MOLECULAR AND CELLULAR BIOLOGY, Mar. 1988, p. 1113–1122
0270-7306/88/031113-10$02.00/0

Vol. 8, No. 3

Characterization of *Gmhsp26-A*, a Stress Gene Encoding a Divergent Heat Shock Protein of Soybean: Heavy-Metal-Induced Inhibition of Intron Processing†

EVA CZARNECKA,[1] RONALD T. NAGAO,[2] JOE L. KEY,[2] AND WILLIAM B. GURLEY[1]*

Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611,[1] and Botany Department, University of Georgia, Athens, Georgia 30602[2]

Received 4 August 1987/Accepted 10 December 1987

We determined the DNA sequence and mapped the corresponding transcripts of a genomic clone containing the *Gmhsp26-A* gene of soybean. This gene is homologous to the previously characterized cDNA clone pCE54 (E. Czarnecka, L. Edelman, F. Schöffl, and J. L. Key, Plant Mol. Biol. 3:45–58, 1984) and is expressed in response to a wide variety of physiological stresses including heat shock (HS). S1 nuclease mapping of transcripts and a comparison of the cDNA sequence with the genomic sequence indicated the presence of a single intron of 388 base pairs. Intron removal from pre-mRNA was preferentially inhibited by treatment of soybean seedlings with either $CdCl_2$ or $CuSO_4$. Analysis of the 5′ termini of transcripts indicated the presence of one major and at least two minor start sites. In each case, initiation occurred 27 to 30 base pairs downstream from a TATA-like motif, and thus each initiation site appears to be promoted by the activity of a separate subpromoter. The three subpromoters are all associated with sequences showing low homology to the HS consensus element of *Drosophila melanogaster* HS genes and are differentially induced in response to various stresses. Within the carboxyl-terminal half of the protein, hydropathy analysis of the deduced amino acid sequence indicated a high degree of relatedness to the small HS proteins. A comparison of the primary amino acid sequence of hsp26-A with sequences of the small HS proteins suggested that this stress protein is highly diverged and may therefore be specialized for stress adaptation in soybean.

The heat shock (HS) response is found in a wide range of organisms and is typically triggered when cells are exposed to hyperthermia or a variety of other stresses (1). In soybean the low-molecular-mass small heat shock proteins (hsps) represent the most abundant class of proteins expressed during HS (23) and include 30 to 50 proteins which range in molecular mass from 15 to 27 kilodaltons (kDa). The low-molecular-mass family of hsps is more diverse than the high-molecular-mass hsps (68, 70, and 83 kDa) and can be subdivided into at least three subgroups on the basis of size and conservation of amino acid sequences (22, 30, 40, 41).

Analysis of hybridization of soybean seedling RNA to cloned HS-specific cDNAs indicates that most representatives of the low-molecular-mass hsps show various degrees of induction by numerous stress agents, but generally at levels much lower than with HS (8; L. Edelman, E. Czarnecka, and J. L. Key, submitted for publication). Although transcripts homologous to clone pCE54 show the greatest increase in abundance in response to elevated temperatures, their pattern of induction differs from the response of typical HS RNAs, since they are present at low to moderate levels in unstressed tissue and markedly increase in response to a wide array of stresses, including treatment with abscisic acid, 2,4-dichlorophenoxyacetic acid, polyethylene glycol, polyamines, sodium arsenite, and heavy metals (8; Edelman et al., submitted). RNA homologous to pCE54 translates into several 27-kDa peptides in vitro, indicating that the gene corresponding to the cDNA represents one gene of a family of four or five closely related genes (8).

We sequenced and transcript mapped a soybean gene (designated *Gmhsp26-A*) homologous to cDNA clone pCE54. We show here that this general stress gene contains a single intron and is related to the small hsp family. In this study, only the treatment of seedlings with $CdCl_2$ or $CuSO_4$ caused the appearance of a second homologous transcript of higher molecular mass that could be detected by Northern blot (RNA blot) analysis (8; Edelman et al., submitted). We show here that the high-molecular-mass transcript represents an unprocessed *Gmhsp26-A* mRNA resulting from the partial processing inhibition of the primary transcript either by cadmium or copper treatment of seedlings.

MATERIALS AND METHODS

Plant material and incubation conditions. Soybean seedlings were grown in the dark for 2 days in moist paper rolls at 28°C (8). All induction treatments of harvested plant material were performed in incubation solution consisting of 1 mM potassium phosphate buffer (pH 6.0) and 1% sucrose with the addition of $CdCl_2$ or $CuSO_4$ where indicated. Treatment conditions of seedlings and isolation of hypocotyl or seedling poly(A)$^+$ RNA were as described previously (8, 16).

cDNA and genomic clones. Plasmid pCE54 (8) was selected in a previous study from a cDNA library (40) prepared from total soybean (*Glycine max* var. Wayne) poly(A)$^+$ RNA isolated from heat-shocked (40°C, 2 h) hypocotyls. Transcripts homologous to this cDNA clone were present in small amounts in unstressed hypocotyls (28°C, 2 h) and were shown to increase in abundance in response to HS (40°C, 2 h) and a variety of other stress treatments. The cDNA clone pCE54 was used as a hybridization probe to screen a lambda 1059 library prepared from total soybean DNA (30). One of

* Corresponding author.
† Florida Agricultural Experiment Station journal series no. 8671.

the selected genomic clones was designated hsY54 and yielded three major fragments ranging in size from ca. 9 to 4 kilobases (kb) when digested with *Bgl*II. Two fragments of 4.0 and 6.0 kb hybridized to the pCE54 insert. The 6.0-kb fragment was subcloned into pUC vectors (46) and subsequently was shown by DNA sequence analysis to contain the complete stress gene *Gmhsp26-A*. The 4.0-kb fragment presumably contained a related member of the *hsp26* gene family and was not investigated further.

DNA sequence determination and transcript mapping. Determination of the DNA sequence of *Gmhsp26-A* was conducted by the chemical cleavage method of Maxam and Gilbert (27) and by the dideoxynucleotide chain-termination technique of Sanger et al. (39). Overlapping deletions generated for sequencing were constructed by the procedure of Dale et al. (10). Sequences from −420 to +1,573 base pairs (bp) were determined for both strands, although only the sense strand was sequenced for the far upstream region (−1,190 to −420 bp).

Transcripts homologous to *Gmhsp26-A* were mapped either by S1 nuclease hybrid protection with 1 to 3 μg of poly(A)$^+$ RNA or by primer extension with 5 to 10 μg of poly(A)$^+$ RNA. S1 nuclease mapping of transcripts was performed by the method of Favaloro et al. (12) with modifications as described previously (9). Primer extension analysis was performed with a synthetic oligonucleotide (5′-GGAACAAGTATAGCTGG-3′) primer complementary to sequences downstream of the 5′ termini at positions +24 to +40. Control, heat-shocked, or cadmium-induced poly(A)$^+$ RNA was annealed after denaturation to 5′-end-labeled primer at room temperature for 16 h in sealed microcapillary tubes. The extension reaction mixtures were incubated with 200 U of Maloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Inc.) for 1 h at room temperature. The in vitro-synthesized products were separated on 8% polyacrylamide-urea gels and exposed with screens at −70°C.

