Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3726–3730, June 1985
Genetics

DNA sequence and transcript mapping of a soybean gene encoding a small heat shock protein

(promoter/stress/*Glycine max*/transcription)

EVA CZARNECKA*, WILLIAM B. GURLEY*, RONALD T. NAGAO†, LUIS A. MOSQUERA*, AND JOE L. KEY†

*Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611; and †Botany Department, University of Georgia, Athens, GA 30602

Communicated by Norman H. Giles, December 12, 1984

ABSTRACT **The DNA sequence of a gene (*Gmhsp17.5-E*) encoding a small heat shock protein of soybean, *Glycine max*, has been determined. Nuclease S1 mapping of the 5′ terminus of the corresponding RNA indicates that the start site for transcription is located 82 bases upstream from the coding region and 24 bases downstream from a "TATA"-like region (-T-T-T-A-A-A-T-A-). The 5′ flanking region of *Gmhsp17.5-E* contains two imperfect dyads that closely resemble regulatory elements present in the promoters of heat-inducible genes of *Drosophila*. One, positioned 18 bases upstream from the TATA-like region, shows 90% homology to the *Drosophila* heat shock consensus sequence. The other overlaps an upstream TATA sequence and is located at position −213. Analysis of the derived amino acid sequence indicates that the protein encoded by *Gmhsp17.5-E* is related structurally to the four small heat shock proteins of *Drosophila*. This relationship is most evident by comparison of hydropathy profiles; they show conservation of several major hydrophilic and hydrophobic regions, which suggests that these proteins have common structural features.**

In *Drosophila*, the four small heat shock proteins (HSPs) hsp22, -23, -26, and -27 represent a single class of proteins. Their relatedness is evidenced by the high degree of amino acid conservation among the group and by the similarity in hydropathy profiles (1). In higher plants, the low molecular weight HSPs represent a complex array of proteins comprising 20–30 members (2–4) which are the predominant proteins synthesized during thermal stress. These plant HSPs are related: hybridization studies with cDNA (5) and genomic clones (unpublished work) indicate that in soybean most members of this class cross-hybridize to various degrees. In addition, some of these cDNAs also share homology with heat shock mRNAs from a variety of other plant species, such as pea, millet, corn, and sunflower (6), as well as tomato and tobacco (unpublished data).

The 5′ flanking sequences of *Drosophila* heat shock genes contain a 14-nucleotide dyad that is located upstream from the "TATA"-like region (7). This dyad is essential in the heat induction of transcription of the *Drosophila hsp70* gene in monkey cells (7, 8) and *Xenopus* oocytes (9, 10) and is necessary for optimal function of an *hsp70-Adh* fused gene in transformed *Drosophila* (11). The mechanism of heat activation of the heat shock promoter seems to be highly conserved among animals, since the cloned *hsp70* gene of *Drosophila* is transcriptionally heat-inducible when introduced into a variety of heterologous systems.

In this study we report the DNA sequence for one of the most actively expressed small *HSP* genes in soybean and map the 5′ and 3′ termini of the RNA product. The 5′ flanking region is shown to contain sequences similar to those present in *Drosophila* heat shock genes. Analysis of the deduced amino acid sequence suggests that the small HSPs found in soybean are not unique to plants but represent a structural and functional class of stress proteins widely dispersed among eukaryotes.

MATERIALS AND METHODS

Soybean Genomic DNA Library. Total soybean (*Glycine max* var. Corsoy) DNA was partially digested with *Mbo* I and ligated into the *Bam*HI site of the cloning vector λ1059 (12). The λ1059 library was constructed by J. Slightom and Y. Ma (Agrigenetics Advanced Research Laboratory, Madison, WI) and screened as described by Nagao *et al.* (13) by using the ^{32}P-labeled insert of the heat shock-specific cDNA clone pFS2019 (5). The λ clone characterized in this study was designated hsE2019. Restriction fragments (see Fig. 1) showing homology to the cDNA pFS2019 were inserted into their respective sites of pBR322 (14) and pUC8 (15) and cloned.

