At elevated temperatures, heat shock protein genes show altered ratios of different RNAs and expression of new RNAs, including several novel HSPB1 mRNAs encoding HSP27 protein isoforms

XIA GAO[1,2], KEYIN ZHANG[1,2], HAIYAN ZHOU[3], LUCAS ZELLMER[4],
CHENGFU YUAN[5], HAI HUANG[6] and DEZHONG JOSHUA LIAO[2,6]

[1]Department of Pathology, Guizhou Medical University Hospital; [2]Key Lab of Endemic
and Ethnic Diseases of The Ministry of Education of China in Guizhou Medical University;
[3]Clinical Research Center, Guizhou Medical University Hospital, Guiyang, Guizhou 550004, P.R. China;
[4]Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA; [5]Department of Biochemistry,
China Three Gorges University, Yichang, Hubei 443002; [6]Center for Clinical Laboratories,
Guizhou Medical University Hospital, Guiyang, Guizhou 550004, P.R. China

Abstract. Heat shock proteins (HSP) serve as chaperones to maintain the physiological conformation and function of numerous cellular proteins when the ambient temperature is increased. To determine how accurate the general assumption that HSP gene expression is increased in febrile situations is, the RNA levels of the HSF1 (heat shock transcription factor 1) gene and certain HSP genes were determined in three cell lines cultured at 37˚C or 39˚C for three days. At 39˚C, the expression of HSF1, HSPB1, HSP90AA1 and HSP70A1L genes demonstrated complex changes in the ratios of expression levels between different RNA variants of the same gene. Several older versions of the RNAs of certain HSP genes that have been replaced by a newer version in the National Center for Biotechnology Information database were also detected, indicating that the older versions are actually RNA variants of these genes. The present study cloned four new RNA variants of the HSP27-encoding HSPB1 gene, which together encode three short HSP27 peptides. Reanalysis of the proteomics data from our previous studies also demonstrated that proteins from certain HSP genes could be detected simultaneously at multiple positions using SDS-PAGE, suggesting that these genes may engender multiple protein isoforms. These results collectively suggested that, besides increasing their expression, certain HSP and associated genes also use alternative transcription start sites to produce multiple RNA transcripts and use alternative splicing of a transcript to produce multiple mature RNAs, as important mechanisms for responding to an increased ambient temperature *in vitro*.

Introduction

Heat shock proteins (HSPs) are encoded by a large family of genes (1-3). The essential function of these proteins is to serve as molecular chaperones that maintain numerous cellular proteins in their normal three-dimensional structures, in order to continue performing their normal functions when the ambient temperature is increased (1-3). However, the majority of heat shock proteins also have numerous other functions and are often overexpressed in various situations of cellular stress (1,2,4-7). For example, when normal cells evolve into cancer cells, the expression of HSPs is often increased, which is usually a mechanism by which cancer cells become resistant to various therapies (5,7).

It is well known that the majority of genes in the human, mouse or rat genome can be expressed as multiple protein isoforms to suit a variety of developmental, physiological or pathological situations (8-11). As we have previously described (10-14), multiple mechanisms for the protein multiplicity exist, including alternative use of transcription start sites to yield different RNA transcripts, alternative use of splicing sites of the same RNA transcript to produce different mRNA variants, and alternative use of different open reading frames (ORFs) of the same mRNA for protein translation. In our previous studies (12-14), proteins were fractionated from several human cell lines using SDS-PAGE, proteins were isolated at certain positions on the gels, and the isolated proteins were identified using a routine liquid chromatography with tandem mass spectrometry (LC-MS/MS) approach.

Correspondence to: Dr Chengfu Yuan, Department of Biochemistry, China Three Gorges University, 8 Daxue Road, Yichang, Hubei 443002, P.R. China
E-mail: yuancf46@ctgu.edu.cn

Dr Dezhong Joshua Liao, Center for Clinical Laboratories, Guizhou Medical University Hospital, 4 Beijing Road, Guiyang, Guizhou 550004, P.R. China
E-mail: djliao@gmc.edu.cn

Notably, it was found that proteins of numerous genes were not supposed to be detected at these positions due to their theoretical molecular masses being either too large or too small for them to appear at the SDS-PAGE positions. Furthermore, proteins produced from the same genes could be simultaneously detected at two or more SDS-PAGE positions. These data suggested that numerous genes may have additional protein isoform(s) in addition to the wild-type (Wt) or the canonical proteins (12-14).

To determine how HSP genes respond to heat stress *in vitro* at the RNA level, the RNA expression of eight arbitrarily selected HSP genes was detected in three human cell lines cultured at 39˚C for three days. This is a less febrile temperature, but severe enough for the majority of patients to decide to see a doctor. As numerous HSP genes are transcriptionally activated by the heat shock transcription factor 1 (HSF1) protein (4,15), the RNA expression of the HSF1 gene was also studied. As some of these genes have multiple RNA variants listed in the National Center for Biotechnology Information (NCBI) database, certain variants other than the Wt or the canonical form were also studied. The results demonstrated that the HSF1 and several HSP genes yielded multiple RNA variants, some of which demonstrate changes in the expression levels when compared with their counterparts at 37˚C. A total of four new mRNA variants of the HSP27-encoding HSPB1 gene that encode three short HSP27 isoforms were identified. Furthermore, reanalysis of certain previous proteomics data also revealed that proteins of certain HSP genes may be detected simultaneously at multiple positions on SDS-PAGE, suggesting that these genes may express multiple protein isoforms.

Materials and methods

Cell lines and cell culture for RNA analyses. The human embryonic kidney (HEK) 293 cell line, the human cervical cancer HeLa cell line and the non-small cell lung cancer H1650 cell line were obtained from the American Type Culture Collection (ATCC). The cells were cultured in water-jacket incubators at 5% CO_2 that were recalibrated regularly to ensure accuracy. All three cell lines were cultured at 37˚C or 39˚C in a Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. A thermometer was placed in the incubator to monitor and guarantee the accuracy and stability of the temperature. After 72 h of culture, the cells reached 70-80% confluence and were harvested with TRIzol® (Invitrogen; Thermo Fisher Scientific, Inc.; cat. no. 15596-026) for RNA isolation.

Reverse transcription-polymerase chain reaction (RT-PCR) assay. Immediately after cells were lysed in TRIzol reagent, total RNA samples were extracted from the lysates according to the instructions for the TRIzol reagent. An aliquot of the RNA was primed with random hexamers or a poly-dT primer during reverse transcription to complementary DNA (cDNA) using MMLV Reverse Transcriptase (Promega Corporation; cat. no. M1705; www.promega.com) according to the reagent manual. The cDNA was amplified using a PCR, in which the cDNA template was melted at 95˚C for 4 min, followed by 34-38 cycles of 95, 58 and 72˚C with each step for 20 sec. Each

PCR was ended by a final extension at 72˚C for 5 min. The primers were all designed to be 20-mers with AT to GC ratios around 1:1, which allowed the PCR conditions to be unified for all target amplicons. However, the number of PCR cycles needed to be optimized based on the general expression level of each individual gene, so that the reaction was terminated within the linear portion of the amplification. Each primer pair used for PCR was determined in such a way that the two primers were separated by one or more large introns to avoid amplification of extant genomic DNA. All the PCR primers used are listed in Table I. In most cases, forward (F) primers were named by calculating the distance (number) from their first nucleotide (nt) to the first nt of the RNA, whereas reverse (R) primers were named by calculating the number of nts from their last nt to the first nt of the RNA. This method of primer naming usually allows for an easy calculation of the PCR amplicon size by subtracting the F primer number from the R primer number, but this calculation will be inaccurate in cases where alternative splicing results in addition or deletion of exon(s). For semi-quantification of the expression level, cDNA loading was normalized using the HRPT1 gene.