RESULTS

DNA sequence analysis. Figure 1 depicts the restriction map of *Gmhsp26-A*, which resides on the 6.0-kb *Bgl*II fragment of genomic clone hsY54. The nucleotide sequence and derived amino acid sequence of *Gmhsp26-A* are shown in Fig. 2. This sequence spanned 1,190 bp of the 5′ flanking region, a transcribed region of 1,335 bp, and 236 bp downstream from the 3′ termini. A single intron of 388 bp splits the protein-coding region between codons +107 and +108. The donor and acceptor splice junction sequences were in agreement with the typical eucaryotic consensus (31). The open reading frame in the mRNA extended for 225 codons and is predicted to code for a 26.0-kDa protein.

The existence of an intron was implied by a comparison of the cDNA (pCE54) sequence with the genomic sequence and was supported by S1 nuclease mapping of the transcript. The incomplete cDNA sequence matched that of the genomic fragment exactly between positions +203 and +393 (first exon) and also between +782 and +1,126 (second exon). The cDNA clone pCE54, which was missing sequences +1 to +202, included most of the protein-coding region and stopped 9 nucleotides short of the translational stop codon.

Inhibition of intron removal by $CdCl_2$ or $CuSO_4$. Transcript mapping by S1 nuclease was performed with hybridization probes that were end labeled upstream, downstream, and internally relative to the intron (Fig. 3D). The results of mapping with a *Hin*dIII-*Bam*HI probe (0.86 kb) 3′ end labeled within the first exon at the upstream *Hin*dIII site are shown in Fig. 3A. A slight expression of *Gmhsp26-A* is evident at 28°C, with a high level of expression seen with HS (40°C) or $CdCl_2$ (200 μM) treatment. The prominent band (296 bp) corresponds in length to the distance from the labeled terminus to the boundary between the first exon and the 5′ border of the intron. RNA from cadmium-treated seedlings produced two protected fragments corresponding to the splice junction and the full-length probe, suggesting that a portion of the transcript retained the intervening sequences. The large fragment corresponding to unspliced transcripts was not detected in control RNA but was seen in heat-shocked RNA after prolonged exposure of the gel.

Transcript mapping (Fig. 3B) with a probe 5′ end labeled in the second exon at the *Bam*HI site also indicated that removal of the intervening sequences was inhibited by $CdCl_2$ or $CuSO_4$. The protected 180-bp fragment corresponded in length to the distance from the labeled terminus to the 3′ border of the intron, and the 961-bp fragment represented unprocessed mRNA (Fig. 3B). The bands visible immediately above the 180-bp band presumably arose from the formation of a DNA hairpin (5′-GTGATTTGCGAAATCAG-3′) structure located within the intron at the acceptor splice junction.

The effect of various $CdCl_2$ concentrations on the inhibition of intron processing is shown in Fig. 3C. The mapping probe (Fig. 3D) was 5′ end labeled at the *Hin*cII site within the intron so that only transcripts containing the unprocessed intron would yield end-labeled protected fragments. Increasing cadmium concentrations resulted in the increased accumulation of unspliced transcripts. Maximum inhibition of intron removal was observed after treatment with 5 mM $CdCl_2$ for 2 h. Note that HS at 40°C resulted in the increased

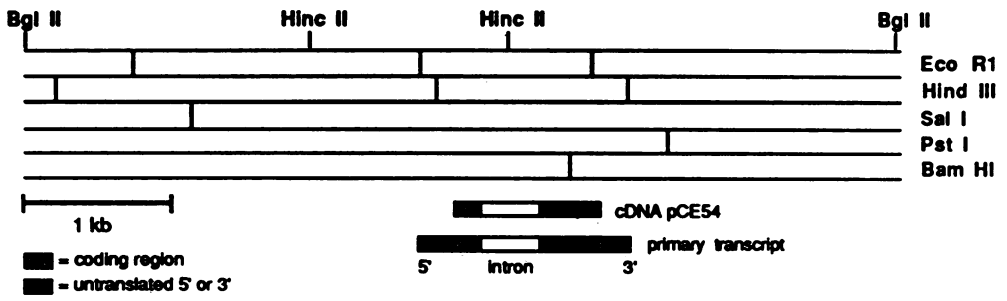


FIG. 1. Restriction map of the cloned *Bgl*II fragment containing stress gene *Gmhsp26-A*. Only selected *Hin*cII sites used for preparation of end-labeled probes are shown. The partial cDNA clone pCE54 extends from +1126 to +203 with a perfect match to the genomic sequence on either side of the intron, which is located from +394 to +781.