Transcript Mapping by Nuclease S1 Hybrid Protection. Total RNA isolation and poly(A)$^+$ RNA purification from control (2 hr at 28°C) and heat-shocked (2 hr at 40°C) intact soybean seedlings was performed as described (16). The 5′ and 3′ termini of soybean mRNA homologous to λ clone hsE2019 were determined by nuclease S1 mapping with end-labeled DNA hybridization probes. The heat shock gene identified in this study was designated *Gmhsp17.5-E* (*Glycine max* heat shock protein, 17.5 kDa). For 5′-end analysis, the BB1.55 fragment was end-labeled at the 5′ termini (17) and redigested with *Eco*RI, and the resulting 250-base-pair (bp) *Bam*HI–*Eco*RI fragment (BEO.25) was isolated. This fragment was hybridized as described by Favaloro *et al.* (18) at various temperatures (42–50°C) to total poly(A)$^+$ RNA (1.2–2.5 μg) from soybean and then digested with nuclease S1 (50–200 units/ml) at 15°C for 30 min. For determination of the 3′ termini, the HB1.45 fragment was end-labeled by filling in the recessed 3′ end with the large fragment of DNA polymerase I (19) and then hybridized to poly(A)$^+$ RNA, and the RNA·DNA hybrid was digested with nuclease S1 (200–800 units/ml). DNA from the protected hybrids was precipitated with isopropanol and detected by autoradiography after fractionation by electrophoresis in 6% polyacrylamide/urea sequencing gels.

DNA Sequencing. Both the dideoxynucleotide chain-termination technique of Sanger *et al.* (20) and the chemical cleavage method of Maxam and Gilbert (17) were utilized for sequencing studies. The DNA sequence was determined for both strands in regions of the gene analyzed by chemical cleavage procedures, by separately labeling fragments at the 3′ and 5′ termini. Restriction fragments HB1.45 and EH1.75 were inserted into the replicative forms of bacteriophage M13 mp9 (21). Overlapping M13 deletions, starting at the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HSP, heat shock protein; hsp*nn*, a specific heat shock protein of *nn* kDa; bp, base pair(s).

*Bam*HI site of the HB1.45 insert, were constructed as described (22) and sequenced by the chain-termination technique (20).

RESULTS

DNA Sequence. Fig. 1 illustrates the restriction map of the 7.0-kilobase *Hin*dIII fragment 2 and indicates the relative position and orientation of the heat shock gene *Gmhsp17.5-E*. The nucleotide sequence of the gene is presented in Fig. 2 along with the derived amino acid sequence. The open reading frame in the transcribed portion of the gene extends for 154 codons and is predicted to code for a 17.5-kDa protein.

The cDNA used in screening to select gene *Gmhsp17.5-E* was sequenced (340 bases) and found to match genomic clone hsE2019 between nucleotides +526 and +866 with the exception of two bases at positions 651 and 755. This minor discrepancy in sequence probably reflects varietal differences, since the cDNA was prepared from RNA isolated from the Wayne cultivar of soybean. This partial cDNA covers the entire noncoding 3′ tail of the *Gmhsp17.5-E* gene and 19 nucleotides within the coding region (Fig. 2).

Determination of the mRNA 5′ Terminus. Restriction fragment BEO.25 was 5′-labeled at the *Bam*HI site and hybridized at 42°C to either control (28°C) or heat shock (40°C) poly(A)$^+$ RNA under conditions (18) favoring the formation of RNA·DNA duplexes. After removal of the single-stranded tails by digestion with nuclease S1, the labeled DNA was analyzed in polyacrylamide/urea gels as shown in Fig. 3. Protected DNA fragments were evident (Fig. 3*A*) only after hybridization to heat shock RNA. One major fragment of 158 nucleotides corresponding to the 5′ terminus was obtained at the lowest concentration (50 units/ml) of nuclease S1. The other, smaller fragments appearing as the level of nuclease S1 was increased to 200 units/ml probably result from nuclease attack at the terminus and other potentially unstable regions of the RNA·DNA duplex. These RNA mapping results indicate that transcription of *Gmhsp17.5-E* is induced in response to elevated temperatures, as shown earlier by RNA blot-hybridization analyses (16), and that the major 5′ terminus is positioned 28–32 bases downstream from the TATA-like motif T-T-T-A-A-A-T-A.

A more detailed analysis of the protected DNA fragment in high-resolution sequencing gels (Fig. 3*B*) was performed, using a sequence ladder of the fragment BEO.25 as a size

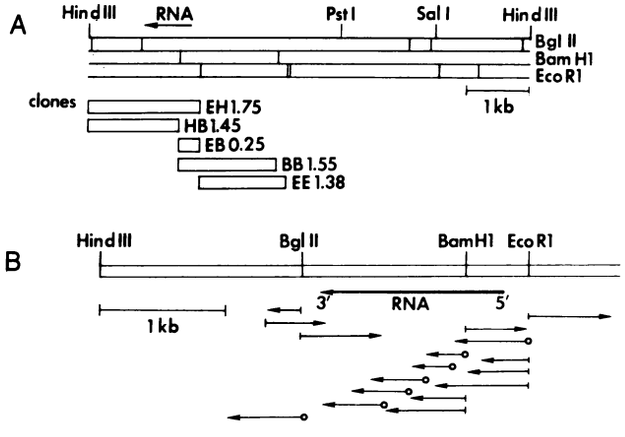

FIG. 1. (*A*) Restriction map of a cloned *Hin*dIII fragment containing heat shock gene *Gmhsp17.5-E*. The position and polarity of the transcribed region is indicated by an arrow. Restriction fragments subcloned for nucleotide sequence determination and transcript mapping are shown below the map. (*B*) The sequencing strategy is denoted by arrows:⟜⟞, dideoxy method with M13 deletion clones;⟞, chemical-cleavage method. kb, Kilobase.

