MOLECULAR AND CELLULAR BIOLOGY, Dec. 1985, p. 3417–3428
0270-7306/85/123417-12$02.00/0

Vol. 5, No. 12

Genes for Low-Molecular-Weight Heat Shock Proteins of Soybeans: Sequence Analysis of a Multigene Family

RONALD T. NAGAO,[1]* EWA CZARNECKA,[2] WILLIAM B. GURLEY,[2] FRITZ SCHÖFFL,[3] AND JOE L. KEY[1]

Botany Department, University of Georgia, Athens, Georgia 30602[1]; Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611[2]; and Fakultät für Biologie (Genetik), Universität Bielefeld, D-4800 Bielefeld 1, Federal Republic of Germany[3]

Received 22 July 1985/Accepted 19 September 1985

Soybeans, *Glycine max*, synthesize a family of low-molecular-weight heat shock (HS) proteins in response to HS. The DNA sequences of two genes encoding 17.5- and 17.6-kilodalton HS proteins were determined. Nuclease S1 mapping of the corresponding mRNA indicated multiple start termini at the 5′ end and multiple stop termini at the 3′ end. These two genes were compared with two other soybean HS genes of similar size. A comparison among the 5′ flanking regions encompassing the presumptive HS promoter of the soybean HS-protein genes demonstrated this region to be extremely homologous. Analysis of the DNA sequences in the 5′ flanking regions of the soybean genes with the corresponding regions of *Drosophila melanogaster* HS-protein genes revealed striking similarity between plants and animals in the presumptive promoter structure of thermoinducible genes. Sequences related to the *Drosophila* HS consensus regulatory element were found 57 to 62 base pairs 5′ to the start of transcription in addition to secondary HS consensus elements located further upstream. Comparative analysis of the deduced amino acid sequences of four soybean HS proteins illustrated that these proteins were greater than 90% homologous. Comparison of the amino acid sequence for soybean HS proteins with other organisms showed much lower homology (less than 20%). Hydropathy profiles for *Drosophila*, *Xenopus*, *Caenorhabditis elegans*, and *G. max* HS proteins showed a similarity of major hydrophilic and hydrophobic regions, which suggests conservation of functional domains for these proteins among widely dispersed organisms.

All groups of organisms investigated undergo a response to high temperature referred to as heat shock (HS) (43). The HS response was first discovered in *Drosophila melanogaster* and has been studied in considerable detail in that organism (4, 43). This response is characterized by control mechanisms which are operative at the levels of both transcription and translation and is generally characterized by the induction of synthesis of a new set of proteins (HS proteins), decreased synthesis of most normal proteins, and the acquisition of thermotolerance to a nonpermissive (or lethal) HS temperature by prior exposure to permissive elevated temperatures. The induction of HS proteins is dependent on the transcriptional activation of a unique set of genes at the elevated or HS temperature. In *D. melanogaster*, four HS proteins in the range of 22 to 27 kilodaltons (kDa) and three high-molecular-weight groups of 68, 70, and 84 kDa are induced (4, 50). The 70-kDa class of HS proteins, arising from three genetic loci, represents a major proportion of total HS-protein synthesis in *D. melanogaster* and several other animal systems. In soybean, the high-molecular-weight HS proteins range from 68 to 110 kDa, and the small HS proteins are grouped between 15 and 27 kDa (22, 23, 26, 52).

The high-molecular-weight HS proteins seem to be highly conserved across a broad spectrum of organisms (21), whereas the small HS proteins show much more diversity (12, 13, 23, 46) in size and amino acid sequence. Sequence conservation between plants and animals among the genes encoding the high-molecular-weight HS proteins is evident by the cross-hybridization of plant genomic clones and poly(A)$^+$ RNAs with *Drosophila* HS cDNA clones (47; J. Roberts, unpublished data) and by cross-reactivity of antibodies (21). The high-molecular-weight HS proteins of plants, in contrast to those of *D. melanogaster* (14, 15, 29), represent a relatively small fraction of total HS-protein accumulation (e.g., soybean [22, 23, 25, 26]). The major HS-protein accumulation in plants is, instead, represented by a complex group of about 20 15- to 18-kDa proteins and approximately 10 20- to 27-kDa proteins based on radioactive amino acid incorporation and Coomassie-stained gel analyses (25). Although all plant species investigated to date synthesize a complex array of 15- to 27-kDa HS proteins, the electrophoretic patterns of these proteins vary among species (22).