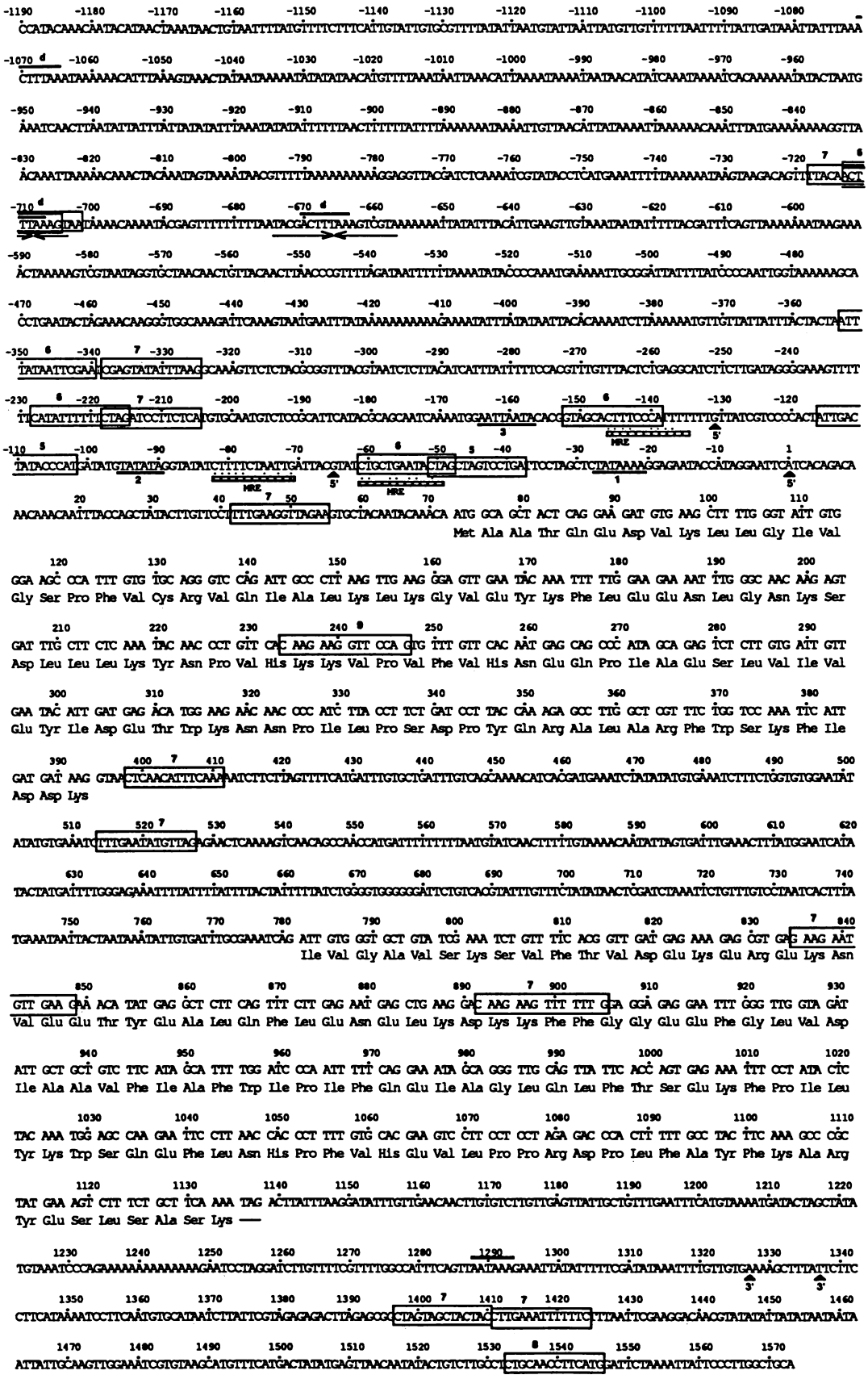


FIG. 2. Nucleotide sequence and derived amino acid sequence of *Gmhsp26-A*. Nucleotides are numbered from the major start site of transcription. Triangles denote 5′ and 3′ termini as determined by S1 nuclease hybrid protection studies and reverse transcriptase primer extension assays. TATA-like motifs are underlined. Regions with homology to the *Drosophila* HSE are blocked with the number of matches out of 10 indicated in boldface type above the box. The striped bars indicate homology (dots) to the mammalian MRE. Arrows, Positions of inverted repeats; d, 6-bp direct repeat. The consensus poly(A)$^+$ signal is centered at +1290.

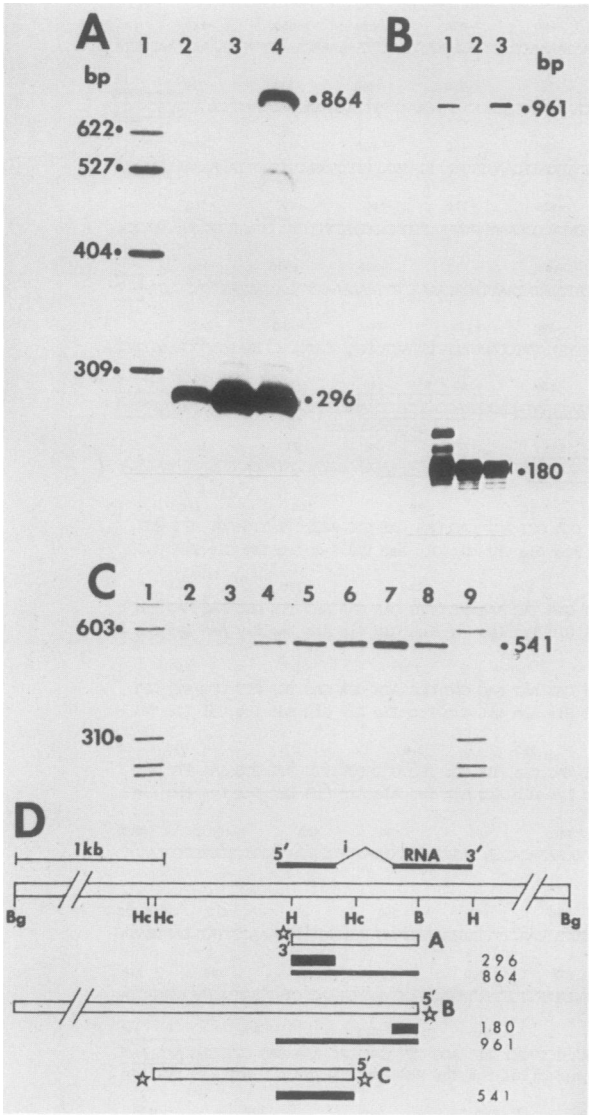


FIG. 3. S1 nuclease transcript mapping of gene *Gmhsp26-A*. Poly(A)$^+$ RNA (3 μg) and 50 U of S1 nuclease per ml were used in hybrid protection assays. (A) 5′ intron border mapped with a 3′-end-labeled DNA probe (*Hin*dIII to *Bam*HI). Markers (lane 1) are *Hpa*II restriction fragments of pBR322. Lanes representing RNA from soybean seedlings exposed to various stresses for 2 h: 2, 28°C; 3, 40°C; and 4, 200 μM $CdCl_2$ at 28°C. (B) 3′ intron border mapped with a 5′-end-labeled *Bgl*II-*Bam*HI DNA fragment as probe. Lanes representing seedling RNA from different 2-h treatments: 1, 250 μM $CuSO_4$ (28°C); 2, 40°C; 3, 200 μM $CdCl_2$ (28°C). (C) Inhibition of *Gmhsp26-A* intron processing by cadmium treatment. The 5′-end-labeled *Hin*cII fragment used as a probe spanned the 5′-flanking region, the first exon, and 148 bp of the intron. Lanes representing RNA from various 2-h treatments: 2, 28°C; 3, 40°C; 4, 200 μM $CdCl_2$; 5, 500 μM $CdCl_2$; 6, 1 mM $CdCl_2$; 7, 5 mM $CdCl_2$; 8, 10 mM $CdCl_2$. All $CdCl_2$ treatments were carried out at 28°C. Protected bands are due to probe hybridization to unprocessed transcripts only. Markers (lane 1) are *Hae*III restriction fragments of φX174 RF DNA. (D) Restriction map of the cloned *Bgl*II fragment indicating the position of *Gmhsp26-A* transcripts. Open bars below delineate end-labeled probes used in S1 nuclease mapping; black bars represent protected fragments.

accumulation of processed transcripts as seen in the increased levels of the 296-base band in Fig. 3A but in only minute amounts of unspliced transcript, indicating that exposure to elevated temperatures had little detrimental effect on the splicing of this RNA.