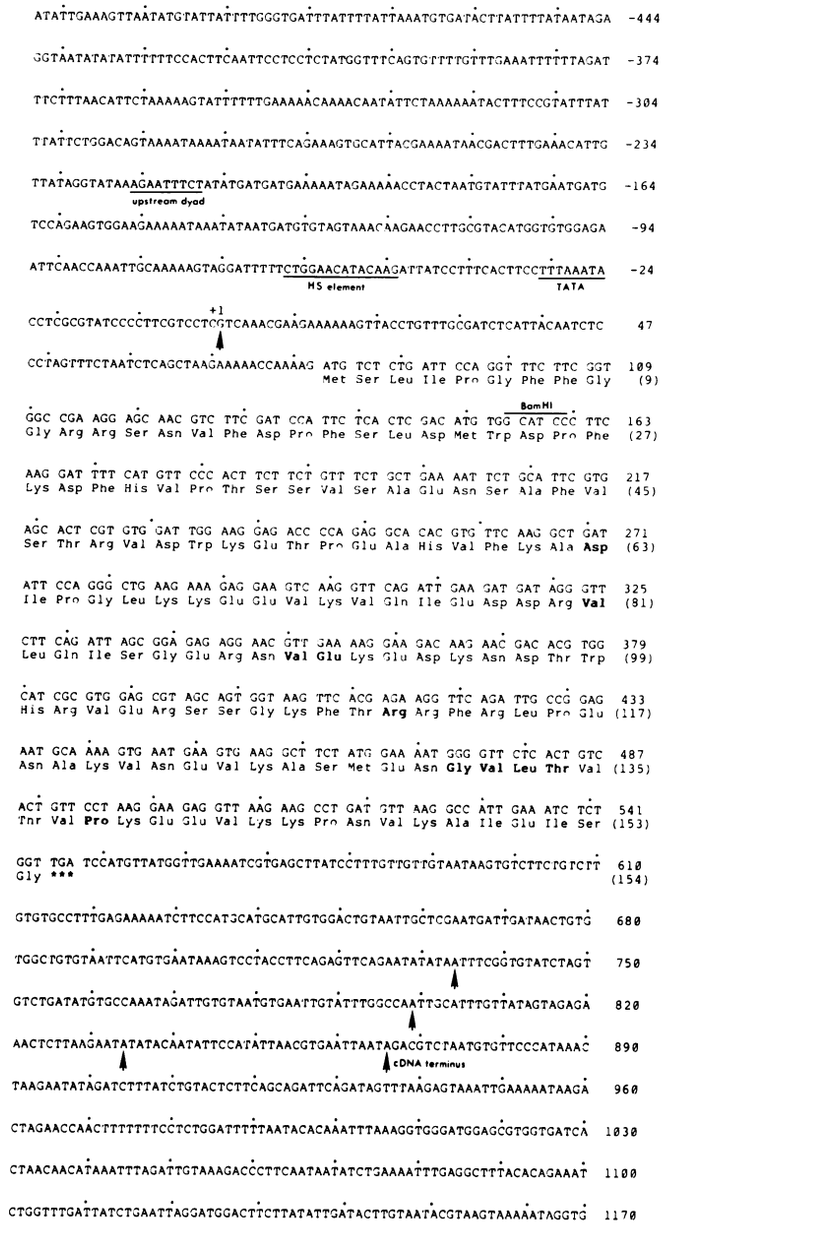

FIG. 2. Complete nucleotide and deduced amino acid sequence of *Gmhsp17.5-E*. Nucleotides are numbered from the start site for transcription, and amino acid residues are numbered (in parentheses) from the Met initiator codon. Arrowheads denote termini of mRNA as determined by nuclease S1 hybrid-protection analysis. The 3′ terminus at position 866 is identical to the cloned cDNA terminus. The TATA-like motif and the *Drosophila* heat shock consensus sequence and far-upstream dyad homologies are underlined.

marker. The uppermost major band (denoted by arrow, Fig. 3*B*) establishes the precise position of the 5′ terminus at the guanylate residue designated +1 in Fig. 2. This determination incorporates a 1.5-nucleotide correction, since fragments generated by nuclease S1 migrate slower than comparable Maxam–Gilbert chemical-cleavage fragments on high-resolution gels (23). The nuclease S1 digestions were conducted at several concentrations of enzyme in order to differentiate between 3′ overhangs of 2 or 3 nucleotides resistant to removal by nuclease S1 and the true start of the RNA·DNA hybrid.

