Nucleic Acids Research

Sequence and organization of genes encoding the human 27 kDa heat shock protein

Eileen Hickey*, Susan E.Brandon, Robert Potter, Gary Stein[1], Janet Stein[1] and Lee A.Weber

Institute of Molecular Biology and Biotechnology, University of South Florida, Tampa, FL 33620 and [1]College of Medicine, University of Florida, Gainesville, FL 32610, USA

ABSTRACT
        The 27 kDa human heat shock protein (hsp27) is encoded by a gene
family of 4 members. Two genomic fragments hybridizing to cDNA encoding hsp27
have been isolated, characterized, and sequenced. One clone is a member of a
cluster of three genes linked within a 14-18 kb region of the genome and
encodes a transcript interrupted by two intervening sequences. A single open
reading frame encodes a polypeptide of 22,300 deduced molecular weight. The 5'
flanking region contains two transcription start sites and sequences homologous
to the Drosophila consensus heat inducible control element. Induction of both
potential transcripts follows heat shock in vivo. Accurate heat inducible
transcription occurs at both start sites after injection into Xenopus oocytes.
The second genomic clone is a processed pseudogene lacking promoter elements
and is unlinked with the other members of the hsp27 gene family. The amino acid
sequence of human hsp27 shows striking homology with mammalian α crystallin,
and contains a region towards the carboxy terminus which shares homology with
the small hsp of Drosophila and other organisms.

INTRODUCTION

    The induction of transcription of the heat shock genes in response to

elevated temperatures, first described in Drosophila (reviewed in 1), has now

been demonstrated to be universal among prokaryotes and eukaryotes (2). The

polypeptides (hsp) encoded by the heat shock genes are conserved to varying

degrees. The nucleotide sequences encoding the 70 kDa hsp, the most conserved

class, cross hybridize with the corresponding sequences of organisms as

phylogenetically distant as insects and bacteria (2). The members of the 80-90

kDa class of hsp of diverse eukaryotes still share antigenic determinants (3).

Similarities among the small, 20-30 kDa hsp of different organisms are only

evident upon examination of the amino acid sequences (4,5). Key regulatory

components of the heat shock response are also conserved. Heat shock genes

from various organisms have been shown to be heat inducible following introduc-

tion into heterologous cells (6-8), and are activated at the temperature

characteristic of the heat shock response of the host cell. Deletion studies

have identified a consensus sequence (Pelham sequence) in the promoter region

required for heat inducibility (6), suggesting conservation of control signals

common to all heat shock genes. The heat shock proteins belong to a larger

group of polypeptides, the stress proteins, that are induced in various
combinations in response to environmental challenges and developmental transi-
tions. The heat shock genes therefore provide an excellent system in which to
study the sequence requirements for coordinate regulation of overlapping
subsets of a dispersed group of genes by different inducers of transcrip-
tion.

The specific functions of the individual hsp have not yet been identified,
but synthesis of the small hsp has been shown to be correlated with the
acquisition of thermotolerance (2,9). In cultured Drosophila cells, ecdysone
induces synthesis of only the low molecular weight class of hsp. Such induced
cells are found to be thermotolerant without prior heat treatment (10). Mutants
of Dictyostelium that are unable to acquire thermotolerance have been shown to
lack the ability to synthesize the small hsp (11).

The mammalian small (27-28 kDa) heat shock protein (hsp27) labels very
weakly with ^{35}S methionine; it was therefore overlooked in early studies and
has only recently been subjected to detailed analysis (12-15). Several isoelec-
tric variants of mammalian hsp27 are found, which also differ slightly in size.
The variants derive in part from differences in phosphorylation state (13,14),
but the possibility that they also represent several unrelated polypeptides can
not yet be excluded. Both the induction of synthesis of hsp27, and the extent
of phosphorylation, are responsive to different environmental stimuli. Syn-
thesis of hsp27 is induced in mammalian cells in the presence of arsenite and
amino acid analogs as well as during hyperthermia (13). Increased phosphoryl-
ation of mammalian hsp27, without increased synthesis of the proteins, follows
serum stimulation of growth, or exposure to tumor promoters or calcium iono-
phores (13). This suggests that covalent modification may regulate the activity
of these polypeptides.

We report here the characterization and sequence of two of the four human
genomic sequences encoding hsp27, one of which is closely linked to two
others. The linked gene encodes 2 overlapping transcripts that have different
5' transcription start sites and contain 2 intervening sequences. The promoter
region of this gene contains sequences homologous to the Drosophila consensus
heat inducible control element. The gene is accurately transcribed in vitro,
and is heat inducible when introduced into Xenopus oocytes. The second of the
genes characterized in this study is unlinked to the others, and is shown to be
a processed pseudogene that contains no intervening sequences or transcription
control signals. The deduced amino acid sequence of human hsp27 is homologous
to the reported sequences of small hsp from a variety of organisms (4,7,16,17).

The amino acid sequence also shares with other small hsp a striking homology to the sequence of mammalian α crystallin (4,16).

MATERIALS AND METHODS

Materials: Restriction enzymes, DNA polymerase, and T_4 polynucleotide kinase were from Bethesda Research Laboratories, reverse transcriptase from BioRad, calf intestine alkaline phosphatase from Boehringer, S1 nuclease from P.L, RNasin from Promega Biotec, and T1 ribonuclease, protease K, and DNAse from Sigma. All sequencing reagents and [35]S dATP were obtained from Amersham, while ^{32}P-labelled ATP, UTP, and dCTP were from ICN. Electrophoresis supplies were DNA grade from BioRad and nitrocellulose filters were from Schleicher and Schuell.

Screening a human genomic library: An Alu I-Hae III human genomic library prepared by Maniatis (18) in phage Charon 4A was screened with the cDNA clone pHS208 (19) that encodes the human hsp27. From one million plaques, four hybridizing phage were isolated, plaque purified, and mapped by restriction digestion and Southern blotting using standard techniques (20). The gene containing fragments were identified by hybridization to the cloned cDNA probe pHS208, and by preferential hybridization to cDNA that was made to RNA from heat shocked cells using short oligonucleotide primers as described (20). Restriction fragments containing the mRNA coding region were subcloned into pUC12.

Southern analysis of genomic sequences: High molecular weight human placenta or sperm DNA was digested with restriction enzymes, and 10μg aliquots electrophoresed on agarose, and blotted onto nitrocellulose (21). The blots were hybridized to the nick-translated cloned pHS208 probe at 68° according to standard procedures, with a final stringent wash for 15 min at 68° in 0.1 x SSC, 0.1 % SDS.