Mapping of 5′ termini with S1 nuclease. A detailed analysis of the 5′ termini of the transcripts was conducted with S1 nuclease, using the *Hin*dIII-*Sal*I 1.6-kb fragment 5′ end labeled at the *Hin*dIII site (within the first exon) as the hybridization probe. The sequence ladder of the probe fragment served as a size marker, and a 1.5-nucleotide correction was incorporated into the estimations of protected fragment size (43). These results confirm that expression of *Gmhsp26-A* is increased by HS conditions and to a lesser extent by cadmium as previously shown (8; Edelman et al., submitted). The pattern of protected hybrids obtained with S1 nuclease mapping indicated the presence of multiple 5′ termini with HS or cadmium treatments (Fig. 4). The three major initiation sites are spaced at 65-bp intervals and are positioned 27 to 30 bp downstream from A+T-rich regions similar to the TATA box of many class II genes (6).

The sequences required for initiating transcription at each of the three sites were designated subpromoters 1, 2, and 3, with subpromoter 1 (major) being most proximal to the ATG start codon. Each terminus showed slight heterogeneity, possibly due to S1 nuclease digestion characteristics. The band migrating between transcripts derived from subpromoters 1 and 2 was assumed to be the result of cleavage of the protected hybrids with S1 nuclease at an A+T-rich region and therefore probably does not represent a true 5′ terminus. Radioactive bands corresponding to each terminus were cut from the gel and counted. The percentage of total transcription originating at each site and the degree of induction

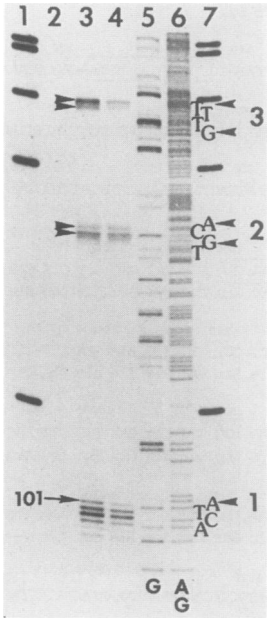


FIG. 4. Nucleotide-level precision in S1 nuclease transcript mapping of 5′ termini of *Gmhsp26-A* RNA. The chemical cleavage-generated sequence ladder of the 5′-end-labeled *Hin*dIII-*Sal*I probe labeled at the *Hin*dIII site was used for size calibration (lanes 5 and 6). The results of S1 nuclease mapping with poly(A)$^+$ RNA isolated from seedlings exposed to various 2-h treatments are presented. Lanes: 2, 28°C; 3, 40°C; 4, 200 μM $CdCl_2$ (28°C). The S1 nuclease concentration was 50 U/ml. Marker lanes (lanes 1 and 7) are *Hae*III restriction fragments of φX174 RF DNA.

TABLE 1. Transcription under different stress conditions[a]

Start site	% Initiation under stress conditions			Induction (fold)	
	28°C	40°C	$CdCl_2$	40°C	$CdCl_2$
3	7	23	18	73	38
2	34	25	32	16	13
1	59	50	50	19	12

[a] The percentage of total transcription under different stress conditions was determined by counting radioactivity from isolated bands cut from the S1 nuclease mapping gel. All treatments were conducted for 2 h. The $CdCl_2$ concentration was 200 μM.

compared with RNA induced at 28°C were calculated and are summarized in Table 1. Transcript 1 predominated under control and stress conditions, representing 50 to 60% of the transcripts. All three transcripts were induced to various degrees by HS and cadmium treatments. Transcript 3 showed the greatest induction with both HS and cadmium.

Mapping the 5′ termini by primer extension. The existence of multiple 5′ termini was further investigated by reverse transcriptase primer extension analysis. The 5′-end-labeled oligomer was hybridized to either control (28°C) or stress-induced soybean poly(A)$^+$ RNA and subsequently extended with Maloney murine leukemia virus reverse transcriptase (Fig. 5). The sizes of the three extended fragments indicate that the multiple bands observed with S1 nuclease analysis represent true 5′ termini. The fraction of the two higher-molecular-mass transcripts is probably grossly underrepresented due to the presence of a potential hairpin structure (−17 to −1) immediately upstream of the major terminus (40-nucleotide band) which resulted in abortive termination of the in vitro extension reaction.

Mapping the 3′ termini with S1 nuclease. The 3′ termini of *Gmhsp26-A* were mapped by S1 nuclease hybrid protection with the *Bam*HI-*Pst*I 0.6-kb fragment 3′ end labeled at the *Bam*HI site as the probe. There were two 3′ termini positioned approximately 370 and 378 bases (±5 bases) downstream of the *Bam*HI site (Fig. 6). Both termini were present in RNA from HS- and cadmium-treated plants. The terminus for the shorter transcript was located within the sequence TGTGAAAA, 33 bases downstream from the add poly(A) signal (5′-AATAAA-3′) established for mammalian genes (36). There was no region of high homology to this consensus region associated with the terminus of the longer transcript. However, sequences showing similarity (AATTA and TATAAA) to the mammalian consensus were located 33 and 18 bases upstream.

Coding-region analysis. The protein encoded by *Gmhsp26-A* is a member of the low-molecular-mass class of hsps. The small hsps of soybean are abundantly expressed at elevated temperatures and consist of 30 to 50 related proteins in the range of 15 to 27 kDa (22, 30, 40). Although a subset of this class with molecular masses from 15 to 18 kDa is greater than 90% homologous in their deduced amino acid sequences, they are less than 20% homologous to the small hsps of other organisms such as *Drosophila melanogaster*, *Xenopus laevis*, and *Caenorhabditis elegans* (3, 9, 19, 30, 38). A linear comparison of amino acids 89 to 140 of soybean hsp26-A with the analogous region (amino acids 73 to 124) of a composite of four soybean hsp17.5 proteins indicated homology of only 14% (Fig. 7). However, a comparison of the same region of hsp26-A with a composite amino acid sequence of the analogous regions in *X. laevis*, *D. melanogaster*, and *C. elegans* small hsps revealed homology of 29%, the same amount seen when the soybean composite hsp17.5 was compared to this same group of hsps from animals.