Determination of the 3′ Termini. To establish the 3′ termini of *Gmhsp17.5-E*, the *Bam*HI–*Hin*dIII 1.45-kilobase fragment (HB1.45) was 3′ end-labeled with ^{32}P and used in nuclease S1 hybrid-protection analyses. As shown in Fig. 4, there

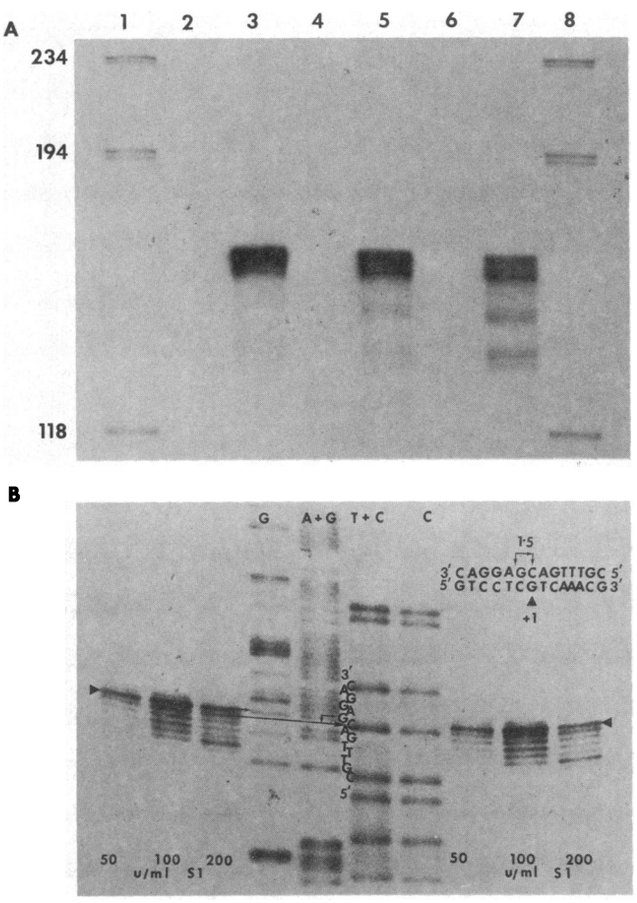

FIG. 3. Nuclease S1 mapping of the 5′ terminus of poly(A)$^+$ RNA homologous to *Gmhsp17.5-E*. (*A*) ^{32}P-labeled DNA protected from nuclease S1 attack by 1 μg of RNA isolated from soybean seedlings incubated at 28°C (lanes 2, 4, and 6) or at 40°C (lanes 3, 5, and 7). Nuclease S1 concentrations were 50 units/ml (lanes 2 and 3), 100 units/ml (lanes 4 and 5), and 200 units/ml (lanes 6 and 7). Markers (lanes 1 and 8) are ^{32}P-labeled *Hae* III digestion products of φX174 DNA; sizes (in nucleotides) are given at left. (*B*) Nucleotide-level precision obtained by using a DNA sequence ladder for size calibration. One microgram of poly(A)$^+$ RNA isolated from seedlings incubated at 40°C was used for each hybrid protection. Nuclease S1 concentrations in units (u)/ml are given below the lanes. See text for explanation of 1.5-nucleotide correction.

seem to be four 3′ termini of *Gmhsp17.5-E*, positioned 580 (major terminus), 645, 680, and 712 bp downstream from the labeled nucleotide of the internal *Bam*HI site. Four major transcripts with total lengths of 734, 799, 834, and 866 bases [excluding the poly(A) tail] are therefore predicted during heat stress and may account for the spread of bands observed in RNA blot analysis (16). Under some conditions of nuclease S1 digestion, weak bands were seen below the major band at 580 bases. We believe that these bands result from the cross-hybridization of RNA transcribed from closely related genes encoding HSPs in the same low molecular weight class (unpublished work). The disappearance of the weak bands as the temperature of hybridization was increased (Fig. 4) supports this assumption.