S1 nuclease analysis and primer extension: S1 nuclease protection analysis was done with uniformly labelled (22) or end-labelled probes (23), according to the protocol described by Favaloro (24). Uniformly labelled probes were obtained by nick-translation, ligation and restriction of the subcloned plasmid DNA, followed by isolation of the insert fragments on gels (20). The 3' end-labelled probes were prepared with Klenow, and the 5' end-labelled probes were first treated with phosphatase and then labelled with T_4 polynucleotide kinase. Probes were denatured at 90° for 10 min, and hybridization continued for 3 hours at 58°. The S1 nuclease digestions were carried out at 22° for 30 min at 200 units/ml. Primer extension analysis was performed as described by Lee and Luse (25). The precise transcription start site was determined by comparison of
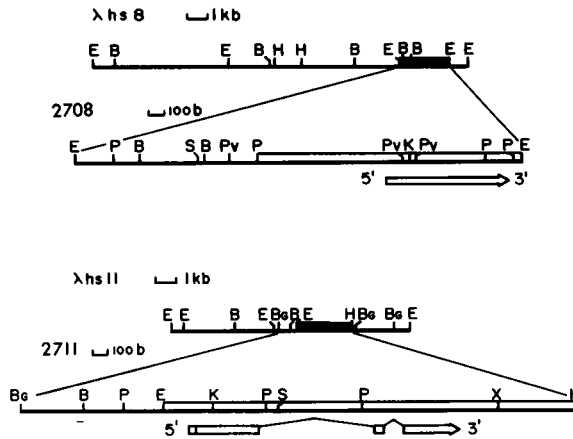
Figure 1 Characterization of human genomic clones encoding hsp27: The maps of the Charon 4A phage λhs8 and λhsII are shown. The filled region indicates the restriction fragment hybridizing to the cloned cDNA and to 42° cDNA. The portion subcloned is shown expanded in the second line. The regions sequenced are marked with an open bar on the map of the subclone. Beneath the maps, the open arrows indicate the S1 nuclease protected regions and the direction of transcription. Restriction enzymes: B-Bam HI, Bg-Bgl II, E-Eco RI, H-Hind III, K-Kpn I, P-Pst I, Pv-Pvu II, S-Sma I, X-Xba I.

the products of primer extension and S1 nuclease protection assays with a seq-
uence ladder. The sequence ladder was derived from dideoxy sequencing in the
antisense direction of the same fragment used as hybridization probe. This DNA,
a Taq I fragment cloned in M13, was cleaved with Taq I following the sequencing
reaction to assure that the sequence ladder and the probe had identical 5'
termini.

Expression in vitro and in vivo: In vitro transcription was assayed in
extracts of HeLa cells prepared according to Manley (26). Linearized plasmid
DNA (500ng) was added to each 10 μl assay, and transcription was carried out in
the presence of ^{32}P labelled UTP. The labelled transcription products from
assays incubated with or without α amanitin were analyzed by electrophoresis
and autoradiography.

In vivo transcription was assayed following injection of 5 ng of plasmid DNA
into the germinal vesicle of Xenopus oocytes as described by Voellmy (8).
Injected oocytes were held at 22° 14-20 hours. For heat induction of transcrip-
tion, oocytes were incubated at 36° for 90 min. RNA was prepared from control
oocytes kept at 22°, or from induced oocytes immediately following heat shock
(8).The presence of transcripts from the introduced plasmid was assayed by S1
nuclease protection analysis using 5' end-labelled probes, or by primer
extension as described above.

Sequencing: Sequencing of restriction fragments subcloned into M13 was per-
formed by the dideoxy termination method (27). For collation of the complete
sequence, the programs of Schwindiger were used (28). Homologies among DNA and
protein sequences were determined with the PCS program (29).

RESULTS

Isolation and characterization of genomic clones: Four phage were isolated from
the Charon 4A library that hybridized to the cDNA clone pHS208. We have shown
by hybridization selection and translation that this cDNA clone is comple-
mentary to a heat inducible RNA that specifically encodes human hsp27 (19).
Restriction mapping revealed that three phage, represented by λhs11, contained
DNA overlapping the same genomic region, while the fourth, λhs8, was quite
different (Figure 1). Each phage contained only one region homologous to hsp27
mRNA (Fig 1, 2711 and 2708). A 2500 bp Eco RI fragment from λhs8 and a 3500 bp
fragment from λhs11 were cloned into plasmid vectors to give subclones 2708 and
2711, respectively (Fig. 1). The subcloned segments were analyzed by S1
protection experiments. The uniformly labelled Eco RI insert from subclone 2708
gave a single 750 bp protected fragment after hybridization with RNA from heat
shocked cells. The Eco RI-Hind III segment from 2711, however, gave protected
fragments of 450, 350, and 65 bp, indicating the presence of two intervening
sequences. The boundaries of the transcription units and intervening sequences
as depicted in Fig. 1 were localized using 5' and 3' end-labelled restriction
fragments as hybridization probes. The precise boundaries of the intervening
sequences and the 3' termini were verified by comparison with the sequence of
the cloned cDNA (see below). Experiments using 5' end-labelled probe fragments
indicated that the gene cloned in 2711 has two transcription start sites
located approximately 50 bp apart (see Fig. 3) Both transcripts can be found
at low levels in control cells, with the shorter transcript more abundant than
the longer one. Levels of the two mRNAs are induced more than 20-fold during
heat shock, without a significant change in proportions (data not shown).
The sequence of the human hsp27 genes: In Figure 2A the sequence of the entire
transcribed region of 2711 plus 160 5' flanking nucleotides and 1000 3' flank-
ing nucleotides, are shown aligned with the homologous sequences of 2708. The
entire sequence of the cDNA clone pHS208 has been determined and it corresponds
precisely to the sequence of 2711 presented in uppercase letters between the
black triangles in Figure 2A. The sites at which transcription initiates in
2711 were determined to within one or two bases by the S1 and primer extension
analysis shown in Figure 3. Because the S1 nuclease gave multiple bands at the

A

2711

-249 gaattcatttgcttttcCTtaacgagaGAAGGTTCCAGatgagggctgaaccctcttcgccccgcccacggcccctgaacgctgggggaggagtgcatgggaggggcg

-140 gccctcaaacgggtcattgcCATTAATAgagacctcaaacaccgcctgctAAAAATACCGACTGGAGGAGCATAAAAGCGCAGCCGAGCCCAGCGCCCCGCACTTTTCT

2708 ...... TTAAAAAAAAAAAAAAATTGA------G--C

Asp
A--CC-----C---- --A--
-30 GAGCAGACGTCCAGAGCAGAGTCAGCCAGC ATG ACC GAG CGC CGC GTC CCC TTC TCG CTC CTG CGG GGC CCC AGC TGG GAC CCC TTC CGC
                                 Met Thr Glu Arg Arg Val Pro Phe Ser Leu Leu Arg Gly Pro Ser Trp Asp Pro Phe Arg

                                                                        Val Ala Val Val Gly His
--- --- --- --- --C --T --- --- --T --- --- --- --- --- --- --- --- G-- --AC
61 GAC TGG TAC CCG CAT AGC CGC CTC TTC GAC CAG GCC TTC GGG CTG CCC CGG CTG CCG GAG GAG TGG TCG CAG TGG TTA GGC
    Asp Trp Tyr Pro His Ser Arg Leu Phe Asp Gln Ala Phe Gly Leu Pro Arg Leu Pro Glu Glu Trp Ser Gln Trp Leu Gly