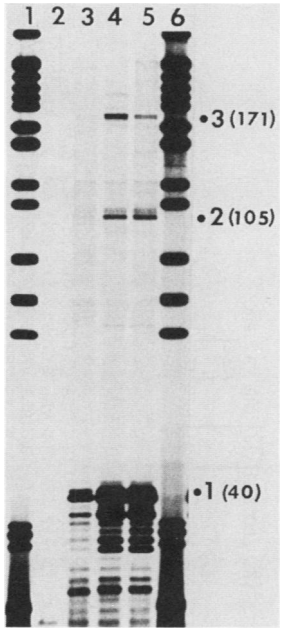


FIG. 5. Reverse transcriptase primer extension mapping of 5′ termini of *Gmhsp26-A* transcripts. Lanes show results obtained with a yeast tRNA control and poly(A)$^+$ RNA (5 μg) isolated from seedlings exposed to various 2-h treatments. Lanes: 2, tRNA; 3, 28°C; 4, 40°C; 5, 200 μM $CdCl_2$ (28°C). End-labeled *Hpa*II restriction fragments of pBR322 were used as size markers in lanes 1 and 6.

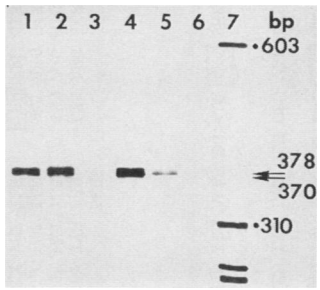
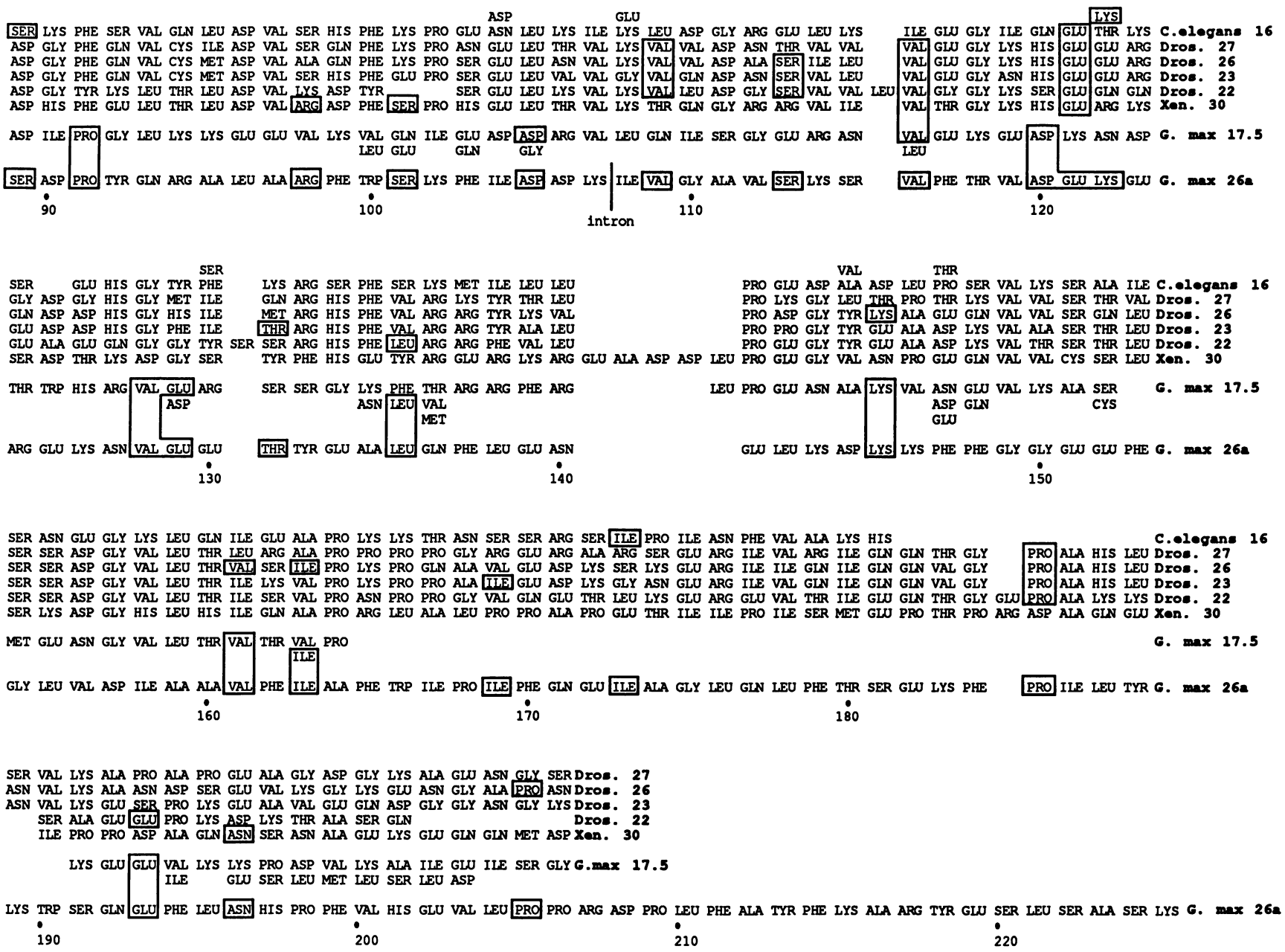


FIG. 6. S1 nuclease transcript mapping of 3′ termini of *Gmhsp26-A* transcripts. The 0.6-kb *Bam*HI-to-*Pst*I fragment was 3′ end labeled at the *Bam*HI site and used as a hybridization probe. Poly(A)$^+$ RNA (1 μg) from soybean seedlings incubated for 2 h at 40°C (lanes 1 to 3) or with 200 μM $CdCl_2$ at 28°C (lanes 4 to 6) was used in S1 nuclease hybrid protection analyses. Hybridization temperatures were 42°C (lanes 1 and 4), 46°C (lanes 2 and 5), and 49°C (lanes 3 and 6). The S1 nuclease concentration was 100 U/ml. Markers (lane 7) are *Hae*III digestion products of φX174 RF DNA.


C.elegans 16
Dros. 27
Dros. 26
Dros. 23
Dros. 22
Xen. 30
G. max 17.5
G. max 26a
intron
C.elegans 16
Dros. 27
Dros. 26
Dros. 23
Dros. 22
Xen. 30
G. max 17.5
G. max 26a
C.elegans 16
Dros. 27
Dros. 26
Dros. 23
Dros. 22
Xen. 30
G. max 17.5
G. max 26a
Dros. 27
Dros. 26
Dros. 23
Dros. 22
Xen. 30
G.max 17.5
G. max 26a

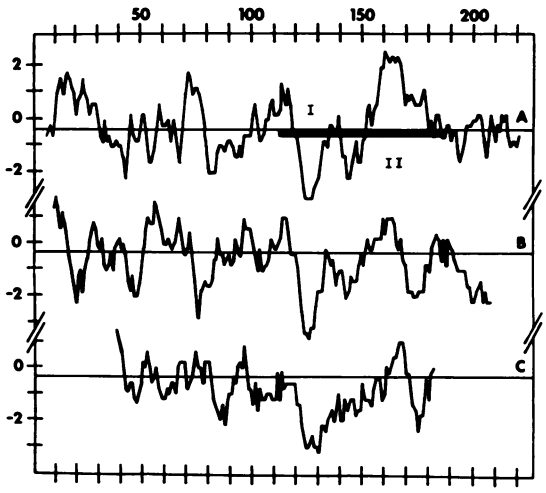


FIG. 8. Hydropathy profiles of soybean hsp26-A (A) and hsp17.5-E (C) and *Drosophila* hsp26 (B). Plots were constructed by the method of Kyte and Doolittle (25). Points above the horizontal line correspond to hydrophobic regions (II), and points below are hydrophilic (I). The plots are aligned along the major hydrophilic peak characteristic of the small hsp class of proteins. The bar in panel A denotes a profile characteristic of low-molecular-mass hsps.