Nested or semi-nested PCR for preliminary identification of PCR amplicons. PCR products were routinely fractionated in a 1% agarose gel that contained 1 µg/ml of ethidium bromide to visualize DNA in the gel during electrophoresis. For those RNAs that demonstrated a questionable or additional band in the gel, a semi-nested or nested PCR was performed to preliminarily determine whether the band was derived from the target gene. For such a PCR, the template DNA was obtained from the DNA band using one of the two methods we described previously (16). In one method, a small blade used for eye surgery was used to excise a small piece (1 mm³ or smaller) of the agarose with the DNA, which was put directly into a 20-µl volume of PCR reagents as the template. The DNA would be released from the gel during the first step of the PCR, which was an incubation at 95˚C for 4 min to melt the DNA, as the agarose gel melts at a temperature of approximately 87˚C. In the second method, a blade was used to make an incision in the gel immediately below the band in question, followed by insertion of a small piece of Whatman filter paper into the incision (Fig. 1). Electrophoresis was then continued for a few more minutes until the DNA had entered into the filter paper. The paper was then removed and placed into an Eppendorf tube containing 25-30 µl of a 10-mT TE [Tris(hydroxymethyl) aminomethane and Ethylene Diamine Tetraacetic Acid] buffer (pH 7.4), followed by elution of the DNA by vortexing the tube for a short time (16). A semi-nested or nested PCR was performed using one or two inner primers and 1-2 µl of the eluted DNA as the template. After the template DNA was melted at 95˚C for 4 min, PCR was allowed for 35 cycles of 95, 58 and 72˚C with each step for 20 sec, followed by a 5-min final extension at 72˚C, as its purpose was to preliminarily determine the identity, but not the expression level, of the template DNA.

T-A cloning, DNA sequencing and sequence analyses. For direct sequencing of a PCR amplicon, the desired DNA band in the agarose gel was excised and purified using a simple method previously described by us, which in

Table I. HSF1 and HSP gene primers used.

Gene	Primer name	Sequence	Primers/amplicon	Detected[a]
HSF1	HSF1F574	5'-ACAGCGTCACCAAGCTGCTG-3'	F574/R1312=739 bps	NM_005526.2; XM_005272315.1; XM_005272316.1; XR_246618.2; XM_005272317.1
	HSF1R1312	5'-TTGTCCAGGCAGGCTACGCT-3'		
	HSF1XM17R	5'-TGGCTGGACTTGGCCATGCG-3'	F574/RXM17R=642 bps	XM_005272317.1
	HSF1R1576	5'-GTGTAGTGCACCAGCTGCTT-3'	F574/R1576=824 bps; 1003 bp	XR_246618.2; NM_005526.2
	HSF1XR26	5'-TAGACATCTGTGGAGTGCGA-3'	F574/XR26=586 bps	XM_005272317.1
HSP90AA1	HSP90F459	5'-AGGAAGCCCCTCTGAAGCCT-3'	F459/R1232=774 bps	NM_001017963.2; XM011536718.2
	HSP90R1232	5'-GTCCTCACTGTGAATGATCC-3'		
	HSP90VF97	5'-GTCGCTATATAAGGCAGGCG-3'	VF97/R1232=614 bps	NM_005348.3
HSPA1A	A1AF171	5'-CTTCTCGCGGATCCAGTGTT-3'	F171/R1104=934 bps	NM_005345.5
	A1AR1104	5'-CAGGGAGTCGATCTCCAGGC-3'		
HSPA1B	A1BF146	5'-CTTGTCGCGGATCCCGTCCG-3'	F146/R1077=932 bps	NM_005346.4
	A1BR1077	5'-CAGGGAGTCGATCTCCAGGC-3'		
HSPA6	HSP6AF280	5'-GTGCGGAAAGGTTCGCGAAA-3'	F280/R1047=768 bps	NM_002155.3
	HSP6AR1047	5'-GAACCGACACATCGAAGGTG-3'		
HSPA1L	A1L-F120	5'-GCTGCGTAATCTGGACGTTT-3'	F120/R899=780 bps	NM_005527.3; XM_005249070.2
	A1L-R899	5'-AGCCTGTTGTCAAAGTCCTC-3'		
	A1LV-F79	5'-GTGCAGTTTGATATTGAGGG-3'	VF79/R899=821 bps	NM_005527.3; XM_005249070.2
HSPB1	HSPB1F19	5'-CTCAAACGGGTCATTGCCAT-3'	F19/R845=827 bps	NM_001540.3
	HSPB1R845	5'-CAAAAGAACACACAGGTGGC-3'		
	HSPB1R313	5'-AGTGTGCCGGATCTCCGAGA-3'	used in 5'-RACE	
	HSPB1F43	5'-AGAGACCTCAAACACCGCCT-3'	used in nested PCR	new variants
	HSPB1F70	5'-ATACCCGACTGGAGGAGCAT-3'		
	HSPB1R788	5'-ATCCGGGCTAAGGCTTTACT-3'		

[a]Because the primers were designed and named years ago based on an earlier version of the RNA, the size of the amplicon and the number of the RNA version may or may not be the same as the current version in the NCBI database.
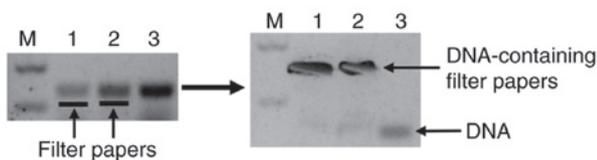
Figure 1. A depiction of one of the methods for isolating a DNA fragment from an agarose gel. A surgical blade was used to make an incision in the gel immediately below the DNA band of interest, and then a small piece of Whatman filter paper (the black dashed lines in lanes 1 and 2 in the left panel) was inserted into the incision. Electrophoresis continues for a few more minutes to allow the DNA to enter into the filter paper, while the DNA in lane 3 serves as a control for further migration (the left panel). The DNA collected in the paper is then eluted from the paper with a 1-mM or 10-mM TE buffer (pH 7.4). M, molecular weight marker.

principle isolates DNA from an agarose gel by centrifuging the DNA-containing gel slice at ~15,743.9 x g for 10 min at room temperature (17). Following being isolated from the gel,

the DNA was precipitated with ethanol and then dissolved with a small volume of water or a 1-mM or 10-mM TE buffer (pH 7.4) to ensure a high concentration of the DNA, which was sent to Genewiz, Inc. (www.genewiz.com) for direct sequencing from both ends using the gene-specific primers used for the PCR. If the data showed a mixture of different sequences, the purified DNA would be cloned into a T-A vector and then transfected into bacteria for amplification, as described in our previous studies (18-21). The desired bacterial clones resulting from the T-A cloning were selected using the 'dirty plasmid' method described previously (22); plasmid DNA was purified from each selected bacterial clone and was sequenced. DNA sequences were analyzed using a ClustalW software (https://www.genome.jp/tools-bin/clustalw), NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and UCSC's BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat). DNAStar 7.1 software (https://www.dnastar.com) was used to identify the start and stop codons, as well as the associated ORFs. Multiple

sequence alignments were performed using the Mega-X (https://megasoftware.net/) or ClustalW software and then edited using the BioEditor 7.2 software (https://bioedit.software.informer.com/7.2), followed by manual editing in a Word document.

5'-RACE assay. A total RNA sample isolated from H1650 cells cultured at 37˚C was used for 5'-RACE assays using a FirstChoice RLM-RACE kit (cat. no. LSAM1700M; Ambion Inc.) according to the manufacturer's protocols and as described previously (18,23). HSP90VF97 and HSPB1R313 were used as the reverse primers for the HSP90AA1 gene and the HSPB1 gene, respectively (Table I). Following fractionation and visualization of the 5'-RACE products in a 1% agarose gel, the band of interest was purified using the method described earlier for the purpose of T-A cloning (17). The purified DNA was directly sequenced; if the sequencing quality was unsatisfactory, the DNA would be cloned into a T-A vector and the selected bacterial clones would then be sequenced as described earlier.

SDS-PAGE and LC-MS/MS analyses. The proteomics data presented in the present study were derived from two previously reported studies, including detailed materials and methods (13,14). The present study reanalyzed the raw LC-MS/MS data from these studies. One study involved HEK293 cells and the metastatic breast cancer MDA-MB231 cell line (obtained directly from the MD Anderson Cancer Center) (14). The other study involved MB231 cells and another human breast cancer cell line, MCF7 (obtained directly from the Michigan Cancer Foundation) (13).