The 734-base transcript (which corresponds to the 580-base protected band) is the most abundant *Gmhsp17.5-E* RNA (Fig. 4). The 3′ terminus is located within the sequence -T-A-A-T-T-T-, 28 nucleotides downstream from the mammalian consensus polyadenylylation signal -A-A-T-A-A-A- (24). The three minor transcripts lack the exact consensus sequence for poly(A) addition. For two of these RNAs, regions of near homology are located upstream from the terminal base: -A-A-T-A-G-A-, 27 bases upstream from the 799-base transcript, and -A-A-T-A-T-A-, 27 bases upstream from the 866-base transcript. The 3′ termini of all four *Gmhsp17.5-E* transcripts are positioned within very A+T-rich, perfect or imperfect, short inverted repeats.

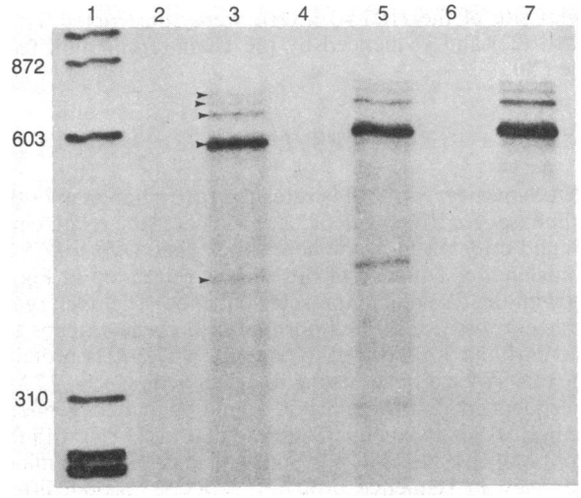

FIG. 4. Nuclease S1 mapping of the 3′ termini of poly(A)$^+$ RNA homologous to *Gmhsp17.5-E*. RNA (1 μg) from soybean seedlings incubated at 28°C (lanes 2, 4, and 6) or 40°C (lanes 3, 5, and 7) was used in nuclease S1 hybrid-protection analyses. Hybridization temperatures were 46°C (lanes 2 and 3), 48°C (lanes 4 and 5), and 52°C (lanes 6 and 7). Nuclease S1 concentration was 200 units/ml. Markers (lane 1) are ^{32}P-labeled *Hae* III digestion products of φX174 DNA; sizes (in nucleotides) are given at left.

Analysis of Deduced Amino Acid Sequence. A comparison of the deduced amino acid sequence of Gmhsp17.5-E (soybean hsp17.5) with those of the four low molecular weight *Drosophila* HSPs and of bovine α-crystallin indicates a degree of conservation both in the hydropathy profiles and in the amino acid sequences of these six proteins. The hydropathy plots (25) of soybean hsp17.5 and *Drosophila* hsp23 and -27 are presented in Fig. 5. Hydropathy profiles (1) of *Drosophila* hsp22, -23, -26, and -27 were compared (data not shown) and four major regions of similarity were identified. Two of these regions (regions 2 and 3, Fig. 5) are also present in hsp16 of *C. elegans* (26), hsp30 of *Xenopus* (27), and bovine α-crystallin (1). The amino-terminal portion (region 1) of three of the four *Drosophila* HSPs (hsp22 is the exception) and the soybean hsp17.5 are hydrophobic. Regions 2 and 3 are present in all of the *Drosophila* small HSPs, α-crystallin, and the Gmhsp17.5-E protein. Since this feature is similar in all six proteins (28), these seem likely to be regions associated with or required for aggregate formation. The fourth region of Gmhsp17.5-E forms a hydrophilic peak located near the carboxyl terminus and is present in all of the *Drosophila* proteins except hsp22 and bovine α-crystallin.

Hydrophobic peak 3 coincides with the region of highest conservation among all three proteins and corresponds to Gmhsp17.5-E residues 124–139. This 16 amino acid sequence is 37–50% conserved in *Drosophila* proteins hsp22, -23, -26, and -27 and 31% conserved in α-crystallin. This hydrophobic area also contains the amino acid sequence -Asn-Gly-Val-Leu-Thr-, which is a near match to the -Asp-Gly-Val-Leu-Thr- sequence appearing in α-crystallin and all four *Drosophila* proteins (1). Alignment of the carboxyl-terminal halves (positions 63–154) of the small HSPs of *Drosophila*, *C. elegans* (26), and *Xenopus* (27) shows that 28% of the soybean amino acids are also found at the same position in at least one of the other proteins.