    Gln Leu Leu Ala Pro Leu Arg Ala Ser Pro Ala   Val   Trp Pro   Ala
--- --T --- --- --CC --- --- --- T-- --- --- --- --- --G G-- --- -- --- - --- --- G-- --- --- ---
144 GGC AGC AGC TGG CCA GGC TAC GTG CGC CCC CTG CCC CCC GCC GCC ATC GAG AGC CCC GCA GTG GCC GCG CCC GCC TAC AGC
    Gly Ser Ser Trp Pro Gly Tyr Val Arg Pro Leu Pro Pro Ala Ala Ile Glu Ser Pro Ala Val Ala Ala Pro Ala Tyr Ser

    Tyr   Gly   Asn   Pro   Ile
TA- --- --- --- --G --- G-- --- --- --- --AT --- --C- --A --- --- --- --- --A --- --- --C A--
225 CGC GCG CTC AGC CGG CAA CTC AGC AGC GGG GTC TCG GAG ATC CGG CAC ACT GAC GAC CGC TGG CGC GTG TCC CTG GAT GTC
    Arg Ala Leu Ser Arg Gln Leu Ser Ser Gly Val Ser Glu Ile Arg His Thr Asp Asp Arg Trp Arg Val Ser Leu Asp Val

    Tyr   Ser   Met
--- T-- --- --- T-C --- --- --- --- --- --- A-- ---
304 AAC CAC TTC GCC CCG GAC GAG CTG ACG GTC AAG ACC AAG GAT GGC GTG GTG GAG ATC ACC ggtgagcccccctgctcctgcaggggagag
    Asn His Phe Ala Pro Asp Glu Leu Thr Val Lys Thr Lys Asp Gly Val Val Glu Ile Thr

394 gaggaggctagcagggcgggcagggccggggcgtgcggttgaaacgggggtcccgggggcctggggagttaaacgttggcccagcaccgggaaaaacaggactcctgat

504 tcccttgctcaggaattgggagtgcgggtcgcttctaagggcgctttctgctctgtaatcccagcgctttgggaggccgagacgggaggatcgcttgaggccaggagttc

614 aagactagcctgggcaacatagcgagacgcgccccccgccccgaccccgcgccattacaaaaaaaaagcaaacaaaaatttttttaaagatcatcgatgaagagagaaa

724 atgcgcttttctacagagtcccccttcccacccacagcccccatccccagataagcggggaggttccctggcgcggtgccagtttctagccgctgagtgggcgtgtgcgcggc

834 tccaagtgcgcctgcgtactgctcactcccagctccgcgccctgctccgttcctcccaaaactctgaatcgaagaactttccggaagtttctgagagcccagaccggcg

944 ggcacgcccccatccccaaccccctctgttaatccctaccagcctgcagtcctggctgcttccaagcaggaggtggggcctctggctagcggggccgaaaaagtcccctc

                      Ser
A-- --- --- --- --- --- --- --- --- --- --- --- --- --- A-- --- --C
1054 ccccgcatgtctgatttccctcttccccccaaa GGC AAG CAC GAG GAG CGG CAG GAC GAG CAT GGC TAC ATC TCC CGG TGC TTC ACG
                                     Gly Lys His Glu Glu Arg Gln Asp Glu His Gly Tyr Ile Ser Arg Cys Phe Thr

1141 CGG AAA TAC ACG tgagtcctggcgccaggtcggggtgggtgggtggcgtggggtgggtcagggaagagggcacagggacccacccggtgtgtaatgtaacgctt
     Arg Lys Tyr Thr

                                Lys
--- --- --- --- --- --- A-G --C --- --- --- --- --- --- --- --- --- --- --- --- ---
1247 gcctttcctctctgcacgtccagg CTG CCC CCC GGT GTG GAC CCC ACC CAA GTT TCC TCC TCC CTG TCC CCT GAG GGC ACA CTG ACC
                            Leu Pro Pro Gly Val Asp Pro Thr Gln Val Ser Ser Ser Leu Ser Pro Glu Gly Thr Leu Thr

                                                                                Leu
--- --- --- --- --- --- --- --- --- --G --- --- --A --- --- --- --- --- --- ---
1336 GTG GAG GCC CCC ATG CCC AAG CTA GCC ACG CAG TCC AAC GAG ATC ACC ATC CCA GTC ACC TTC GAG TCG CGG GCC CAG CTT
     Val Glu Ala Pro Met Pro Lys Leu Ala Thr Gln Ser Asn Glu Ile Thr Ile Pro Val Thr Phe Glu Ser Arg Ala Gln Leu

          Leu Gly Ala Ala Lys Ser Asp GlnSerAlaAlaLys   ACCC
--- --CT- GG- --- --G --- --- --C C--T--A --- --- --- --- --- --CCGG -C- --- --- --A--C- --- --- --- ---
1415 GGG GGC AGA AGC TGC AAA ATC CGA TGA GACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCCTCCCCCGCCA
     Gly Gly Arg Ser Cys Lys Ile Arg *

---------- --- ---GG --- --- --- ---G --- --- --- --- ---G --- --- CAAAAATTTAAAAAAAAAATA......
1514 CCTGTGTGTTCTTTTGATACATTTATCTTCTGTTTTTCTCAAATAAGTTCAAAGCAACCACCTGTCactggcccaggccctggtgtttgtggaaggaagcctcaggcac
                                                                    *   *

1624 ctgccatttgctggctttcaggagtcatctttgctcaggcccgtgctgggccatgtgggtacactggtgtaggttgctggacacaggctgactcactcacatccataaagacag

1734 aggtcttagggccgggcgcagtggctcatacctacaatcccagcactttgggggggttgaagcaggaggaggtgcttgaagccaagagttctagaccagcctggacaacata

1844 gtaagactgtctctaaaaaataaaaattaggcagggtggtactgcacgcctgtagtcccagctactcaggaggctgaggcaggaggatcgcttgagcccagagttgtgaa

1954 ggtacagtgagctaacatcgtgccattgcactccagcctgggcaacagaacaagatcctgtctcaaaacaaccaaaagcccagagagagaagagtgagacccccatctttaa

2064 aagaaaaaaaaaaaggtcatgattgcaaggtcacgattgcaattaaaactgtaaggtgggaaggaggaggaggaaataagagaagcacctgaggcttgagttctcaggagc

2174 acctaggttgggtcccaggtgaaggggcacagaggtaattgcacctcagagctgatgggaggattactatgtca

B   2708

-900 ctgcagtccagtgggctcttccaagggaaaaaggaacaacccgtgtggtggggagtggaggggccggacagaaaggggcttcctgactcattcgtttaaggcggcattca

-790 tcttctaggctataaagtgtacacattaagggcagaattttaacacaaagcgactacttccctgtgtcttagtctcctcatctgtaaatgggactattacagtacctg

-680 ctttttaggattaaaacagaattaatttatataaagcacgttgtactatgtgcctgacacataagataagaaagcactatgtaagttatagttactagtgtccatgcatt

-570 caataaatgctttaatatgatacacatttgtacactccaaagaataattacacagcaaatatctatgtaac**c**ggctgggtgcatctcgtcacgcctgtaatcccagcact

-460 ttgggaggccgagtcgggtggatcactgaggtcaggagttcgagaccagcctggccaacatggtgaaaccccctctctactaaaatgagaaaaattagccaggcatggtg