In the two studies, cells were cultured at 37˚C under the same conditions as described earlier. Cells at ~80% confluence were lysed using a buffer (24) containing 1x protease inhibitor cocktail (Sigma-Aldrich; Merck KGaA) (13,14), followed by centrifugation at 18,892.7 x g for 20 min at 4˚C. The supernatant was collected as the protein sample and its concentration was determined using bicinchoninic acid reagents (Pierce; Thermo Fisher Scientific, Inc.). The proteins were diluted with a 5X Western blotting loading buffer (250 mM Tris-HCl, 500 mM DTT, 10% SDS, 2% 2-mercaptoethanol, 50% glycerol, and 0.5% bromophenol blue). After boiling for 4 min, the proteins were loaded at 50 μg (14) or 70 μg (13) per lane into a 10% SDS-PAGE gel that was 2-cm longer than other gels made using ordinary mini-gel casting systems (13,14). Electrophoresis was then performed to fractionate the proteins. In one study, gel strips approximately 2-mm wide were excised at the 26-kDa and 40-kDa positions, respectively, guided by pre-staining protein markers loaded also into the gel (14,19). Similarly, in another study gel strips were excised at the 72-kDa, 55-kDa and 48-kDa positions (13). Inside the gel strips, the proteins were dehydrated using acetonitrile (ACN), reduced, alkylated with 10 mM dithiothreitol and 55 mM iodoacetamide, and then digested to peptides using trypsin. The peptides were extracted with ACN containing 0.1% formic acid (FA), and were then vacuum-dried. The peptides were dissolved in 0.1% FA, loaded into a nano RP column (5-μm Hypersil C18, 75 mm x 100 mm; Thermo Fisher Scientific, Inc.), and eluted using ACN. Different eluted fractions were delivered into a Q-Exactive mass spectrometer (Thermo Fisher Scientific, Inc.)
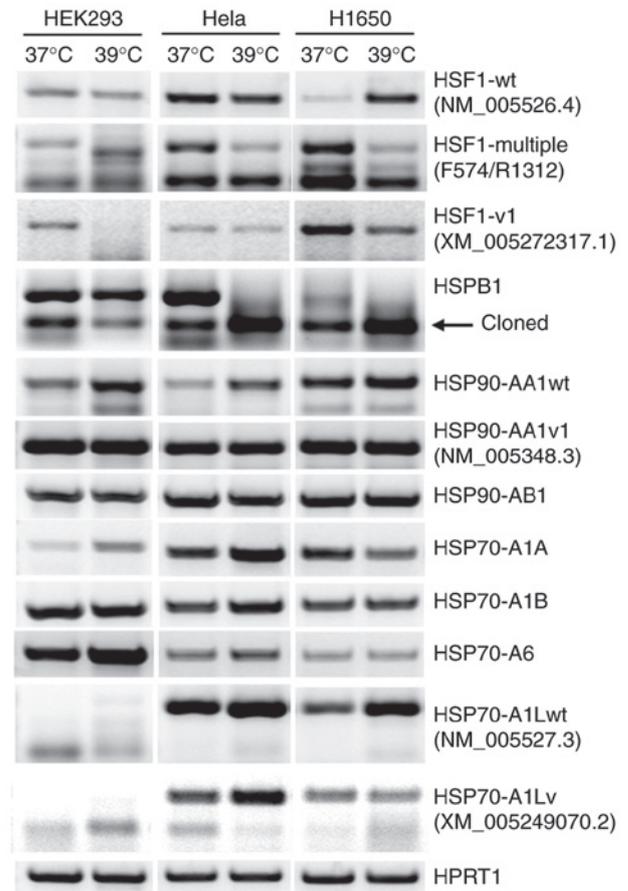
Figure 2. Reverse transcription-polymerase chain reaction amplicons of RNAs from HSF1 and several HSP genes visualized in agarose gels, with the HPRT1 gene as a cDNA loading control. The arrow indicates the band of excised and purified HSPB1 for T-A cloning. Wt, wild-type; v or v1, variant or variant 1.

as described previously (13,14). With the higher-energy collisional dissociation (HCD) as the MS/MS acquisition method, raw MS/MS data were converted into an MGF format using Proteome Discoverer 1.2 (Thermo Fisher Scientific, Inc.). With Mascot v2.3.01 (www.matrixscience.com), the MGF files were matched to the human SwissProt database for protein identification.

Results

Complicated changes in the ratios of expression levels between RNA variants of HSP genes and HSF1 at 39˚C. As shown in Fig. 2, RT-PCR detected the RNAs of the HSP90AB1, HSP70A1A, HSP70A1B and HSP70A6 genes as a single band at the anticipated molecular weights indicated in Table I. However, one or more additional bands of the HSF1, HSPB1, HSP90AA1 and HSP70A1L genes were also detected (Fig. 2).

The NCBI database currently lists 11 HSF1 RNAs. The NM_005526.4 RNA is the only NM (normalized) sequence that is designated herein as the Wt. Whether the remaining 10 RNAs really exist remains unknown, as they are either XM (predicted mRNA) or XR (predicted non-coding RNA) sequences. The primers used in the present study were designed based on a previous version of the NM sequence (NM_005526.2). RT-PCR using the primer pair F574/R1576

Figure 3. Alignment of the first exon shown in all four versions of the NM_005348 mRNA of the HSP90AA1 gene. This exon has 138 nts in version 3 (NM_005348.3), which has been confirmed by a 5'-RACE assay, but it had only 60 nts in the first and second versions and has only 59 nts in the current version (NM_005348.4), due to the lack of the nt C (shaded) present in the three previous versions. The reverse transcription-PCR with the primer pair VF97 (underlined)/R1232 detected the band shown in Fig. 2; sequencing of the PCR product and a 5'-RACE assay confirmed the existence of the VF97 sequence therein. PCR, polymerase chain reaction; nt, nucleotide.

detected the Wt RNA, which was further confirmed by direct sequencing of the PCR amplicon. Notably, this RNA exhibited a markedly increased expression at 39˚C only in H1650 cells, while exhibiting a slight decrease in expression in HEK293 and Hela cells. With the primer pair F574/R1312 designed to amplify several RNA variants, RT-PCR produced multiple bands that likely included, besides possible artifacts, some of the predicted RNAs or their alternative splicing variants that are currently unknown. Notably, the responses of these PCR amplicons to the increased temperature were quite different from each another and from one cell line to another, and the results were too complex to distinguish one amplicon from another (Fig. 2). RT-PCR using primer pair F574/XR26 resulted in a single band, which was confirmed by direct sequencing as XM_005272317.1 RNA (designated herein as variant 1). This RNA showed different extents of decrease in the three cell lines (Fig. 2). In all three cell lines, the HSP90AA1 gene demonstrated increased expression in its Wt form, while the expression of its variant remained unchanged (Fig. 2).

The first exon of the current version (NM_005348.4) of HSP90AA1 mRNA is much shorter than its counterpart in a previous version (NM_005348.3), and lacks the sequence of the primer, FV97 (Table I), which was designed several years ago. However, RT-PCR using the primer pair VF97/R1232 yielded an amplicon whose size was anticipated based on the NM_005348.3. Direct sequencing of this amplicon confirmed that it contained the VF97 sequence, and the 5'-RACE results further confirmed that its 5'-end matched with that of the NM_005348.3 (Fig. 3). These data together indicated that although the current version is also authentic and should be designated as Wt RNA, the previous version remains as a variant despite having been abandoned by NCBI.

Identification of four HSPB1 RNA variants that encode HSP27 protein isoforms. At present, only one RNA of the HSPB1 gene is listed in the NCBI database. Its current version is NM_001540.5 (Fig. 4), which lacks the first 115 nts of a previous version (NM_001540.3) that contains three of our forward primers (F19, F43 and F70) designed a few years ago (Table I). However, in the majority of the samples studied, RT-PCR using the primer pair F19/R845 yielded not only the anticipated band, but also an additional band below it, which was ~400 base pairs smaller (Fig. 2). Compared with its counterpart at 37˚C, this smaller amplicon showed decreased expression in HEK293 cells but increased expression in HeLa and H1650 cells at 39˚C. Below this smaller amplicon is another weak and even smaller band in the HEK293 and HeLa

cells cultured at 37˚C. Nested RT-PCRs with inner primer pairs F43-R778 and F70-R778 (Table I) may continue to result in these two smaller bands (data not shown), suggesting that they may be unreported HSPB1 RNA variants.

The second band was excised from H1650 cells cultured at 39˚C from an agarose gel, the DNA was purified and it was cloned into a T-A vector. Sequencing over 20 resultant plasmid clones yielded five different HSPB1 RNA sequences, each of which appeared in at least two clones, including the Wt RNA and four previously unreported variants coined as 716A1-15 (MW881014), 716A1-16 (MW881015), 716A1-20 (MW881016) and cloned sequence-1 (MW881017) (Fig. 5). These RT-PCR and sequencing results involving F19, F43 and F70 primers suggested that NM_01540.3 RNA is actually a variant of HSPB1 RNA relative to NM_001540.5, and that the cloned band is a mixture of different heterodimers formed between these five HSPB1 RNAs. A 5'-RACE assay using R313 as the reverse primer also confirmed that the 5'-end of the HSPB1 RNAs is identical to that of the NM_001540.3 sequence. Furthermore, NM_01540.5 lacks the last four nts and the poly-A tail encompassed by NM_01540.3, but the PCR using cDNA samples primed by a poly-dT primer detected the Wt form and the cloned band, which also supported the conclusion that NM_01540.3 is an authentic RNA variant.