Even though the conservation in amino acid sequences between hsp26-A and members of the small hsp family was low, analysis of hydropathy profiles revealed a very striking structural relationship between these proteins of soybean and other organisms. The relatedness of these proteins was seen most clearly in the presence of a major hydrophilic region (Fig. 8, region I) followed by a characteristic hydrophobic peak (Fig. 8, region II) in the carboxyl half of the proteins. The clear similarity in hydropathy profiles, contrasted with the small amount of amino acid conservation, suggests that hsp26-A is a highly diverged member of the low-molecular-mass family of hsps. It is interesting that the intron is positioned near the boundary between the major hydrophilic peak (amino acids 120 to 150) characteristic of the small hsps (45) and the highly divergent amino-terminal half of the protein, suggesting that this region may be the boundary between structural or modular domains (13–15).

DISCUSSION

Inhibition of intron removal. Previous studies have shown that *Gmhsp26-A* is induced by a variety of stresses, including HS and heavy metals (8; Edelman et al., submitted). Of those agents causing transcriptional induction, cadmium, copper, and arsenite treatments result in accumulation of both normal-size and higher-molecular-mass RNA homologous to *Gmhsp26-A* cDNA (pCE54) (8; Edelman et al., submitted). In the present study, we showed an inhibition of splicing by cadmium or copper, suggesting that the higher-molecular-mass RNA visualized by Northern blot analysis is a precursor mRNA containing the unprocessed intron. Although cadmium and copper inhibited intron removal, processing was barely affected by elevated temperatures. This result differs from the splicing disruption of *Drosophila hsp83* and alcohol dehydrogenase gene transcripts (*hsp70* promoter joined to the alcohol dehydrogenase coding region) at 38°C, in which inhibition of splicing correlates with the accumulation of hsps and exposure to elevated temperatures (47). The block in the processing of *Gmhsp26-A* RNA appeared to be only partial under the conditions tested, with the ratio of unprocessed to processed transcripts unchanged with treatments up to 0.5 mM $CdCl_2$. The inhibition of processing was not due to the general lethality of the 2-h cadmium treatment, since in a previous study both processed and unprocessed *Gmhsp26-A* RNAs continued to accumulate during treatments with up to 1 mM $CdCl_2$ for 6 h or with 0.1 mM $CdCl_2$ for 12 h (Edelman et al., submitted).

The block in processing is consistent with earlier observations from the analysis of in vitro translation products of hybrid-selected RNA isolated from cadmium-treated seedlings (8). In the previous study, clone pCE54 was used to hybrid select poly(A)$^+$ RNA for in vitro translation. The appearance of two or three additional low-molecular-mass peptides (13 to 15 kDa) on two-dimensional gels correlates with cadmium treatment of seedlings. The predicted size of the protein translated from mature *Gmhsp26-A* mRNA is 26 kDa, but translation of the pre-mRNA should only include the first exon and 15 codons of the intron, resulting in a truncated peptide of 14.1 kDa. The predicted size of peptides terminated within the intron coincided well with the size of in vitro-translated peptides, further supporting the identification of the high-molecular-mass transcripts as the precursor to *Gmhsp26-A* mRNA.

The inhibition of splicing by treatment with metals such as cadmium and copper may be widespread in higher plants. Winter et al. (J. A. Winter, R. Wright, N. Duck, C. Gasser, D. Rochester, R. Fraley, and D. Shah, submitted for publication) have evidence indicating that splicing of a petunia *hsp70* gene containing a single intron is inhibited by $CdCl_2$ concentrations as low as 50 μM. In addition, we have examined transgenic expression of *Gmhsp26-A* in sunflower crown gall tumors and found a cadmium inhibition of splicing similar to that in soybean seedlings. For petunia *hsp70*, the ratio of processed to unprocessed transcripts did not change with prolonged cadmium treatment or at elevated levels (500 μM) of cadmium. The inability of cadmium to completely block intron removal in soybean, sunflower, and petunia may mean that in plants multiple pathways of splicing exist which can be differentiated by their sensitivity to heavy metals. It is not clear whether the inhibition of intron splicing by $CdCl_2$ is restricted to specific classes of HS transcripts or whether inhibition is a general phenomenon affecting most intron-containing RNA in plant cells.

Promoter structure. There were three start sites for *Gmhsp26-A* transcription as established by S1 nuclease mapping and reverse transcriptase primer extension analysis. The existence of three initiation sites for transcription suggests that this promoter may actually be a composite of three subpromoters. Each of the initiation sites was located

FIG. 7. Comparative analysis of the amino acid sequences of the carboxyl half of small hsps. The sequences have been aligned to obtain maximal homology with hsp26-A. Numbers correspond to the amino acid position in hsp26-A. The sequences of *C. elegans* and *X. laevis* (Xen.) represent a composite of two cDNAs (3, 39), and the sequence of *G. max* hsp17.5 is a composite of four small soybean hsps (9, 30, 41). Additional amino acids below the *G. max* hsp17.5 composite sequence represent deviations from the consensus. Boxes indicate homology of hsp26-A with any of the other hsps. Dros., *D. melanogaster*.

27 to 31 bp downstream from an A+T-rich motif similar to the TATA box. The actual sequence of the TATA motif varied from TATAAAAG for site 1 (proximal) to TATATAGGTATAT for site 2 and to AATTAATA for site 3. Although the overall homology between the 5'-flanking region of *Gmhsp26-A* and other HS genes from soybean and *D. melanogaster* was very low, the TATA sequence for site 1 (most active) was identical to those of *Drosophila HSP26* and *HSP27* (18, 44).

An alternative view of the significance of the multiple transcriptional start sites of *Gmhsp26-A* is that they may simply be the result of spurious starts adjacent to strong promoter elements that normally mediate the initiation of transcription through the proximal TATA (site 1). This view was supported by the presence of a redundant AUG in the leader of the transcript originating from the distal start site (subpromoter 3), suggesting that this transcript may be very inefficiently translated.