-350 gcacatgcctgtagtcccagctactcgggaggcttaggcaggagaatcacttgaactggggaagcgaaggtggcagtgagctgagattgcgccattacactccagcctgg

-240 gcaacagagcgagactccgtctctaaataaataaataaagcaaaagaaatctatgtaaccatttcttgaacagccgacgcgcacctgcaatcccagcagttttgcaaggc

-130 gaggcgggaggatcgcttgaggtcaggagtttgagaccggcatgggcaatacagcaatacaccatctcttaaaaaaaaaaaaaaatt**GACACTTGTCCAAGCAGACGTC**

-20   **CCCAGCAGAGCCAGCCAGC**ATG

Figure 2 A. The aligned DNA sequences of 2711 and the homologous regions of 2708: The numbered DNA sequence of 2711 is shown, with +1 being the start of the initiator ATG. Upper case letters indicate sequences present in the processed mRNA; the amino acid sequence is aligned below the DNA. The sequence of 2708 is aligned above that of 2711; dashes indicate identity of nucleotides; spaces have been introduced to maximize matching. Only the amino acids in 2708 which differ from 2711 are shown, above the sequences. The common control sequences are shaded. The Pelham sequence is underlined. Large arrows indicate transcription start sites. The black triangles indicate the boundaries of the cDNA sequence cloned in pHS208. The dotted underlines and overlines, and the horizontal arrow indicate recognized sequences described in the text. The small double brackets over the sequence in IVS I and in the 3' flanking region delineate regions of homology to the human Alu repeat.
B. The 5' flanking sequences of the pseudogene 2708: Sequences 5' to the 2708 regions in Fig 4A are shown. Upper case letters indicate homology with 2711, and nucleotide +1 is at the ATG. The double small brackets enclose the Alu repeat. Dotted underlines indicate the polymerase III promoter. The flanking direct repeat is marked with dotted overlines.

A+T rich first cap site at all digestion temperatures and nuclease concentrations assayed, the termini indicated by primer extension were used in assigning the cap sites (Fig. 2A, large arrows). The resulting mRNAs have 5' untranslated leaders of 40 (major transcript) and 90 (minor transcript) nucleotides. The transcribed region contains two intervening sequences of 723 and 120 nucleotides, bounded by appropriate splice signals. The precise boundaries of the intervening sequences were determined by comparison with the sequence of the cDNA clone pHS208. The first ATG in both transcripts is designated base +1 in the numbering of 2711 shown in Figure 2A. This ATG opens a reading frame that extends to a stop codon at nucleotide 1437, and predicts a polypeptide of 199 amino acids with one internal methionine, and a deduced molecular weight of 22,300. All other potential reading frames would encode polypeptides no longer than 47 amino acids. The 3' terminal nucleotide, corresponding to the polyadenylation site of the hsp27 cDNA, occurs 20 bases downstream from a consensus AATAAA polyadenylation signal. The two complete spliced transcripts are 825 and 775 nucleotides long and have 139 3' noncoding nucleotides. Particular G + T rich sequences twenty to thirty-eight bases 3' to the AATAAA have been reported
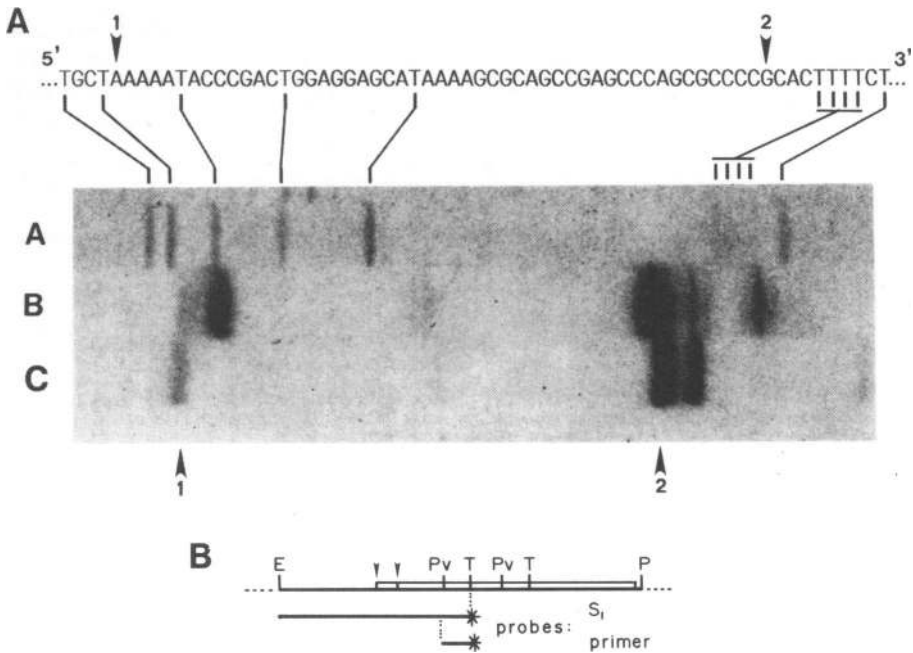
Figure 3. Determination of the transcription start site in 2711: S1 nuclease protection and primer extension were used to identify the precise transcription start site as described. A: T lane of the sequence ladder; B: S1 nuclease protected fragments; C: primer extension products. The sequence of the region shown is aligned above the figure; arrows indicate the start sites. The source of the probes is shown below the figure. The map shows the portion of 2711 containing the complete first exon, and 250 flanking bases. E-Eco RI, P-Pst I, Pv-Pvu II, T-Taq I. Labelling was at the Taq I site as shown.

to be important for efficient processing of mRNAs (30). A similar sequence is found thirty-five nucleotides downstream from the AATAAA in 2711 (Fig. 2A, between the stars).

The 5' flanking sequences of 2711 contain TATA boxes (CATAA) 22 nucleotides upstream from each transcription start site (Figure 2A, shaded). One hundred twenty-one nucleotides 5' to the distal cap site is a sequence homologous to the Drosophila consensus heat control element (CTGGAAT[N]TTCTAGA) (6), with 8 nucleotides separating the initial CT from the remaining sequence (underlined). Seven bases downstream there is an abbreviated repeat of this element (also underlined). A dotted overline in the same region marks an inverted complement of the first consensus sequence. From nucleotides -190 to -180, underlined with an arrow, is an inverted copy of the consensus Sp1 binding site (31).

Both the first intervening sequence and the 3' flanking sequences contain regions of homology with the human Alu repeat family (32). The double brackets over the sequences in Figure 2A mark the boundaries of a sequence in IVS I which corresponds to one-half of the Alu repeat, and of a 3' flanking sequence which contains an entire copy of Alu but is not flanked by the usual short direct repeats, and contains no apparent polymerase III promoter sequences.