DISCUSSION

We have determined the DNA sequence and mapped the transcript of the soybean heat shock gene *Gmhsp17.5-E* to

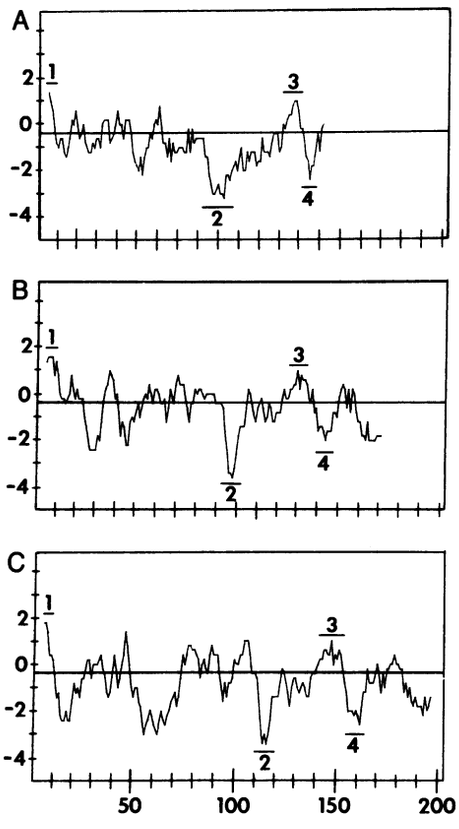

FIG. 5. Hydropathy profiles of soybean hsp17.5 (*A*) and of *Drosophila* hsp23 (*B*) and hsp27 (*C*). Plots were constructed (see ref. 25) by progressively moving along the amino acid sequence and averaging the hydropathy index for each stretch of 9 amino acids. Points above the horizontal line correspond to hydrophobic domains, whereas points below correspond to hydrophilic domains. Regions 1–4 are delineated by numbered bars and represent similar structural features of the proteins.

better understand the mechanisms operative in the heat shock response and the development of thermal tolerance in higher plants. At present *Drosophila* heat shock genes serve as a model and point of reference for all studies dealing with heat shock gene structure. A comparison of *Gmhsp17.5-E* with *Drosophila* heat shock genes indicates that promoter design and certain aspects of protein structure appear to have been maintained in both animal and higher plant evolution. Examination of the 5′ flanking region of this plant gene indicates a remarkable degree of similarity between the presumptive promoter structure of plant heat shock genes and those of *Drosophila*. In addition, analysis of the derived amino acid sequence of soybean hsp17.5 and the corresponding hydropathy profile reveals a structural relationship to the small HSPs of *Drosophila*.

Although actual sequence homology within the 5′ flanking regions of *Gmhsp17.5-E* and *Drosophila* heat shock genes is very limited, the presence of common sequence elements in both systems suggests that the overall promoter architecture may be similar. The strongest evidence for conservation of promoter structure is homology in *Gmhsp17.5-E* to the heat shock response element of *Drosophila*. This element in *Drosophila* forms an inverted repeat and is located 15–28 bp from the 5′ end of the TATA box. The consensus sequence for this regulatory element is -C-T-n-G-A-A-n-n-T-T-C-n-A-G- (n, any deoxynucleoside) (1) with ≥80% homology seen in six of the seven *Drosophila* genes characterized. As in *Drosophila*, this dyad element in *Gmhsp17.5-E* is contained within a larger inverted repeat extending in the 5′ direction. In heterologous systems, an 8-out-of-10 base match has been shown to be sufficient for heat activation of transcription (7–9). From these same studies, the effectiveness of this element was shown to be also strongly dependent on its proximity to the 5′ end of the TATA box, with a spacing of 10–30 bp being optimal. In the soybean *Gmhsp17.5-E* gene, the sequence -C-T-g-G-A-A-c-a-T-a-C-a-A-G- shows a 90% match with the *Drosophila* consensus sequence and is located in the 5′ flanking region at positions −49 to −62. In addition, the dyad element in *Gmhsp17.5-E* is located 17 bp upstream from the TATA-like motif, falling well within the spacing requirements for heat inducibility seen in *Drosophila*.

The soybean *Gmhsp17.5-E* gene has an A+T-rich region (-T-T-T-A-A-A-T-A-) similar to the TATA box (29) characteristic of many eukaryotic genes. Justification for classifying this sequence as a TATA box is based on its position 24 bases upstream from the transcriptional start site and on its neighboring sequence environment. The significance of this sequence is further supported by its presence at essentially the same position in two other closely related soybean heat shock genes (unpublished work).