The heat shock element (HSE) has been found in all eucaryotic HS genes analyzed to date and consists of a 14-bp imperfect repeat, 5'-CTnGAAnnTTCnAG-3' (35). This dyad confers heat inducibility if placed 15 to 28 bp upstream of the TATA box of heterologous promoters (29, 34, 35). In most cases, an 8-of-10-bp match with the consensus is required for thermal induction; however, a 6-of-10-bp match can function when additional copies are present (2, 4, 35). Although only two TATA-proximal HSEs are all that is needed for the *Drosophila HSP70* promoter (12), it appears that HSEs located outside the optimal spacing window (13 to 19 bp upstream of TATA) are less efficient and require additional TATA-proximal HSEs for full activity. This principle seems to apply to the *Drosophila HSP22*, *HSP23*, and *HSP26* promoters, which require three HSEs for efficient induction (24, 28, 33). In addition to their activity as TATA-proximal elements, HSEs also appear to have enhancerlike qualities as demonstrated by their bidirectional stimulation of heat-inducible transcription over considerable distances in the *Xenopus hsp70* promoter (5, 21).

In *Gmhsp26-A*, several overlapping HSE-like sequences were present upstream of each of the three putative TATA boxes. The homology to the *Drosophila* HSE was only 5 of 10 and 6 of 10 bp matches for subpromoters 1 and 2, respectively. Subpromoter 3 was the most inducible by heat and $CdCl_2$ (Table 1), whereas subpromoters 1 and 2 were less activated by these stresses, showing similar induction patterns. Subpromoter 3 displayed a cluster of four overlapping HSEs with various degrees of homology (5, 7, and 6 of 10 bp matches) to the *Drosophila* consensus. Although no single HSE-like sequence showed a very high degree of homology to the *Drosophila* consensus, it may be possible that low homology is partially compensated by the presence of multiple overlapping copies. This prediction is consistent with the findings of Kay et al. (21), which suggest that the number of HSEs can be a major determinant of promoter strength.

The overall structure of the *Gmhsp26-A* promoter appears to be more complex than that of a typical HS gene. The relative inducibility of transcription from each of the subpromoters differed depending on the stress treatment (Table 1), suggesting that each subpromoter may represent a specialization toward a specific set of stress conditions. Since a large number of stress conditions stimulate *Gmhsp26-A* transcription, an understanding of its regulation may reveal new information about a common set of regulatory proteins and *cis* elements that are responsible for the transcriptional induction of genes regulated by generalized physiological stress. Alternatively, this type of evolutionary solution (multiple subpromoters) for the design of a generalized stress promoter may indicate that there is no single set of transcriptional factors that activate genes in response to multiple stresses. The underlying mechanisms of transcriptional activation that function in the highly specialized stress responses, such as HS and heavy-metal induction, may not be entirely compatible when combined into a single promoter containing one TATA motif. It remains to be seen what role the HS transcription factor (32) and various other stress-specific factors play in the basal and stress-inducible transcription of this gene.

In animals, heavy metals such as cadmium are known to induce the synthesis of metallothioneins, which are small, cysteine-rich proteins that tightly bind heavy-metal ions. Regulatory metal-response elements (MREs) have been identified in the 5'-flanking regions of mouse and human metallothionein genes (7, 17, 20, 42, 45) and, if present in multiple copies, have been found to confer metal regulation to a heterologous gene. The MRE sequences can act as inducible transcription enhancers when located either upstream or downstream of heterologous genes (42).

Even though typical metallothionein proteins have not been found in plants, the cadmium-responsive promoter of *Gmhsp26-A* may share common features with the metallothionein gene promoter of animals. A search for sequence homology with the mammalian MRE (5'-CPyTTTGC PuPyPyCG-3') (45) in the 5'-flanking sequences of *Gmhsp26-A* identified several regions of limited homology. A region with a 7-of-12 bp match overlaps the HSE-like sequence of subpromoter 1, and two other regions, each with an 8-of-12-bp match, are located between the TATA motif and the transcriptional initiation site for subpromoters 2 and 3. In mammalian genes, redundant MREs are usually found upstream of position −40 bp (7, 20). In other members of the soybean small hsps, MRE-like sequences are located between the TATA and the cap site (9, 30). Although the possibility of conservation of promoter structure in heavy-metal-regulated genes exists between plants and animals, the question of whether these sequences function in a similar manner in higher plants has not been addressed.

Relationship of hsp26-A to other hsps. Not only is soybean hsp26-A highly diverged from other members of the small hsps in amino acid sequence, but its occurrence in response to environmental stress also differs from that of most other hsps in soybean. For example, it is present in unstressed tissue (2 h, 28°C), it is induced to relatively low levels by HS when compared with other hsps, and it is induced by many agents that are not associated with HS. Additionally, hsp26-A does not preferentially localize to organelle fractions during HS like other small hsps (26) but remains soluble in the cytoplasm, suggesting that its mode of action may differ from that of other small hsps. A similar stress induction pattern is seen with the maize *hsp70* gene expressed in mesocotyls, but in the case of hsp70, the protein shows relatively little divergence from that found in other organisms (37). The similarity in hydropathy profiles between soybean hsp26-A and the small hsps suggests that hsp26-A is related to this class of proteins but that it has diverged through evolution to assume functions beneficial in response to a wide variety of stresses in soybean in addition to HS.

ACKNOWLEDGMENTS

We thank Sabita Bandyopadhyay for excellent technical assistance and Dulce Barros and Paul C. Fox for critical review of this

manuscript. We also thank C. Hampton McRae for sequencing part of the far upstream region and Luis Mosquera for providing preliminary results on expression of *Gmhsp26-A* in sunflower tumors.

This work was supported by contracts to W.B.G. and J.L.K. from Lubrizol Genetics (Agrigenetics Research Associates, Ltd.).