The hsp27 gene sequences cloned in 2708 clearly indicate a processed pseudogene. The intervening sequences present in 2711 are absent from 2708, and the closed splice sites correspond precisely to the corresponding sequences of the cDNA clone. Also, 2708 is flanked at the 5' and 3' boundaries with a poly A sequence and lacks the 5' leader sequences corresponding to the distal transcription start site in 2711. The homologous region of 2708 differs from the 2711 transcript by 14.3%, of which 10% consist of single base differences and small deletions and additions. The remainder is the result of one 20 base and one 12 base segment present in 2711 and absent in 2708. Of the single base changes, 24 are transversions, and 30 are transitions. A total of 6 base differences in 2708 are insertions, and in addition to the stretches mentioned above, 9 bases are deleted. The 3' terminus is also shortened by about 7 nucleotides with respect to 2711. The 2708 sequence encodes an open reading frame of 193 amino acids with 23% differing from 2711, 7% due to nucleotide changes in codons, and the rest due to frame shifts.

Figure 2B shows 820 5' flanking nucleotides of 2708, including the poly A and the 2711-homologous sequences up to the ATG (nucleotide +1). There are no 5' flanking sequences similar to polymerase II transcription control elements. However, between bases -497 and -190 there is a complete member of the Alu repeat family (double brackets), flanked by an 11 base direct repeat (dotted overline), and containing sequences corresponding to the split polymerase III promoter (dotted underline) (33).

Expression of the cloned genes in vitro and in vivo: Both forms of the hsp27 gene contain sequences that act as promoters in the in vitro transcription system described by Manley (26). DNA from 2711 linearized with Hind III produces a 2.2kb transcript in vitro (Figure 4A). This agrees with the start site predicted by the sequence, and by the results of S1 analysis. Transcription from 2711 is completely inhibited by α amanitin, confirming that the gene contains an active polymerase II promoter. In contrast, there is no evidence for polymerase II transcription of 2708. Transcription from Eco RI digested 2708 DNA is only slightly sensitive to 5 μg/ml of α amanitin and labels an RNA of about 1.2 kb, 450 nucleotides longer than predicted by S1 mapping. RNA of
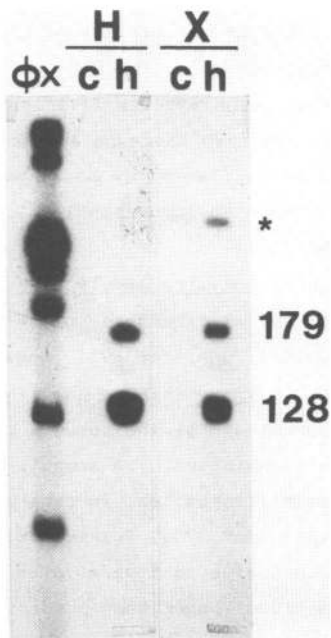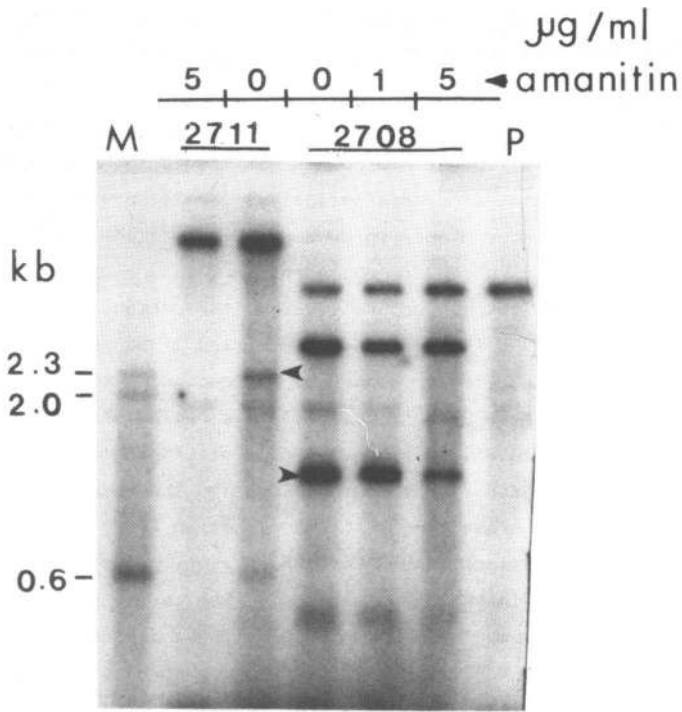
this size, however, would result from initiation at the polymerase III promoter identified in the 5' flanking sequences of 2708. Northern blot analysis of RNA from HeLa cells incubated at 42° for 3 hr, hybridized with pHS208, shows one band of 960 nucleotides, appropriate for a polyadenylated transcript from 2711 (19, 20). No evidence of an 1.2 kb mRNA is found, indicating that transcription initiating at the polymerase III promoter in vivo does not continue through the regions of 2708 that are homologous to hsp27 mRNA.

The hsp 27 gene in 2711 also functions in vivo when injected into the Xenopus oocyte (Figure 3B). S1 nuclease protection studies of the RNA isolated from heat stressed Xenopus oocytes shows that transcription is initiated accurately at both start sites on the introduced gene, in approximately the same proportions as is seen in human cells. No protected fragments correspond-ing to the 2711 transcripts are found in oocytes held at the control tempera-ture. Northern transfer analysis of RNA from injected oocytes shows that approximately 50% of the transcripts from 2711 are processed to a size which comigrates on agarose gels with authentic human hsp27 mRNA, while the remainder appears in a diffuse region of higher molecular weight (not shown).

Homologies between hsp27 and other low molecular weight heat shock proteins and mammalian α crystallins: The amino acid sequence of human hsp27 shows two large regions with more than 50% homology to mammalian α crystallins, separated by a non-homologous region 14 amino acids longer in hsp27 than in the crystal-lins (Fig. 5A). Fifty percent of the residues between amino acids 18 and 59 of hsp27 are found in the same positions in either mouse or bovine α crystallin. In a region between human hsp27 amino acid 84 and 178, there is 64 % agreement with the sequence of the α crystallins. Included in this region is a stretch of 11 contiguous amino acids identical to the corresponding sequence of bovine α crystallin B.

Comparison of the sequences of human hsp27 and small hsp from Drosophila (16,17), Xenopus (5), Caenorhabditis (4), and soybean (34) indicates that the carboxy terminal half of the polypeptide is the most highly conserved. Within

Figure 4 A. In vitro expression of genomic sequences encoding hsp27: Autoradio-graph of transcription products from Hind III cut 2711 and Pst I cut 2708 incubated in the Manley system in the presence of the indicated amounts of α amanitin. Arrows indicate the polymerase II transcript from 2711, and the polymerase III transcript from 2708. M- Hind III lambda marker, P- control transcription assay containing pBR322.
B. Heat inducible transcription of 2711 in Xenopus oocytes: Transcription products from plasmid DNA introduced into oocytes as described in Methods were analysed by S1 nuclease protection. Fragments protected by RNA isolated from oocytes were compared with those protected by HeLa RNA. H – RNA from HeLa cells; X - oocyte RNA; c - incubation at control temperature; h - heat shocked cells or oocytes. Numbers indicate length in bases of protected fragments. Star indicates position of probe fragment.