The four new HSPB1 RNA variants were lacking a part of the middle region of the NM_01540.3 sequence (Fig. 4). The 716A1-15 RNA has a deletion of 426 nts, including 211 nts from exon 1, the whole 64-nt exon 2 and 151 nts from exon 3. The 716A1-16 RNA has a deletion of 431 nts, including 293 nts from exon 1, the whole 64-nt exon 2 and 74 nts from exon 3. The 716A1-20 RNA has 404 nts deleted, including 377 nts from exon 1 and 27 nts from exon 2, whereas the cloned sequence-1 has 459 nts deleted, including 436 nts from exon 1 and 23 nts from exon 2.

The ORF encoding the Wt HSP27 protein in NM_001540.3 has two additional short ORFs encoding non-HSP27 peptides (Figs. 5 and 6). Furthermore, there is an additional ORF that begins with an ATG start codon at the 3'-end, but extends beyond the stop codon of the ORF for the HSP27 (Fig. 5). The peptide sequences encoded by these short ORFs are shown in Fig. 6. NM_001540.5 lacks the first 115 nts and thus lack the first short ORF (Fig. 5). All four novel RNAs lack the canonical ATG for translation of the Wt HSP27 protein. However, deletion of the middle region leads to the formation of one or two new ORFs that encode new HSP27 isoforms containing either the N- or C-terminal region of the HSP27 protein. The 716A1-15 and the cloned sequence-1 encode the

```
NM_001540.3  GCATGGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAGGAGCATAAAAGCGCAGCCGAGCCCAGCGCCCCGCACT
716A1-15     GCATGGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAGGAGCATAAAAGCGCAGCCGAGCCCAGCGCCCCGCACT
716A1-16     GCATGGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAGGAGCATAAAAGCGCAGCCGAGCCCAGCGCCCCGCACT
716A1-20     GCATGGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAGGAGCATAAAAGCGCAGCCGAGCCCAGCGCCCCGCACT
Sequence-1   GCATGGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAG------------------------------------
NM_001540.5  --------------------------------------------------------------------------------------------------------------GCACT

NM_001540.3  TTTCTGAGCAGACGTCCAGAGCAGAGTCAGCCAGCAATGACCGAGCGCCGCGTCCCCTTCTCGCTCCTGCGGGGCCCCAGCTGGGACCCCTTCCGCGACTGGTACCCGCATAGCCGCCTCT
716A1-15     TTTCTGAGCAGACGTCCAGAGCAGAGTCAGCCAGCAATGACCGAGCGCCGCGTCCCCTTCTCGCTCCTGCGGGGCCCCAGCTGGGACCCCTTCCGCGACTGGTACCCGCATAGCCGCCTCT
716A1-16     TTTCTGAGCAGACGTCCAGAGCAGAGTCAGCCAGCAATGACCGAGCGCCGCGTCCCCTTCTCGCTCCTGCGGGGCCCCAGCTGGGACCCCTTCCGCGACTGGTACCC--------------
716A1-20     TTTCTGAGCAGACGTCCAGAGC--------------------------------------------------------------------------------------------------
Sequence-1   ----------------------------------------------------------------------------------------------------------------------
NM_001540.5  TTTCTGAGCAGACGTCCAGAGCAGAGTCAGCCAGCAGACCGAGCGCCGCGTCCCCTTCTCGCTCCTGCGGGGCCCCAGCTGGGACCCCTTCCGCGACTGGTACCCGCATAGCCGCCTCT

NM_001540.3  TCGACCAGGCCTTCGGCTGCCCCGGCTGCCGGAGGAGTGGTCGCAGTGGTTAGGCGGCAGCAGCTGGCCAGGCTACGTGCGCCCCCTGCCCCCCGCCGCCATCGAGAGCCCCGCAGTGG
716A1-15     TCGACCAGGCCTTCGGCTGCCCCGGCTGCCGGAGGAGTGGTCGCAGTGGTTAGGCGGCAGCAGCTGGCCAG-------------------------------------------------
716A1-16     ----------------------------------------------------------------------------------------------------------------------
716A1-20     ----------------------------------------------------------------------------------------------------------------------
Sequence-1   ----------------------------------------------------------------------------------------------------------------------
NM_001540.5  TCGACCAGGCCTTCGGCTGCCCCGGCTGCCGGAGGAGTGGTCGCAGTGGTTAGGCGGCAGCAGCTGGCCAGGCTACGTGCGCCCCCTGCCCCCCGCCGCCATCGAGAGCCCCGCAGTGG

NM_001540.3  CCGCGCCCGCCTACAGCCGCGCGCTCAGCCGGCAACTCAGCAGCGGGGTCTCGGAGATCCGGCACACTGCGGACCGCTGGCGCGTGTCCCTGGATGTCAACCACTTCGCCCCGGACGAGC
716A1-15     ----------------------------------------------------------------------------------------------------------------------
716A1-16     ----------------------------------------------------------------------------------------------------------------------
716A1-20     ----------------------------------------------------------------------------------------------------------------------
Sequence-1   ----------------------------------------------------------------------------------------------------------------------
NM_001540.5  CCGCGCCCGCCTACAGCCGCGCGCTCAGCCGGCAACTCAGCAGCGGGGTCTCGGAGATCCGGCACACTGCGGACCGCTGGCGCGTGTCCCTGGATGTCAACCACTTCGCCCCGGACGAGC

NM_001540.3  TGACGGTCAAGACCAAGGATGGCGTGGTGGAGATCCACGGCAAGCACGAGGAGCGGCAGGACGAGCATGGCTACATCTCCCGGTGCTTCACGCGGAAATACACGCTGCCCCCCGGTGTGG
716A1-15     ----------------------------------------------------------------------------------------------------------------------
716A1-16     ----------------------------------------------------------------------------------------------------------------------
716A1-20     ---------------------------------------------------------------------ATGGCTACATCTCCCGGTGCTTCACGCGGAAATACACGCTGCCCCCCGGTGTGG
Sequence-1   -------------------------------------------------------------------GAGCATGGCTACATCTCCCGGTGCTTCACGCGGAAATACACGCTGCCCCCCGGTGTGG
NM_001540.5  TGACGGTCAAGACCAAGGATGGCGTGGTGGAGATCCACGGCAAGCACGAGGAGCGGCAGGACGAGCATGGCTACATCTCCCGGTGCTTCACGCGGAAATACACGCTGCCCCCCGGTGTGG

NM_001540.3  ACCCCACCCAAGTTTCCTCCTCCCTGTCCCCTGAGGGCACACTGACCGTGGAGGCCCCCATGCCCAAGCTAGCCACGCAGTCCAACGAGATCACCATCCCAGTCACCTTCGAGTCGCGGG
716A1-15     ----------------------------------------------------------------------------------------------------------------------
716A1-16     ---------------------------------------------------------------CCATGCCCAAGCTAGCCACGCAGTCCAACGAGATCACCATCCCAGTCACCTTCGAGTCGCGGG
716A1-20     ACCCCACCCAAGTTTCCTCCTCCCTGTCCCCTGAGGGCACACTGACCGTGGAGGCCCCCATGCCCAAGCTAGCCACGCAGTCCAACGAGATCACCATCCCAGTCACCTTCGAGTCGCGGG
Sequence-1   ACCCCACCCAAGTTTCCTCCTCCCTGTCCCCTGAGGGCACACTGACCGTGGAGGCCCCCATGCCCAAGCTAGCCACGCAGTCCAACGAGATCACCATCCCAGTCACCTTCGAGTCGCGGG
NM_001540.5  ACCCCACCCAAGTTTCCTCCTCCCTGTCCCCTGAGGGCACACTGACCGTGGAGGCCCCCATGCCCAAGCTAGCCACGCAGTCCAACGAGATCACCATCCCAGTCACCTTCGAGTCGCGGG

NM_001540.3  CCCAGCTTGGGGGGCCCAGAAGCTGCAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCCTCCCCCGCCACCTGTGTGTTC
716A1-15     -------------------AAGCTGCAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCCTCCCCCGCCACCTGTGTGTTC
716A1-16     CCCAGCTTGGGGGGCCCAGAAGCTGCAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCCTCCCCCGCCACCTGTGTGTTC
716A1-20     CCCAGCTTGGGGGGCCCAGAAGCTGCAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCCTCCCCCGCCACCTGTGTGTTC
Sequence-1   CCCAGCTTGGGGGGCCCAGAAGCTGCAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCCTCCCCCGCCACCTGTGTGTTC
NM_001540.5  CCCAGCTTGGGGGGCCCAGAAGCTGCAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCCTCCCCCGCCACCTGTGTGTTC

NM_001540.3  TTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACCTGTCAAAAAAAAAAAAAAAA
716A1-15     TTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACCTGTCAAAAAAAAAAAAAAAA
716A1-16     TTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACCTGTCAAAAAAAAAAAAAAAA
716A1-20     TTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACCTGTCAAAAAAAAAAAAAAAA
Sequence-1   TTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACCTGTCAAAAAAAAAAAAAAAA
NM_001540.5  TTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACC--------------------
```

