & Wu, C. (1990) *Cell* **63,** 1085–1097.
12. Scharf, K.-D., Rose, S., Zott, W., Schoff, F. & Nover, L. (1990) *EMBO J.* **9,** 4495–4501.
13. Rabindran, S. K., Giorgi, G., Clos, J. & Wu, C. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6906–6910.
14. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11,** 1475–1489.
15. Gallo, G. J., Schuetz, T. J. & Kingston, R. E. (1991) *Mol. Cell. Biol.* **11,** 281–288.
16. Tempst, P. & Riviere, L. (1989) *Anal. Biochem.* **183,** 290–300.
17. Tempst, P., Link, A. J., Riviere, L. R., Fleming, M. & Elicone, C. (1990) *Electrophoresis* **11,** 537–553.
18. Ghosh, S., Gifford, A. M., Riviere, L. R., Tempst, P., Nolan, G. P. & Baltimore, D. (1990) *Cell* **62,** 1019–1029.
19. Seed, B. & Aruffo, A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3365–3369.
20. Workman, J. L., Roeder, R. G. & Kingston, R. E. (1990) *EMBO J.* **9,** 1299–1308.
21. Wu, B., Hunt, C. & Morimoto, R. I. (1985) *Mol. Cell. Biol.* **5,** 330–341.
22. Greene, J. M. & Kingston, R. E. (1990) *Mol. Cell. Biol.* **10,** 1319–1328.
23. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56,** 777–783.
24. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89.
25. Hunkapiller, M. W., Lujan, E., Ostrander, F. & Hood, L. E. (1983) *Methods Enzymol.* **91,** 227–236.
26. Harlow, E. & Lane, D. (1988) *Antibodies: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 122–123.
27. Goldenberg, C. J., Luo, Y., Fenna, M., Baler, R., Weinmann, R. & Voellmy, R. (1988) *J. Biol. Chem.* **263,** 19734–19739.
28. Kingston, R. E., Schuetz, T. J. & Larin, Z. (1987) *Mol. Cell. Biol.* **7,** 1530–1534.
29. Gilman, M., Wilson, R. & Weinberg, R. (1986) *Mol. Cell. Biol.* **6,** 4305–4316.
30. Larson, J. S., Schuetz, T. J. & Kingston, R. E. (1988) *Nature (London)* **335,** 372–375.
31. Blake, M. J., Gershon, D., Fargnoli, J. & Holbrook, N. J. (1990) *J. Biol. Chem.* **265,** 15275–15279.