A

B

Figure 5 A. Comparison of the amino acid sequences of human hsp27 and mammalian
α crystallins: The amino acid sequence of human hsp27 (HS) has been aligned
with that of murine α crystallin A (A), and bovine α crystallin B (B). Hyphens
were introduced to maximize homologies between the crystallins; spaces were
introduced to gain the best alignment with the heat shock protein sequence.
Amino acid homologies between hsp27 and either crystallin are capitalized and
shaded.
B. Comparison of the regions of homology of small hsp from several organisms:
The region of 78 amino acids of human hsp27 showing greatest homology to the
amino acid sequences of other heat shock proteins is shown aligned with the
corresponding sequences from Drosophila (Dros), Xenopus, Caenorhabditis
(Caeno), and α crystallin (Crys). Alignment was maximized by introduction of
spaces (hyphens), and deletions (letters below the lines). Homologies are
indicated by shaded upper case letters. Arrows indicate the boundaries of the
second exon in human hsp27.
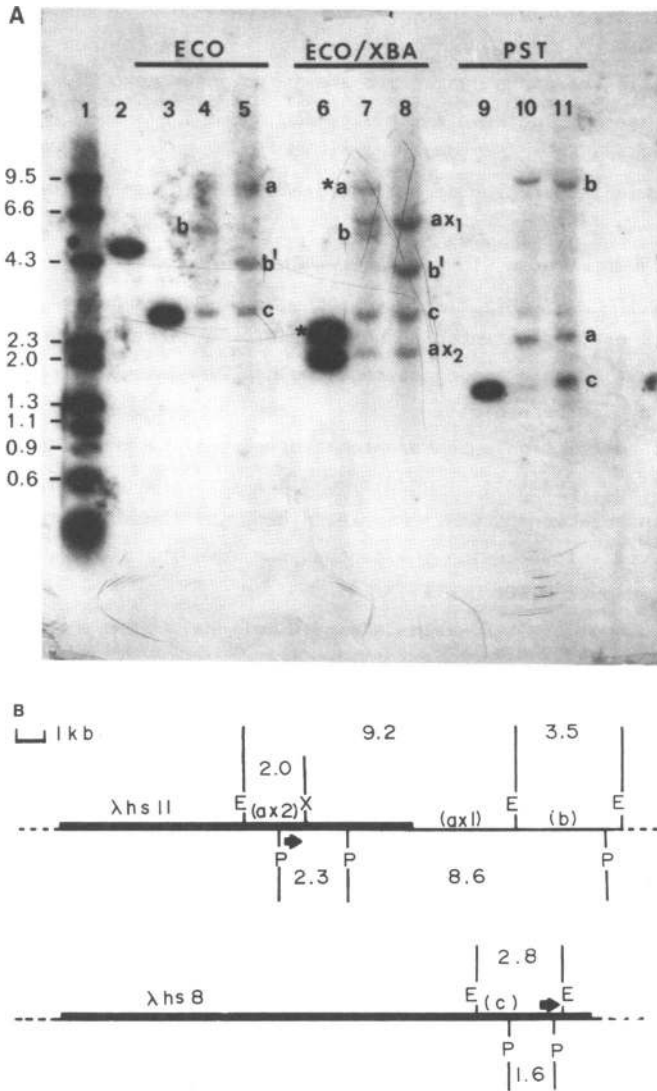
Figure 6 A. Southern analysis of genomic sequences encoding hsp27: Human sperm DNA (lanes 4, 7, 10) or placental DNA (lanes 5, 8, 11) was digested with the indicated restriction enzymes, and Southern blots were hybridized to nick-translated pHS208. Appropriate digests of the genomic subclones were run in parallel: lane 3, Eco RI cut 2708, lane 6, Eco RI-Xba I cut 2711, lane 9, Pst I cut 2708. Lane 1 - lambda markers, lane 2 - linearized pHS208. The stars indicate incomplete digestion products.
B. Map of the organization of human hsp27 genes: The heavy bars indicate the limits of the DNA pieces cloned in phage λhs8 and λhs11, as indicated. Short arrows show the location and direction of transcription of the cloned regions hybridizing to pHS208. Only restriction enzymes used in Fig. 6A are indicated; the numbers between the vertical lines give the fragment sizes in kb. The location of the four members of the gene family are indicated on the map by the labels they are given in the Eco-Xba digest in Fig 4A, Lanes 7&8.

the region between amino acids 100 and 178 of human hsp27, the small hsp
presented in Figure 5B show from a minimum of 42% to a maximum of 50% homology.
The soybean 17.5 kDa hsp shows only a short region of homology, with 8 of 14
amino acids identical between amino acid 158 and 171 of hsp27 (not shown). The
5 amino acid sequence, Asp-Gly-Val-Leu-Thr, conserved in all Drosophila small
hsp, lies within this region of the soy protein, with Asn substituted for Asp
(34). In human hsp27 the changes in this region have been conservative; the
equivalent sequence reads Glu-Gly-Thr-Leu-Thr.

Organization of hsp27 genes within the genome: We have isolated a heat induci-
ble gene and a pseudogene encoding human hsp27. To determine whether there are
additional human genomic sequences encoding hsp27, human DNA from two sources
was analyzed by Southern blotting. Human placental and sperm DNA were digested
with specific restriction enzymes which did not cut the pHS208 cDNA clone, and
which gave unique fragments of predicted sizes from the genomic subclones that
hybridize with the cloned cDNA. Parallel digests of the subclones were electro-
phoresed for comparison beside the genomic DNA, and the blot was probed with
nick-translated pHS208 (Figure 6).

An Eco RI digest of λhs8 yields a single internal 2.8 kb fragment that
hybridizes with pHS208. In λhs11, however, the hybridizing Eco RI fragment
contains a linker-generated site adjacent to a phage arm, and the actual size
of the equivalent genomic Eco RI fragment can not be predicted. However,
digestion of λhs11 with both Eco RI and Xba I gives an internal hybridizing
fragment of 2.0 kb. Pst I digestion of λhs8 and λhs11 gives hybridizing
fragments of 1.6 and 2.3 kb respectively. Figure 6A shows the presence of both
gene arrangements in each preparation of human DNA. Hybridization of the hsp27
cDNA probe to Eco RI digested DNA reveals 3 hybridizing fragments, each
containing no more than 1 or 2 gene copies (quantification not shown). The
fragments are labelled a, b, and c in order of decreasing size. Fragment c
comigrates with the 2.8 kb Eco RI fragment from λhs8. In the Eco RI-Xba I
digest, the fragments are given the label of the Eco RI fragment to which they
correspond. In this digest, a 2.0 kb hybridizing fragment is seen (ax2), that
is derived from Eco RI fragment a, and comigrates with the Eco RI-Xba I
fragment of λhs11. The predicted 2.3 and 1.6 kb fragments are also present in
the Pst I digest, labelled a (containing the 2711 gene) and c (containing the
2708 gene). Thus, there does not appear to have been any gene rearrangement
during cloning and amplification.