Figure 4. Alignment of the four assembled new HSPB1 RNA variants with the NM_01540.3 and NM_01540.5 sequences. The four new sequences have been assembled with sequences prior to F19 (confirmed with a 5'-RACE assay) and after R788 based on NM_01540.3. The ATG start codon and TAA stop codon of the ORF encoding the Wt HSP27 protein are boldfaced, italicized and underlined with a wavy line. The 64-nt exon 2 of NM_001540.3 is underlined to separate each exon. The deletion of part of exon 2 in the 716A1-20 and the cloned sequence-1 leads to the translation initiation of HSP27 from a downstream in-frame ATG (shaded), engendering a peptide containing the last 37 AAs of the HSP27. Wt, wild-type.

same HSP27 isoform, whereas the 7161A-16 encodes two isoforms (Figs. 5 and 6). Furthermore, each of these four novel RNAs encodes several other short ORFs, some of which also appear in the NM_001540.3 RNA.

Detection of proteins from the same HSP genes at multiple positions of SDS-PAGE. If HSP genes respond to increased ambient temperatures primarily by changes in the ratios of expression levels between different RNA variants (including some new ones), they may express multiple protein isoforms. Therefore, the proteomics data from two previous studies was reanalyzed to determine whether certain HSP or associated genes express multiple protein isoforms. In one study, proteins from 12 HSP genes were detected in MB231 and MCF7 cells at the 72-kDa position of SDS-PAGE (Table II), of which the HSP90AA1, STIP1, SHPA8, HSP90AB1, TRAP1 and HSPH1 genes have multiple protein isoforms listed in the NCBI database, with their numbers of amino acids (AAs) presented in Table II. The largest and smallest isoforms of these proteins

range between approximately 98 kDa (HSP90AA1) and 53.5 kDa (HSPA8). However, it remains unknown whether the largest or the smallest are truly expressed and detected because, following excision from the SDS-PAGE gel at the indicated position (72-kDa), our approach uses short peptide(s) to predict the existence of the proteins. Notably, several detected genes have only one protein form listed in the NCBI database, which is too small to be detected at 72-kDa. For instance, the HSPD1 gene was detected, but NCBI demonstrates that its protein has only 573 AAs (less than 61.1 kDa), indicating that either HSPD1 has an unidentified larger isoform or the canonical protein has undergone substantial post-translational modifications that greatly hinder its migration in the SDS-PAGE, as previously inferred (12-14).

The protein products of 9 to 13 HSP or associated genes that were detected at 72-kDa were also detected at the 55-kDa and/or 48-kDa positions in MB231 or MCF7 cells. Notably, the majority of these genes have isoforms much larger than 55 or 48 kDa and were already detected at the 72-kDa position

>NM_001540.3
GCATGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAGGAGCATAAAAGCGGCAGCCGAGCCCAGCGGC
CCCGCACTTTTCTGAGCAGACGTCCAGAGCAGAGTCAGCCAGCCATGACCGAGCGCCGCGTCCCCTTCTCGCTCCTGCGGGGGCCCCAGCTGGGACCCCTTCCGCGACTGGTAC
CCGCATAGCCGCCTCTTCGACCCAGGCCTTCGGGCTGCCCCGGCTGCCGGAGGAGTGGTGCGCAGTGGTTAGGCGGCAGCAGCTGGCCAGGCTACGTGCGCCCCCTGCCCCCCG
CCGCCATCGAGAGCCCCGCAGTGGCCGCGCCCGCCTACAGCCGCGCGCTCAGCCGGCAACTCAGCAGCGGGGTCTCGGAGATCCGGCACACTGCGGACCGCTGGCGCGGTGTC
CCTGGATGTCAACCACTTCGCCCCGGACGAGCTGACGGTCAAGACCAAGGATGGCGTGGTGGAGATCACCGGCAAGCACGAGGAGCGGCCAGGACGAGCATGGCTACATCTCC
CGGGTGCTTCACGCGGGAAATACACGCTGCCCCCCGGTGTGGACCCCACCCAAGTTTCCTCCTCCTGTCCCCTGAGGGCACACTGACCGTGGAGGCCCCCATGCCCAAGCTAG
CCACGCAGTCCAACGAGATCACCCATCCCAGTCACCTTGAGTCGCGGGCCCAGCTTGGGGGCCCAGAAGCTGCAAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCC
GGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCTCCCCCGCCCACCTGTGTGGTTCTTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACCTGTC
AAAAAAAAAAAAAAAAAA

>NM_001540.5
GCACTTTTCTGAGCAGACGTCCAGAGCAGAGTCAGCCAGCCATGACCGAGCGCCGCGTCCCCTTCTCGCTCCTGCGGGGGCCCCAGCTGGGACCCCTTCCGCGACTGGTACCCG
CATAGCCGCCTCTTCGACCCAGGCCTTCGGGCTGCCCCGGCTGCCGGAGGAGTGGTGCGCAGTGGTTAGGCGGCAGCAGCTGGCCAGGCTACGTGCGCCCCCTGCCCCCCGCCG
CCATCGAGAGCCCCGCAGTGGCCGCGCCCGCCTACAGCCGCGCGCTCAGCCGGCAACTCAGCAGCGGGGTCTCGGAGATCCGGCACACTGCGGACCGCTGGCGCGGTGTCCCT
GGATGTCAACCACTTCGCCCCGGACGAGCTGACGGTCAAGACCAAGGATGGCGTGGTGGAGATCACCGGCAAGCACGAGGAGCGGCCAGGACGAGCATGGCTACATCTCCCGG
TGCTTCACGCGGGAAATACACGCTGCCCCCCGGTGTGGACCCCACCCAAGTTTCCTCCTCCTGTCCCCTGAGGGCACACTGACCGTGGAGGCCCCCATGCCCAAGCTAGCCA
CGCAGTCCAACGAGATCACCCATCCCAGTCACCTTCGAGTCGCGGGCCCAGCTTGGGGGCCCAGAAGCTGCAAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGA
TGCCCACCCCTGCTGCCGCCACTGGCTGTGCTCCCCCGCCCACCTGTGTGGTTCTTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACC

>20160716A1-20
GCATGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACGCCGCCTGCTAAAAATACCCGACTGGAGGAGCACAAAAGCGGCAGCCGAGCCCAGCGGC
CCCGCACTTTTCTGAGCAGACGTCCAGGGCATGGCTACATCTCCCGGTGCTTCACGCGGGAAATACACGCTGCCCCCCGGTGTGGACCCCACCCAAGTTTCCTCCTCCTGTC
CCCTGAGGGCACACTGACCGTGGAGGCCCCCATGCCCAAGCTAGCCACGCAGTCCAACGAGATCACCCATCCCAGTCACCTTGAGTCGCGGGCCCAGCTTGGGGGCCCAGAA
GCTGCAAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCCTCCCCCGCCACCTGTGTGTTCTTTTGATACATT
TATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACCTGTCAAAAAAAAAAAAAAAAAAAA

>Cloned sequence-1
GCATGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAGGAGCATGGCTACATCTCCCGGTGCTTCAC
GCGGGAAATACACGCTGCCCCCCGGTGTGGACCCCACCCAAGTTTCCTCCTCCTGTCCCCTGAGGGCACACTGACCGTGGAGGCCCCCATGCCCAAGCTAGCCACGCAGTCC
AACGAGATCACCCATCCCAGTCACCTTCGAGTCGCGGGCCCAGCTTGGGGGCCCAGAAGCTGCAAAAATCCGATGAGACTGCCGCCAAGTAAAGCCTTAGCCCGGATGCCCACC
CCTGCTGCCGCCACTGGCTGTGCTCCCCCGCCCACCTGTGTGTTCTTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTTCAAAGCAACCACCTGTCAAAAAAAAAAAA
AAAAAAA