LITERATURE CITED

1. **Ashburner, M., and J. J. Bonner.** 1979. The induction of gene activity in *Drosophila* by heat shock. Cell **17**:241–254.
2. **Ayme, A., R. Southgate, and A. Tissières.** 1985. Nucleotide sequences responsible for the thermal inducibility of the *Drosophila* small heat-shock protein genes in monkey COS cells. J. Mol. Biol. **182**:469–475.
3. **Bienz, M.** 1984. Developmental control of the heat shock response in *Xenopus*. Proc. Natl. Acad. Sci. USA **81**:3138–3142.
4. **Bienz, M.** 1985. Transient and developmental activation of heat-shock genes. Trends Biochem. Sci. **10**:157–161.
5. **Bienz, M., and H. R. B. Pelham.** 1986. Heat shock regulatory elements function as an inducible enhancer in the *Xenopus hsp70* gene and when linked to a heterologous promoter. Cell **45**:753–760.
6. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
7. **Carter, A. D., B. K. Felber, M. J. Walling, M.-F. Jubier, C. J. Schmidt, and D. H. Hamer.** 1984. Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc. Natl. Acad. Sci. USA **81**:7392–7396.
8. **Czarnecka, E., L. Edelman, F. Schöffl, and J. L. Key.** 1984. Comparative analysis of physical stress responses in soybean seedlings using cloned heat shock cDNAs. Plant Mol. Biol. **3**: 45–58.
9. **Czarnecka, E., W. B. Gurley, R. T. Nagao, L. A. Mosquera, and J. L. Key.** 1985. DNA sequence and transcript mapping of a soybean gene encoding a small heat shock protein. Proc. Natl. Acad. Sci. USA **82**:3726–3730.
10. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing corn mitochondrial 18S rDNA. Plasmid **13**: 31–40.
11. **Dudler, R., and A. A. Travers.** 1984. Upstream elements necessary for optimal function of the *hsp70* promoter in transformed flies. Cell **38**:391–398.
12. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
13. **Gilbert, W.** 1978. Why genes in pieces? Nature (London) **271**:501.
14. **Go, M.** 1981. Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature (London) **291**:90–92.
15. **Go, M.** 1983. Modular structural units, exons, and function in chicken lysozyme. Proc. Natl. Acad. Sci. USA **80**:1964–1968.
16. **Gurley, W. B., E. Czarnecka, R. T. Nagao, and J. L. Key.** 1986. Upstream sequences required for efficient expression of a soybean heat shock gene. Mol. Cell. Biol. **6**:559–565.
17. **Haslinger, A., and M. Karin.** 1985. Upstream promoter element of the human metallothionein-II_A gene can act like an enhancer element. Proc. Natl. Acad. Sci. USA **82**:8572–8576.
18. **Ingolia, T. D., and E. A. Craig.** 1981. Primary sequence of the 5′ flanking regions of the *Drosophila* heat shock genes in chromosome subdivision 67B. Nucleic Acids Res. **9**:1627–1642.
19. **Ingolia, T. D., and E. A. Craig.** 1982. Four small *Drosophila* heat shock proteins are related to each other and to mammalian a-crystallin. Proc. Natl. Acad. Sci. USA **79**:2360–2364.
20. **Karin, M., A. Haslinger, H. Holtgreve, R. I. Richards, P. Krauter, H. M. Westphal, and M. Beato.** 1984. Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-II_A gene. Nature (London) **308**:513–519.
21. **Kay, R. J., R. J. Boissy, R. H. Russnak, and E. P. M. Candido.** 1986. Efficient transcription of a *Caenorhabditis elegans* heat shock gene pair in mouse fibroblasts is dependent on multiple promoter elements which can function bidirectionally. Mol. Cell. Biol. **6**:3134–3143.
22. **Key, J. L., W. B. Gurley, R. T. Nagao, E. Czarnecka, and M. A. Mansfield.** 1985. Multigene families of soybean heat shock proteins. NATO Adv. Studies Inst. Ser. A Life Sci. **83**:81–100.
23. **Key, J. L., C.-Y. Lin, and Y.-M. Chen.** 1981. Heat shock proteins of higher plants. Proc. Natl. Acad. Sci. USA **78**:3526–3530.
24. **Klemenz, R., and W. J. Gehring.** 1986. Sequence requirement for expression of the *Drosophila melanogaster* heat shock protein *hsp22* gene during heat shock and normal development. Mol. Cell. Biol. **6**:2011–2019.
25. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
26. **Lin, C.-Y., J. K. Roberts, and J. L. Key.** 1984. Acquisition of thermotolerance in soybean seedlings. Plant Physiol. **74**:152–160.
27. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
28. **Mestril, R., D. Rungger, P. Schiller, and R. Voellmy.** 1985. Identification of a sequence element in the promoter of the *Drosophila melanogaster hsp22* gene that is required for heat activation. EMBO J. **4**:2971–2976.
29. **Mirault, M.-E., R. Southgate, and E. Delwart.** 1982. Regulation of heat-shock genes: a DNA sequence upstream of *Drosophila hsp70* genes is essential for their induction in monkey cells. EMBO J. **1**:1279–1285.
30. **Nagao, R. T., E. Czarnecka, W. B. Gurley, F. Schöffl, and J. L. Key.** 1985. Genes for low-molecular-weight heat shock proteins of soybeans: sequence analysis of a multigene family. Mol. Cell. Biol. **5**:3417–3428.
31. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.
32. **Parker, C. S., and J. Topol.** 1984. A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an *hsp70* gene. Cell **37**:273–283.
33. **Pauli, D., A. Spierer, and A. Tissières.** 1986. Several hundred base pairs upstream of *Drosophila hsp23* and *hsp26* genes are required for their heat induction in transformed flies. EMBO J. **5**:755–761.
34. **Pelham, H. R. B.** 1982. A regulatory upstream promoter element in the *Drosophila hsp70* heat-shock gene. Cell **30**:517–528.
35. **Pelham, H. R. B., and M. Bienz.** 1982. A synthetic heat-shock promoter element confers heat-inducibility on the herpes simplex virus thymidine kinase gene. EMBO J. **1**:1473–1477.
36. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
37. **Rochester, D. E., J. A. Winter, and D. M. Shah.** 1986. The structure and expression of maize genes encoding the major heat shock protein, hsp70. EMBO J. **5**:451–458.
38. **Russnak, R. H., D. Jones, and E. P. M. Candido.** 1983. Cloning and analysis of cDNA sequences coding for two 16 kilodalton heat shock proteins (hsps) in *Caenorhabditis elegans*: homology with the small hsps of *Drosophila*. Nucleic Acids Res. **11**:3187–3205.
39. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
40. **Schöffl, F., and J. L. Key.** 1982. An analysis of mRNAs for a group of heat shock proteins of soybean using cloned cDNAs. J. Mol. Appl. Genet. **1**:301–314.
41. **Schöffl, F., E. Raschke, and R. T. Nagao.** 1984. The DNA sequence analysis of soybean heat-shock genes and identification of possible regulatory promoter elements. EMBO J. **3**: 2491–2497.
42. **Serfling, E., A. Lubbe, K. Dorsch-Häsler, and W. Schaffner.** 1985. Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes.

EMBO J. **4**:3851–3859.

43. **Sollner-Web, B., and R. H. Reeder.** 1979. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis*. Cell **18**:485–499.
44. **Southgate, R., A. Ayme, and R. Voellmy.** 1983. Nucleotide sequence analysis of the *Drosophila* small heat shock gene cluster at locus 67B. J. Mol. Biol. **165**:35–57.
45. **Stuart, G. W., P. F. Searle, H. Y. Chen, R. L. Brinster, and R. D. Palmiter.** 1984. A 12-base-pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc. Natl. Acad. Sci. USA **81**:7318–7322.
46. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
47. **Yost, H. J., and S. Lindquist.** 1986. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell **45**:185–193.