The data in Figure 6A also reveal two more members of the gene family
homologous to pHS208, that are linked with the gene cloned in λhs11 in a

contiguous region of the genome of between 14 and 18 kb. The third form of the
gene appears as fragment b in the Eco RI digest, and reveals a restriction site
polymorphism between the two sources of human DNA. Fragment b is unchanged in
the Eco RI-Xba I digest, but Eco RI fragment a is cleaved by Xba I into two
hybridizing fragments. Only one hybridizing fragment, ax2, can be derived from
the DNA region cloned in λhs11 (see Figure 1). Therefore, the new fragment,
ax1, represents a fourth separate sequence homologous to pHS208, that is linked
to the λhs11 sequence within an 8.6 kb Eco RI fragment in the genome. The ax1
sequence and the Eco RI b sequence appear to be included together in the 9.2 kb
Pst I b fragment. Thus, the Eco RI b sequence is also linked to the λhs11 se-
quence in the genome within a fragment of at least 14 kb, and no more than 18
kb. The map of the gene organization derived from Southern analysis is shown in
Figure 6B. The λhs8 pseudogene does not appear to be linked with any of the
other hsp27 genes.

DISCUSSION

     The general picture of the structure and organization of heat shock genes,
as derived from the Drosophila system, is one of clustered families of related
genes lacking intervening sequences. The Drosophila hsp70 genes are found at
two loci on the same chromosome, one containing a pair of genes, the other a
cluster of four (35). The small hsp genes are clustered within a region of 11
kb (36-38). The clustering of small hsp genes within the genome now appears to
be common to many organisms. Four small hsp genes of Caenorhabditis are linked
in a remarkable head to head inverted repeat (4). Two Xenopus small hsp genes
have also been isolated as a tandem array (7). The unlinked member of the human
hsp27 gene family proved to be a processed pseudogene. In Drosophila, the only
hsp gene found to contain an intervening sequence is the single copy hsp83
gene. This gene is expressed in unstressed cells as well as during heat shock
(2). Intervening sequences are also present in the "cognate" genes that encode
proteins homologous to hsp70, but are not heat inducible (39). It initially
appeared likely, therefore, that intervening sequences in heat shock genes
characterized constitutively expressed homologues of the hsp, or heat inducible
genes with high levels of expression at control temperatures. As the heat shock
genes of a wider variety of organisms have been investigated, a more complex
picture has emerged. The small hsp genes of nematodes, which are under strict
temperature regulation, have been shown to be interrupted (4), and we present
here the sequence of a heat inducible human hsp27 gene which gives a spliced
transcript. The rapidity with which the hsp mRNAs reach the cytoplasm has been

attributed to the abbreviated processing required for unspliced mRNAs (40).
There is evidence that processing and/or transport of most non heat shock mRNAs
is inhibited during heat shock (40, our unpublished results). This suggests
that there may be a specific mechanism by which the hsp27 transcript circum-
vents this inhibition.

The 5' regulatory sequences of the gene cloned in 2711 show characteris-
tics common to most published heat shock gene sequences (2,4,6,7,16,17,35,41-
43). Multiple copies of the Pelham consensus heat inducible control element are
present: one quite exact, one inverted, and one abbreviated. This agrees with
the complexity of this region found by Voellmy for the human hsp70 gene (41).
The presence of more than one Pelham sequence has been reported to be required
for efficient heat inducible expression of hsp genes introduced into homologous
cells (44,45). The human hsp27 gene also contains an inverted copy of the
sequence TGGGCGGGGC, corresponding to the Sp1 protein binding site necessary
for transcription of SV40 genes, that has also been found upstream from many
eukaryotic genes (31). In SV40 this sequence functions in either orientation,
since it is required both for early and late transcription. The key element of
this sequence consists of GGGCGG, which is also found repeated in the 5'
sequences flanking the human hsp70 gene (31,41). The TATAA sequences of 2711
correspond to the type (CATAAA) found in the majority of the globin genes (46).

At the 3' terminus the sequence GCCCT is found 10 nucleotides downstream
from the polyadenylation site. This sequence is common to the 3' termini of
many genes, however, no function has been determined for it (47). On the other
hand, a sequence similar to GGTGTTGG found 24-38 bases downstream from the
AATAAA has been shown in deletion studies with the HSV thymidine kinase gene to
be important for efficient processing and polyadenylation (30). This sequence
has been identified in a number of viral and cellular genes (30), and is
present in the human hsp27 gene at the appropriate site.

Pseudogenes, which have been reported in most gene families, appear to arise
either as tandem repeats of a functional gene that are subsequently inactivated
(48,49), or by reverse transcription of mRNA and insertion of the cDNA into the
genome (50). The Xenopus small hsp pseudogene described by Bienz is of the
first type; it is linked with functional hsp genes, but has a divergent Pelham
sequence, encodes a polypeptide quite different from the normal small hsp, and
lacks a functional promoter (7). In contrast, the human hsp27 pseudogene
described here is of the second type, a processed pseudogene, characterized by
precise excision of intervening sequences from the coding region, termination
of the sequence at the polyadenylation site, and presence of a direct repeat

flanking the pseudogene sequence. The presence of the long open reading frame
encoded in 2708, unexpected in a pseudogene, is particularly unusual given the
divergence between the 2708 and 2711 sequences. It is possible that the
presence of the potential coding region in 2708 may reflect a sequence present
in one of the hsp27 genes that have not yet been studied.

Hsp27 is a major cellular phosphoprotein, and has been reported to be
phosphorylated primarily on serine residues (51). Interestingly, the small heat
shock protein encoded in 2711 contains 21 serine residues (10%), as well as 13
threonine residues and 5 of tyrosine, giving a protein with almost 20% of its
amino acids susceptible to phosphorylation. The presence of the single methion-
ine residue in the polypeptide encoded by 2711 was surprising, since in our
hands labelling of hsp27 with ^{35}S methionine _in vivo_ was not detected (12).
Similar lack of labelling of human low molecular weight hsp with methionine has
also been reported elsewhere (13). A recent report, however, shows that a low
level of incorporation of methionine into a form of hsp27 can be seen following
prolonged labelling of human lymphocytes at elevated temperature (52). The
polypeptides encoded in the remaining two hsp27 genes may entirely lack
methionine, which could mask the incorporation of label into the polypeptide
encoded by the 2711 gene.

It is striking that the homology between human hsp27 and mammalian α
crystallins is much more extensive than that between hsp27 and small hsp of
lower organisms. The derivation of crystallins from the duplication of part of
an ancestral small hsp has been suggested (53). Regions of amino acid alignment
between hsp27 and α crystallin are found along almost the entire length of the
crystallin. This suggests that the α crystallin gene may have been derived from
a complete hsp gene that had an amino terminal region differing in sequence
from earlier small hsp genes. The human small hsp would thus be closer evolu-
tionarily to α crystallins than to invertebrate small hsp. Since crystallins
have been shown to have highly conserved exon-intron distribution (53), the
location of an intervening sequence at the same position in the Caenorhabditis
hsp and the α crystallin gene is interesting. This position is quite different
from that of either IVS in the human gene. The crystallin-like arrangement of
exons may yet be found in one of the human hsp27 genes which has not been
characterized. It is also possible that the Caenorhabditis gene is closer in
structure, if not in sequence, to the ancestral form from which the crystallins
are derived. Where present, the intervening sequences delineate protein regions
which are most highly conserved (4, Fig 5). In Caenorhabditis the splice
junction precedes the segment of greatest homology to the crystallins, and to

small hsp of other organisms. In the human hsp gene, the second exon corres-
ponds to the region most similar to other small hsp and to the α crystallins.
The amino acid sequence encoded in exon 2 forms a highly hydrophilic domain,
which might be expected to be found on the exterior of the protein. The
conservation of this domain may reflect its importance for the interaction
of small hsp with other cellular components during stress. Analysis of the
sequence and structure of the remaining human hsp27 genes, and those of other
vertebrates, will clarify evolutionary relationships and identify the conserved
domains that are likely to be of functional significance.