Both *Drosophila* and soybean contain a distinctive palindromic sequence in the far-upstream portion of some heat shock genes. This sequence in *Gmhsp17.5-E*, -A-G-A-A-T-T-T-C-T-, is nearly identical to the *Drosophila* sequence -A-G-A-A-A-T-T-T-C-T-, which is present once in *hsp26* and *hsp22* and twice in *hsp70* (1). For the *Drosophila* heat shock genes, the spacing upstream from the TATA box ranges from 153 to 256 bp. In *Gmhsp17.5-E* this far-upstream dyad element is similarly spaced 182 bp from the 5′ end of the TATA-like motif. An additional similarity between the two systems is the overlap of the far-upstream dyad with a second TATA motif in *Gmhsp17.5-E* and in *hsp26* and *hsp70* of *Drosophila*. Although this far-upstream sequence has not been shown to be essential in either *Drosophila* or heterologous expression systems, its presence at similar locations in both *Drosophila* and soybean heat shock genes suggests that it may play a common role in the regulation of the heat shock response in these two organisms.

Repeated -C-T- sequences, which correlate with nuclease S1-sensitive sites in the upstream parts of the promoters of the *hsp22*, *hsp26*, and *hsp70* genes of *Drosophila* (30) are not present in *Gmhsp17.5-E*. However, two short blocks of alternating purine-pyrimidine sequences centered at comparable positions (−132 and −107) have the potential to form Z-DNA (31) and may therefore have a similar effect on chromatin structure. It is also interesting to note that the TATA-proximal block (at position −112) is adjacent to the purine-rich sequence -G-G-T-G-T-G-G-A-G-A-A-T-T-C- (positions −113 to −90), which shows considerable homology (11 bases out of 14) to the simian virus 40 enhancer core (-G-G-T-G-T-G-G-A-A-A-G-T-C-C-) (32).

The sequences surrounding the transcriptional initiation site of the soybean *Gmhsp17.5-E* gene show a general similarity to other characterized eukaryotic genes. Although the nucleotide at position +1 differs from other characterized heat shock genes by the presence of a guanine instead of the typical adenine, this initiation point lies within the sequence -T-C-G-T-C-A-, which shows limited homology (3 out of 7 bases) to both the mammalian cap-site consensus sequence -Y-C-A-T-T-C-R- (Y, pyrimidine nucleoside; R, purine nucleoside) (33) and the *Drosophila* heat shock cap-site consensus sequence (G/C)-(A/T)-C-A-G-(A/T) (34). A more striking homology (5 out of 7 bases) is found between the cap site (-C-G-T-C-A-) of the mouse metallothionein-IIA gene (35) and the analogous site of the soybean *Gmhsp17.5-E* gene.

Another example of sequence redundancy in *Gmhsp17.5-E* is present in the region between the TATA motif and the start site for transcription. Sequences with homology (9 out of 13) to the mammalian metal-ion response element (MRE)

are repeated twice in this 23-nucleotide stretch. The MRE motif (T/C)-(G/T)-C-G-N_n-C-C-C-G-G-N_n-C-(T/C)-C ($n = 0$ or 1) has been shown by deletion analysis to be essential for transcriptional induction of the human metallothionein-IIA gene by cadmium ions (36). By analogy, it is tempting to speculate that this sequence may be involved either in the general or metal-ion-specific regulation of *Gmhsp17.5-E*, since transcripts from this gene and other soybean stress genes are induced to moderate levels in response to cadmium (16).

Many of the proteins induced by thermal stress are conserved to varying degrees among a wide spectrum of organisms. The highest degree of evolutionary conservation exists in the HSP 68- to 70-kDa group and in the HSP 80- to 90-kDa group (37). The third class of HSPs, those from 12- to 30-kDa, is also induced upon heat shock in many organisms. In *Drosophila*, the small HSP genes are clustered on an 11-kilobase stretch of DNA, share DNA sequence homology, and encode proteins that show a considerable degree of amino acid conservation. Outside of *Drosophila*, cross-hybridization and antigenic relatedness has not been demonstrated between major taxonomic groups, although hsp16 of *C. elegans* (26) and hsp30 of *Xenopus* (28) share considerable amino acid homology. From the results obtained in this study, we propose that the small HSPs of *Drosophila* and soybean comprise a widely dispersed class of thermally induced proteins that are related by the criteria of amino acid conservation and similarity in functional domains revealed by analysis of hydropathy profiles. It is not possible, based on the low homologies between small HSPs, to tell whether these proteins have diverged greatly from a single progenitor protein or are simply analogous proteins that arose through convergent evolution.