>20160716A1-15
GCATGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAGGAGCATAAAAGCGGCAGCCGAGCCCAGCGGC
CCCGCACTTTTCTGAGCAGACGTCCAGAGCAGAGTCAGCCAGCCATGACCGAGCGCCGCGTCCCCTTCTCGCTCCTGCGGGGGCCCCAGCTGGGACCCCTTCCGCGACTGGTAC
CCGCATAGCCGCCTCTTCGACCCAGGCCTTCGGGCTGCCCCGGCTGCCGGAGGAGTGGTGCGCAGTGGTTAGGCGGCAGCAGCTGGCCAGAAGCTGCAAAAATCGATGAGACTG
CCGCCAAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCTCCCCCGCCACCTGTGTGTTCTTTTGATACATTTATCTTCTGTTTTTCTCAAATA
AAGTTCAAAGCAACCACCTGTCAAAAAAAAAAAAAAAAAAAA

>20160716A1-16
GCATGGGAGGGGCGGCCCTCAAACGGGTCATTGCCATTAATAGAGACCTCAAACACCGCCTGCTAAAAATACCCGACTGGAGGAGCATAAAAGCGGCAGCCGAGCCCAGCGGC
CCCGCACTTTTCTGAGCAGACGTCCAGAGCAGAGTCAGCCAGCCATGACCGAGCGCCGCGTCCCCTTCTCGCTCCTGCGGGGGCCCCAGCTGGGACCCCTTCCGCGACTGGTAC
CCCCATGCCCAAGCTAGCCACGCAGTCCAACGAGATCACCCATCCCAGTCACCTTCGAGTCGCGGGCCCAGCTTGGGGGCCCAGAAGCTGCAAAAATCCGATGAGACTGCCGCC
AAGTAAAGCCTTAGCCCGGATGCCCACCCCTGCTGCCGCCACTGGCTGTGCTCCCCCGCCACCTGTGTGTTCTTTTGATACATTTATCTTCTGTTTTTCTCAAATAAAGTT
CAAAGCAACCACCTGTCAAAAAAAAAAAAAAAAAAAA

Figure 5. ORFs of NM_001540.3, NM_001540.5 and the four HSPB1 RNA variants cloned in the present study. Each ORF is shaded, underlined, or boldfaced and italicized. The longest ORF in NM_001540.3 encodes the Wt HSP27 protein (the shaded 618 nts), which contains two short ORFs encoding non-HSP27 peptides. Two new variants we cloned (i.e. the 716A1-20 and cloned sequence-1) have the same ORFs. None of the four new variants encodes an ORF longer than 200 nts. ORFs, open reading frames; Wt, wild-type; nt, nucleotide.

(Table II), implying that they are detected simultaneously at two or three positions on SDS-PAGE. On the other hand, proteins from the HSPE1 gene were also detected at the 55-kDa and 48-kDa positions, despite the fact that the gene has only one protein of 102 AAs (10.9 kDa) listed in the NCBI database (Table II).

In a separate experiment, proteins from 8-12 HSP or associated genes were detected in MB231 and HEK293 cells at the 40-kDa and/or 26-kDa positions. However, according to the NCBI database, these proteins varied between 98-53.5 kDa (Table II). The majority of these genes were also detected in MB231 and MCF7 cells at the 72-kDa, 55-kDa and/or 48-kDa positions in the aforementioned study (Table II). By contrast, HSPE1, whose protein should be 10.9 kDa, was also detected at the 26-kDa in HEK293 cells.

Discussion

Although cells may behave differently *in vitro* compared with *in vivo*, it was worth noting that at 39˚C, the Wt forms of the majority of the genes studied did not show a pronounced increase at the RNA level. Furthermore, even the Wt mRNA of the HSF1 gene, whose protein is the master transactivator of the majority of HSP genes, did not demonstrate a universal rise. Associated with these findings were complex changes in the ratios of expression levels between different RNA variants of HSF1 and several HSP genes. These results collectively suggested that regulation of alternative transcriptional initiation to generate different RNA transcripts and regulation of alternative splicing to convert the same transcripts into different mature RNA variants are two important mechanisms by which cells adapt to increased ambient temperatures. These mechanisms collectively lead to complex changes in the ratios of expression levels between different RNA variants of the same gene, including certain new ones that may not be expressed under physiological conditions. Given that HSPs have versatile functions, these mechanisms may confer onto cells tactics to promptly adapt to a sudden increase in ambient temperature and other microenvironmental changes under hostile conditions. These RNA changes, as well as other molecular mechanisms that are not discussed herein, render cells more flexible in dealing with stressful situations.

Over recent decades, NCBI has updated its database by deleting the part of an RNA sequence that houses our RT-PCR

```
NM_001540.3:
1.  MGRGGPQTGHCH.
2.  MTERRVPFSLLRGPSWDPFRDWYPHSRLFDQAFGLPRLPEEWSQWLGGSSWPGYVRPLPPAAIESPAVAAPAYSRALSRQLSSGVSEIRHTAD
    RWRVSLDVNHFAPDELTVKTKDGVVEITGKHEERQDEHGYISRCFTRKYTLPPGVDPTQVSSSLSPEGTLTVEAPMPKLATQSNEITIPVTFE
    SRAQLGGPEAAKSDETAAK.(full-length HSP27 protein)
3.  MSTTSPRTS.
4.  MAWWRSPASTRSGRTSMATSPGASRGNTRCPPVWTPPKFPPPCPLRAH.
5.  MRLPPSKALARMPTPAAATGCASPATCVFF.

NM_00154.5:
1.  MTERRVPFSLLRGPSWDPFRDWYPHSRLFDQAFGLPRLPEEWSQWLGGSSWPGYVRPLPPAAIESPAVAAPAYSRALSRQLSSGVSEIRHTAD
    RWRVSLDVNHFAPDELTVKTKDGVVEITGKHEERQDEHGYISRCFTRKYTLPPGVDPTQVSSSLSPEGTLTVEAPMPKLATQSNEITIPVTFE
    SRAQLGGPEAAKSDETAAK.(full-length HSP27 protein)
2.  MSTTSPRTS.
3.  MAWWRSPASTRSGRTSMATSPGASRGNTRCPPVWTPPKFPPPCPLRAH.
4.  MRLPPSKALARMPTPAAATGCASPATCVFF.

716A1-20 and cloned sequence-1:
1.  MGRGGPQTGHCH.
2.  MATSPGASRGNTRCPPVWTPPKFPPPCPLRAH.
3.  MPKLATQSNEITIPVTFESRAQLGGPEAAKSDETAAK. (the last 37 AAs of HSP27)
4.  MRLPPSKALARMPTPAAATGCASPATCVFF.

716A1-15:
1.  MGRGGPQTGHCH.
2.  MTERRVPFSLLRGPSWDPFRDWYPHSRLFDQAFGLPRLPEEWSQWLGGSSWPEAAKSDETAAK. (The first 51 AAs belong to
    the N-terminus of HSP27)
3.  MRLPPSKALARMPTPAAATGCASPATCVFF.

716A1-16:
1.  MGRGGPQTGHCH.
2.  MTERRVPFSLLRGPSWDPFRDWYPHAQASHAVQRDHHPSHLRVAGPAWGPRSCKIR.(The first 25 AAs belong to the N-
    terminus of HSP27)
3.  MRLPPSKALARMPTPAAATGCASPATCVFF.
```

Figure 6. Translation of ORFs in NM_001540.3, NM_001540.5, and the four new RNA variants cloned in the present study. The shaded sequences are identical to the N- or C-terminal part of the Wt HSP27 protein. ORFs, open reading frames; Wt, wild-type.

primers, despite the existence of the missing part being confirmed by nested RT-PCR and/or sequencing of the RT-PCR amplicons, including in the cases of HSP90AA1 and HSPB1 shown in the present study. A suggestion is made to our RNA research peers that numerous older versions of RNA sequences may be expressed in certain situations and thus should be regarded as RNA variants. Since these earlier versions may still be retrieved from NCBI in the majority of cases, it is strongly recommended that these earlier versions should be retrieved and consulted when studying RNA sequences.