*To whom correspondence should be addressed

References:
 1. Ashburner, M., and Bonner, J. J.,(1979) Cell 17:241-254
 2. Schlesinger, M., Ashburner, M., and Tissieres, A. (eds) (1982) Heat Shock:
    from Bacteria to Man Cold Spring Harbor Laboratory, N. Y.
 3. Kelley, P. M., and Schlesinger, M., (1982) Mol. Cell. Biol. 2:267-274
 4. Russnak, R. and Candido, E. P. M., (1985) Mol. Cell. Biol. 5:1268-1278
 5. Bienz, M. (1984) Proc. Nat. Acad. Sci. USA 81:3138-3142
 6. Pelham, H. R. B. (1982) Cell 30:517-528
 7. Bienz, M., (1984) EMBO J. 3:2477-2483
 8. Voellmy, R., and Rungger, D., (1982) Proc. Nat. Acad. Sci. USA 79:1776-1780
 9. Petersen, N. S., and Mitchell, H. K., (1981) Proc. Nat. Acad. Sci. USA 78:
    1708-1711
10. Berger, E. M., and Woodward, M. P., (1983) Exp. Cell Res. 147:437-442
11. Loomis, W. F., and Wheeler, S., (1980) Dev. Biol. 79:399-408
12. Hickey, E. D., and Weber, L. A., (1982) Biochemistry 21:1513-1521
13. Welch, W. J., (1985) J. Biol. Chem. 260:3058-3062
14. Kim, Y.-J., Shuman, J., Sette, M., and Przybyla, A. (1983) J. Cell Biol.
    96:393-400
15. Atkinson, B. G., and Pollock, M. (1982) Can. J. Biochem. 60:316-327
16. Ingolia, T. D., and Craig, E. A. (1982) Proc. Nat. Acad. Sci. USA 79:2360-
    2364
17. Southgate, R., Ayme, A., and Voellmy, R. (1983) J. Mol. Biol. 165:35-57
18. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., and Maniatis, T.
    (1978) Cell 15:1157-1174
19. Hickey, E., Brandon, S. E., Sadis, S., Smale, G., and Weber, L. A. (1985)
    GENE (in press)
20. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning,
    A Laboratory Manual Cold Spring Harbor Laboratory, N. Y.
21. Southern, E. (1975) J. Mol. Biol. 98:503-517
22. Berk, A. J., and Sharp, P. A. (1978) Proc. Nat. Acad. Sci. USA 75:1274-
    1278
23. Weaver, R. F., and Weissman, C. (1979) Nucl. Acids Res. 7:1175-1193
24. Favaloro, J., Treisman, R., and Kamen, R. (1980) Methods Enzymol. 65:718-
    749
25. Lee, D. C., and Luse, D. S. (1982) BRL Focus 4:1-3
26. Manley, J. L., Fire, A., Samuels, M., and Sharp, P. A. (1983) Methods
    Enzymol. 101:568-582

27. Messing, J. (1983) Methods Enzymol. 101:20-78
28. Schwindiger, W. F., and Warner, J. R. (1984) Nucl. Acids Res. 12:601-604
29. Lagrimini, L. M., Brentano, S. T., and Donelson, J. E. (1984) Nucl. Acids Res. 12:605-614
30. Cole, C. N., and Stacy, T. P. (1985) Mol. Cell. Biol. 5:2104-2113
31. Dynan, W. S., and Tjian, R. (1985) Nature 316:774-778
32. Deininger, P., Jolly, D., Rubin, C., Friedman, T., and Schmid, C. (1981) J. Mol. Biol. 151:17-33
33. Perez-Stable, C., Ayres, T. M., and Shen, C.-K. J. (1984) Proc. Nat. Acad. Sci. USA 81:5291-5295
34. Czarnecka, E., Gurley, W. B., Nagao, R. T., Mosquera, L. A., and Key, J. (1985) Proc. Nat. Acad. Sci. USA 82:3726-3730
35. Hackett, R. W., and Lis, J. (1981) Proc. Nat. Acad. Sci. USA 78:6196-6200
36. Craig, E. A., and McCarthy, B. J. (1980) Nucl. Acids Res. 8:4441-4457
37. Corces, V., Holmgren, R., Freund, R., Morimoto, R., and Meselson, M. (1980) Proc. Nat. Acad. Sci. USA 77:5390-5393
38. Voellmy, R., Goldschmidt-Clermont, M., Southgate, R., Tissieres, A., Levis, R., and Gehring, W. (1981) Cell 23:261-271
39. Ingolia, T. D., and Craig, E. A. (1982) Proc. Nat. Acad. Sci. USA 79:525-529
40. Mayrand, S., and Pederson, T. (1983) Mol. Cell. Biol. 3:161-171
41. Voellmy, R., Ahmad, A. Schiller, P., Bromley, P., and Rungger, D. (1985) Proc. Nat. Acad. Sci. USA 82:4949-4953
42. Ingolia, T. D., and Craig, E. A. (1981) Nucl. Acids Res. 9:1627-1640
43. Holmgren, R., Corces, V., Morimoto, R., Blackman, R., and Meselson, M. (1981) Proc. Nat. Acad. Sci. USA 78:3775-3778
44. Amin, J., Mestril, R., Lawson, R., Klapper, H., and Voellmy, R. (1985) Mol. Cell. Biol. 5:197-203
45. Dudler, R., and Travers, A. (1984) Cell 38:391-398
46. Efstradiatis, A., Posakony, J. W., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., DeRiel, J. K., Forget, B., Weissman, S., Slighton, J., Blechl, A., Smithies, O., Baralle, F., Shoulders, C., and Proudfoot, N. (1980) Cell 21:653-668
47. Frankis, R., and Paddock, G. V. (1982) J. Mol. Biol. 157:681-686
48. Proudfoot, N., Shander, M., Manley, J., Gefter, M., and Maniatis, T. (1980) Science 209:1329-1335
49. Leder, P., Hansen, J. N., Konkel, D., Leder, A., Nishioka, Y., and Talkington, C. (1980) Science 209:1336-1342
50. Hollis, G.F., Hieter, P., McBride, O. W., Swan, D., and Leder, P. (1982) Nature 296:321-325
51. Kim, Y.-J., Shuman, J., Sette, M., and Przybyla, A. (1984) Mol. Cell. Biol. 4:468-474
52. Rodenhiser, D., Jung, J. H., and Atkinson, B. G. (1985) Can. J. Biochem. 63:711-722
53. Quax-Jeuken, Y., Quax, W., van Rens, G., Khan, P. M., and Bloemendal, H. (1985) Proc. Nat. Acad. Sci. USA 82:5819-5823