In soybean, there are 20 or more small HSPs that can be grouped into at least three multicomponent families and range in size from 15 to 24 kDa (unpublished work). In higher plants, it appears that the majority of proteins synthesized during thermal stress are members of this class. The selective localization of small HSPs in soybean to ribosomal and nuclear fractions is correlated with the acquisition of thermal tolerance (38). The striking similarity of the hydropathy profile of the Gmhsp17.5-E protein of soybean to those of the low molecular weight HSPs of *Drosophila*, together with the presence of identical amino acids at certain positions, suggests that the Gmhsp17.5-E protein is a member of a conserved class of heat (or stress)-inducible proteins that may be widely dispersed among eukaryotic organisms.

We thank Beth Laughner for technical assistance and L. Edelman for providing a sample of soybean poly(A)$^+$ RNA. This work was supported by contracts from Agrigenetics Research Associates Ltd. This paper is Florida Agricultural Experiment Station Journal Series no. G072.

1. Southgate, R., Ayme, A. & Voellmy, R. (1983) *J. Mol. Biol.* **165,** 35–37.
2. Key, J. L., Lin, C.-Y. & Chen, Y. M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3526–3530.
3. Altschuler, M. & Mascarenhas, J. P. (1982) *Plant Mol. Biol.* **1,** 103–115.
4. Cooper, P., Ho, T. D. & Hauptmann, R. M. (1984) *Plant Physiol.* **75,** 431–441.
5. Schoffl, F. & Key, J. L. (1982) *J. Mol. Appl. Genet.* **1,** 301–314.
6. Key, J. L., Czarnecka, E., Lin, C.-Y., Kimpel, J., Mothershed, C. & Schoffl, F. (1983) in *Current Topics in Plant Biochemistry and Physiology*, eds. Randall, D. D., Blevins, D. G., Larson, R. L. & Rapp, B. J. (Univ. of Missouri Press, Columbia, MO), Vol. 2, pp. 107–117.
7. Pelham, H. R. B. (1982) *Cell* **30,** 517–528.
8. Mirault, M.-E., Southgate, R. & Delwart, E. (1982) *EMBO J.* **1,** 1279–1285.
9. Pelham, H. R. B. & Bienz, M. (1982) *EMBO J.* **1,** 1473–1477.
10. Voellmy, R. & Rungger, D. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1776–1780.
11. Dudler, R. & Travers, A. A. (1984) *Cell* **38,** 391–398.
12. Karn, J., Brenner, S., Barnett, L. & Cesareni, G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5172–5176.
13. Nagao, R. T., Shah, D. M., Eckenrode, V. K. & Meagher, R. B. (1981) *DNA* **1,** 1–9.
14. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L. & Boyer, H. W. (1977) *Gene* **2,** 95–113.
15. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
16. Czarnecka, E., Edelman, L., Schoffl, F. & Key, J. L. (1984) *Plant Mol. Biol.* **3,** 45–58.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
18. Favaloro, J., Treisman, R. & Kamen, R. (1980) *Methods Enzymol.* **65,** 718–749.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 113.
20. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143,** 161–178.
21. Messing, J. (1983) *Methods Enzymol.*, eds. Wu, R., Grossman, L. & Moldave, K. (Academic, New York), Vol. 101, pp. 20–78.
22. Barnes, W. M. & Bevan, M. (1983) *Nucleic Acids Res.* **11,** 349–368.
23. Sollner-Webb, B. & Reeder, R. H. (1979) *Cell* **18,** 485–499.
24. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
25. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
26. Russnak, R. H., Jones, D. & Candido, E. P. M. (1983) *Nucleic Acids Res.* **11,** 3187–3205.
27. Bienz, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3138–3142.
28. Ingolia, T. D. & Craig, E. A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2360–2364.
29. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
30. Mace, H. A. F., Pelham, H. R. B. & Travers, A. A. (1983) *Nature (London)* **304,** 555–557.
31. Nordheim, A. & Rich, A. (1983) *Nature (London)* **303,** 674–679.
32. Laimins, L. A., Khoury, G., Gorman, C., Howard, B. & Gruss, P. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6453–6457.
33. Busslinger, M., Portmann, R., Irminger, J.-C. & Birnstiel, M. L. (1980) *Nucleic Acids Res.* **8,** 957–977.
34. Ingolia, T. D. & Craig, E. A. (1981) *Nucleic Acids Res.* **9,** 1627–1642.
35. Glanville, N., Durnam, D. M. & Palmiter, R. D. (1981) *Nature (London)* **292,** 267–269.
36. Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M. & Beato, M. (1984) *Nature (London)* **308,** 513–519.
37. Kelley, P. M. & Schlesinger, M. J. (1982) *Cell. Mol. Biol.* **2,** 267–274.
38. Lin, C.-Y., Roberts, J. K. & Key, J. L. (1984) *Plant Physiol.* **74,** 152–160.