In the realm of RNA research, certain studies consider arbitrarily those RNAs with their largest ORF shorter than 100 codons or 300 nts as long non-coding RNAs (25-29), while other studies define long non-coding RNAs more stringently with 200 nts as the arbitrary cut-off (30-32). The present study (8,10,33), along with many others (34,35), argue against these existing definitions as numerous studies have demonstrated that very short peptides [as short as 11 AAs encoded by only 33 nts (36-39)] have important functions. Therefore, numerous RNAs, including virtually all mRNAs, may function through translation into proteins or short peptides and through one of the several known non-coding mechanisms (including miRNA or siRNA mechanisms), with the steroid receptor RNA activator being the first known example of such dual mechanisms (40). The number of pure long non-coding RNAs, defined as those that function only via a non-coding mechanism, is much smaller than has been claimed in the literature. In fact, few reports to date have carefully studied long non-coding RNAs for short peptide expression. The dichotomy of RNAs into long non-coding ones and coding ones is too arbitrary (41) and may be greatly misleading. These general conceptions lead us to consider that the four HSPB1 RNA variants identified herein should be classified as mRNAs if the deduced peptides are expressed, although the longest ORF they encompass is shorter than 200 nts.

The possible translation of an HSP27-containing peptide from the 716A1-16 RNA raises concerns about the definition of 'protein isoforms' used in the present. This is because only 25 of the 56 AAs encoded by the 716A1-16 belong to HSP27, and this short peptide (if it is expressed) may have quite a different function compared with Wt HSP27, which encompasses 205 AAs. By contrast, it is not uncommon for two protein isoforms encoded by two differently spliced mRNAs of the same gene to have opposite functions. For example, the Bcl-xL (the long form of Bcl-X) promotes cell survival, while the Bcl-xS (the short form lacking one exon) promotes cell death; however, the long and short forms of the FANCL gene exhibit the opposite phenomenon (42). Therefore, this deduced HSP27-containing peptide may be considered by others as a non-HSP27 protein, despite it being considered an HSP27 isoform in the present study. This problem requires further discussion in the RNA research fraternity to reach a common definition of 'protein isoforms', since the human genome may encode an astounding number of such similar/dissimilar ORFs.

Reanalysis of previous proteomics data has revealed that numerous HSP or associated genes have their protein products detected simultaneously at multiple positions on SDS-PAGE, including some positions with much larger or smaller

Table II. Heat-shock and associated proteins detected at different positions of SDS-PAGE.

Position[a]	Cell line	Proteins detected (isoform variation)[b]
72-kD (65-80 kD)	MB231	HSPD1 (573); *HSP90AA1* (*854*,732,853); STIP1 (590,543,519); HSPA9 (679); *HSPA8* (*493*,646); HSP90B1 (803); HSPA5(654); HSP90AB1 (724,714,676); HSPA1A (641); TRAP1 (651,704,564); HSPA4 (840); HSPA1L (641)
	MCF7	HSPD1; TRAP1; HSPA2 (639); HSP90AA1; HSP90AB1; STIP1; HSPA9; HSPA5; *HSPA8*; HSPA1A; HSP90B1; HSPA6 (643); *HSPH1* (851,858,782,*860*,814,809,853,816,736,780,807)
55-kD (50-60 kD)	MB231	HSPD1; *HSP90AA1*; HSP90AB1; *HSPE1* (*102*); HSP90B1; HSPA8; HSPA14 (509); HSPA5; CDC37 (429,378); HSPA9; HSPA1A; AHSA1 (203,338)
	MCF7	HSPD1; STIP1; HSPA5; *HSPA8*; HSP90AB1; *HSPH1*; TRAP1; HSPA1A; HSPA1L; HSP90AA1; HSPA9
48-kD (43-53 kD)	MB231	HSPD1; *HSP90AA1*; HSP90AB1; *HSPE1*; HSPA5; HSPA8; CDC37; HSP90B1; HSPA1A or HSPA1L; HSPA9
	MCF7	HSPD1; HSP90AA1; HSP90AB1; HSPA1A; *HSPA8*; HSPA9; HSPA14(509); *HSPH1*; HSPA1L
40-kD	MB231	*HSPA8*; HSPD1; HSPA9; HSPA5; *HSPBP1* (*359*,405,431,327)
(36-44 kD)	HEK293	HSPD1; AHSA1; HSP90B1; TRAP1; HSP90AB1; HSPA9; HSPA5; HSP90AA1; *HSPA8*; HSPBP1; STIP1; HSPA1A; HSPA4; *HSPH1*; CDC37
26-kD (22-30 kD)	HEK293	HSPD1; HSP90AB1; HSPB1 (205); HSP90B1; *HSP90AA1*; TRAP1; HSPA8; HSPA5; HSPA1A; HSPA1L; HSPA9; *HSPE1*; AHSA1

[a]Approximately 10% divergence of the molecular weight is parenthesized. [b]Molecular weights (in kD) of protein isoforms of the gene are given in parenthesis. The gene and the largest or smallest isoform it produces are italicized and boldfaced.

molecular masses than their theoretical counterparts (Table II). Explanations for those appearing at a relatively low position on the SDS-PAGE include that either they are isoform(s) smaller than the Wt or they are degraded debris. Explanations for those appearing at a much larger position include that either they are unknown isoforms or they are known isoforms, but have undergone multiple post-translational modifications, including the recently reported lysine lactation (43). At present, there is still lacking a reliable high-throughput technique to rectify the authenticity of those proteins that appear at the 'wrong' positions on SDS-PAGE and to rectify the three putative HSP27 isoforms described in the present study. These technical hurdles include the lack of isoform-specific antibodies to determine protein identity using Western blotting, and the fact that different isoforms sharing parts of the AA sequence may differ in conformation and thus have different antibody epitopes.

In conclusion, the RT-PCR results demonstrated that HSF1 and several HSP genes respond to increased ambient temperatures in cell culture through complex changes in the ratios of expression levels between different RNA variants, including certain previously unreported variants and older versions of the RNAs that have been replaced by newer versions. Among these variants are four novel RNA variants of the HSPB1 gene that may encode three short HSP27 protein isoforms. In line with these observations, reanalysis of proteomics data from two previous studies also reveals that proteins of certain HSP or associated genes may be detected simultaneously at multiple positions on SDS-PAGE, suggesting that these genes are likely to be expressed in multiple isoforms. Among the limitations of the present study is the lack of determination of the time course of the alterations in the expression of individual HSP genes at the full range of febrile temperatures, i.e. from 38˚C or 42˚C.

Using alternative transcription-initiation sites and alternative RNA splicing to produce different HSF1 RNAs and HSP genes may be important molecular mechanisms that allow cells to respond to an elevated ambient temperature *in vitro*.

Availability of data and materials

The sequences cloned in the present study have been documented in public databases: MW881014 for 7161A-15, MW881015 for 7161A-16, MW881017 for 7161A-20, and MW881017 for cloned sequence 1. The sequences can be retrieved by entering the corresponding access number into the NCBI website for nucleotides (www.ncbi.nlm.nih.gov/nuccore/?term=)

Authors' contributions

XG performed part of the study and drafted the manuscript. KYZ and HYZ performed part of sequence analysis and

bioinformatics analysis while preparing the figures and tables. LZ reanalyzed and corrected all open reading frames of the RNA sequences as well as helped draw conclusions while performing English editing of the manuscript. CFY performed some bench experiments and helped conceptualize the manuscript. HH participated in conceptualization and discussion of the study and helped draw conclusions. DZL conceptualized and finalized the manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Chaudhury S, Keegan BM and Blagg BSJ: The role and therapeutic potential of Hsp90, Hsp70, and smaller heat shock proteins in peripheral and central neuropathies. Med Res Rev 41: 202-222, 2021.
2. Dabbaghizadeh A and Tanguay RM: Structural and functional properties of proteins interacting with small heat shock proteins. Cell Stress Chaperones 25: 629-637, 2020.
3. Bascos NAD and Landry SJ: A history of molecular chaperone structures in the protein data bank. Int J Mol Sci 20: 6195, 2019.
4. Masser AE, Ciccarelli M and Andreasson C: Hsf1 on a leash-controlling the heat shock response by chaperone titration. Exp Cell Res 396: 112246, 2020.
5. Ahmed K, Zaidi SF, Mati-Ur-Rehman, Rehman R and Kondo T: Hyperthermia and protein homeostasis: Cytoprotection and cell death. J Therm Biol 91: 102615, 2020.
6. Inia JA and O'Brien ER: Role of heat shock protein 27 in modulating atherosclerotic inflammation. J Cardiovasc Transl Res 14: 3-12, 2021.
7. Prince TL, Lang BJ, Guerrero-Gimenez ME, Fernandez-Munoz JM, Ackerman A and Calderwood SK: HSF1: Primary factor in molecular chaperone expression and a major contributor to cancer morbidity. Cells 9: 1046, 2020.
8. He Y, Yuan C, Chen L, Lei M, Zellmer L, Huang H and Liao DJ: Transcriptional-Readthrough RNAs reflect the phenomenon of 'A Gene Contains Gene(s)' or 'Gene(s) within a Gene' in the human genome, and thus are not Chimeric RNAs. Genes (Basel) 9: 40, 2018.
9. He Y, Yuan C, Chen L, Liu Y, Zhou H, Xu N and Liao DJ: While it is not deliberate, much of today's biomedical research contains logical and technical flaws, showing a need for corrective action. Int J Med Sci 15: 309-322, 2018.
10. Jia Y, Chen L, Ma Y, Zhang J, Xu N and Liao DJ: To know how a gene works, we need to redefine it first but then, more importantly, to let the cell itself decide how to transcribe and process its RNAs. Int J Biol Sci 11: 1413-1423, 2015.
11. Liu X, Wang Y, Yang W, Guan Z, Yu W and Liao DJ: Protein multiplicity can lead to misconduct in western blotting and misinterpretation of immunohistochemical staining results, creating much conflicting data. Prog Histochem Cytochem 51: 51-58, 2016.
12. Qu J, Zhang J, Zellmer L, He Y, Liu S, Wang C, Yuan C, Xu N, Huang H and Liao DJ: About three-fourths of mouse proteins unexpectedly appear at a low position of SDS-PAGE, often as additional isoforms, questioning whether all protein isoforms have been eliminated in gene-knockout cells or organisms. Protein Sci 29: 978-990, 2020.
13. Yan R, Zhang J, Zellmer L, Chen L, Wu D, Liu S, Xu N and Liao JD: Probably less than one-tenth of the genes produce only the wild type protein without at least one additional protein isoform in some human cancer cell lines. Oncotarget 8: 82714-82727, 2017.
14. Zhang J, Lou X, Shen H, Zellmer L, Sun Y, Liu S, Xu N and Liao DJ: Isoforms of wild type proteins often appear as low molecular weight bands on SDS-PAGE. Biotechnol J 9: 1044-1054, 2014.
15. Taga A and Cornblath DR: A novel HSPB1 mutation associated with a late onset CMT2 phenotype: Case presentation and systematic review of the literature. J Peripher Nerv Syst 25: 223-229, 2020.
16. Gao X, Zhang K, Lu T, Zhao Y, Zhou H, Yu Y, Zellmer L, He Y, Huang H and Joshua Liao D: A reassessment of several erstwhile methods for isolating DNA fragments from agarose gels. 3 Biotech 11: 138, 2021.
17. Sun Y, Sriramajayam K, Luo D and Liao DJ: A quick, cost-free method of purification of DNA fragments from agarose gel. J Cancer 3: 93-95, 2012.
18. Sun Y, Cao S, Yang M, Wu S, Wang Z, Lin X, Song X and Liao DJ: Basic anatomy and tumor biology of the RPS6KA6 gene that encodes the p90 ribosomal S6 kinase-4. Oncogene 32: 1794-1810, 2013.
19. Sun Y, Lou X, Yang M, Yuan C, Ma L, Xie BK, Wu JM, Yang W, Shen SX, Xu N and Liao DJ: Cyclin-dependent kinase 4 may be expressed as multiple proteins and have functions that are independent of binding to CCND and RB and occur at the S and G 2/M phases of the cell cycle. Cell Cycle 12: 3512-3525, 2013.
20. Yang M, Sun Y, Ma L, Wang C, Wu JM, Bi A and Liao DJ: Complex alternative splicing of the smarca2 gene suggests the importance of smarca2-B variants. J Cancer 2: 386-400, 2011.
21. Yang M, Wu J, Wu SH, Bi AD and Liao DJ: Splicing of mouse p53 pre-mRNA does not always follow the 'first come, first served' principle and may be influenced by cisplatin treatment and serum starvation. Mol Biol Rep 39: 9247-9256, 2012.
22. Yang W, Chen L, Yan F, Xie BK, Liao Y and Liao DJ: Dirty plasmid used as a quick and low-cost method to identify bacterial colonies with a recombinant plasmid. Indian J Applied Res 6: 649, 2016.
23. Wu J, Wu SH, Bollig A, Thakur A and Liao DJ: Identification of the cyclin D1b mRNA variant in mouse. Mol Biol Rep 36: 953-957, 2009.
24. Bollig-Fischer A, Thakur A, Sun Y, Wu J and Liao DJ: The predominant proteins that react to the MC-20 estrogen receptor alpha antibody differ in molecular weight between the mammary gland and uterus in the mouse and rat. Int J Biomed Sci 8: 51-63, 2012.
25. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC and Giraldez AJ: Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J 33: 981-993, 2014.
26. Cheng H, Chan WS, Li Z, Wang D, Liu S and Zhou Y: Small open reading frames: Current prediction techniques and future prospect. Curr Protein Pept Sci 12: 503-507, 2011.
27. Kageyama Y, Kondo T and Hashimoto Y: Coding vs non-coding: Translatability of short ORFs found in putative non-coding transcripts. Biochimie 93: 1981-1986, 2011.
28. Landry CR, Zhong X, Nielly-Thibault L and Roucou X: Found in translation: Functions and evolution of a recently discovered alternative proteome. Curr Opin Struct Biol 32: 74-80, 2015.
29. Pauli A, Valen E and Schier AF: Identifying (non-)coding RNAs and small peptides: Challenges and opportunities. Bioessays 37: 103-112, 2015.
30. Signal B, Gloss BS and Dinger ME: Computational approaches for functional prediction and characterisation of long noncoding RNAs. Trends Genet 32: 620-637, 2016.
31. Jia H, Osak M, Bogu GK, Stanton LW, Johnson R and Lipovich L: Genome-wide computational identification and manual annotation of human long noncoding RNA genes. RNA 16: 1478-1487, 2010.
32. Samudyata, Castelo-Branco G and Bonetti A: Birth, coming of age and death: The intriguing life of long noncoding RNAs. Semin Cell Dev Biol 79: 143-152, 2018.
33. Yuan C, Han Y, Zellmer L, Yang W, Guan Z, Yu W, Huang H and Liao DJ: It is imperative to establish a pellucid definition of chimeric RNA and to clear up a lot of confusion in the relevant research. Int J Mol Sci 18: 714, 2017.
34. Reisacher C and Arbibe L: Not lost in host translation: The new roles of long noncoding RNAs in infectious diseases. Cell Microbiol 21: e13119, 2019.
35. Ruiz-Orera J, Messeguer X, Subirana JA and Alba MM: Long non-coding RNAs as a source of new peptides. Elife 3: e03523, 2014.

36. Chu Q, Ma J and Saghatelian A: Identification and characterization of sORF-encoded polypeptides. Crit Rev Biochem Mol Biol 50: 134-141, 2015.
37. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S and Kageyama Y: Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. Nat Cell Biol 9: 660-665, 2007.
38. Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, Kobayashi S, Payre F and Kageyama Y: Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. Science 329: 336-339, 2010.
39. Ladoukakis E, Pereira V, Magny EG, Eyre-Walker A and Couso JP: Hundreds of putatively functional small open reading frames in Drosophila. Genome Biol 12: R118, 2011.
40. Chooniedass-Kothari S, Emberley E, Hamedani MK, Troup S, Wang X, Czosnek A, Hube F, Mutawe M, Watson PH and Leygue E: The steroid receptor RNA activator is the first functional RNA encoding a protein. FEBS Lett 566: 43-47, 2004.
41. Bogard B, Francastel C and Hube F: Multiple information carried by RNAs: Total eclipse or a light at the end of the tunnel? RNA Biol 17: 1707-1720, 2020.
42. Yuan C, Xu N and Liao J: Switch of FANCL, a key FA-BRCA component, between tumor suppressor and promoter by alternative splicing. Cell Cycle 11: 3356, 2012.
43. Zhang D, Tang Z, Huang H, Zhou G, Cui C, Weng Y, Liu W, Kim S, Lee S, Perez-Neut M, et al: Metabolic regulation of gene expression by histone lactylation. Nature 574: 575-580, 2019.