Because of the distinct abundance and complexity of low-molecular-weight HS proteins in plants, our efforts have concentrated on the analysis of mRNA induction and the isolation and characterization of genes for this class of HS proteins. With HS-specific cDNA clones as hybridization probes, HS mRNAs corresponding to the small HS proteins are detectable within 3 to 5 min after HS (44). Liquid hybridization studies indicate that about 20 of these mRNAs accumulate to 20,000 copies each per cell within 2 h at 40°C (44). Considerable homology among this class of HS mRNAs is demonstrated by two-dimensional hybrid select and translation analyses with HS cDNA clones. Clone pCE53 or pFS2005 yields 13 proteins of 15 to 18 kDa, whereas pFS2019 yields a single 18-kDa protein within the same group (24, 44). Similar analyses with HS cDNA clone pFS2033 show three proteins of 21 to 24 kDa (23, 44), whereas pCE54 hybrid select translates into five 27-kDa proteins (12). These analyses and Southern hybridization analyses (23, 45) indicate the existence of several HS gene families within the low-molecular-weight group of proteins in soybean, with some families comprising only a few members and others up to 13 closely related proteins.

* Corresponding author.

As in other systems (4, 43), the HS genes of soybean are also induced to various degrees by such agents as arsenite, cadmium, and various environmental stresses (12). We see no evidence for the expression of soybean HS proteins in the 15- to 24-kDa groups under normal developmental or hormone-induced states. However, the 27-kDa group of HS proteins in soybean is expressed at control temperatures (28°C) and is induced 5- to 20-fold by elevated temperatures and a variety of other stresses (12). A similar type of enhanced expression has been observed for some of the *Drosophila* genes (36).

In this report and one by Czarnecka et al. (13), we present the DNA sequence analyses and transcript mapping results for three closely related genes within the 15- to 18-kDa group of HS proteins. A comparison of DNA sequences in the 5′ flanking region of the soybean genes with the corresponding regions of *Drosophila* HS genes reveals a striking similarity between plants and animals in the presumptive promoter structure of thermoinducible genes. The relationship of the low-molecular-weight HS proteins of soybean to the four small HS proteins of *D. melanogaster* is characterized by a comparison of the deduced amino acid sequences and by analysis of hydropathy plots.

MATERIALS AND METHODS

Restriction endonucleases, T4 DNA ligase, and DNA polymerase I large fragment were obtained from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), New England BioLabs, Inc. (Beverly, Mass.), and New England Nuclear Corp. (Boston, Mass.). Calf intestinal phosphatase was purchased from Boehringer Mannheim Biochemicals, (Indianapolis, Ind.). Polynucleotide kinase was obtained from Pharmacia, Inc. (Piscataway, N.J.). [α-^{32}P]dNTP and [γ-^{32}P]ATP were purchased from New England Nuclear. Chemicals used for DNA sequencing were from vendors recommended by Maxam and Gilbert (31). X-ray film, X-Omat AR-5, was supplied as long rolls by the Eastman Kodak Co. (Rochester, N.Y.). Acrylamide was purchased from Kodak and purified as described by Maniatis et al. (30) with the inclusion of a charcoal decolorization step (5 g/liter). All other chemicals were reagent grade unless otherwise stated.

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 λ_{1059} (20). The soybean λ_{1059} genomic library was constructed by J. Slightom and Y. Ma, Agrigenetics Advanced Research Laboratory, Madison, Wis. (10, 48). Screening was as described by Nagao et al. (34) with purified HS cDNA inserts labeled by nick translation (30) with ^{32}P. Lambda clones characterized in this study were designated GmhsλL (λL), GmhsλM (λM), and GmhsλE (λE). HS genes contained within these clones were subcloned into pUC9 (51) and designated *Gmhsp17.6-L* (*G. max* HS protein, 17.6 kDa), *Gmhsp17.5-M*, and *Gmhsp17.5-E* (corresponding plasmid clones are abbreviated pL, pM, and pE, respectively).

Subcloning. Southern hybridization with cDNA insert (pFS2005 [44]) was used to identify appropriate fragments from an *Eco*RI restriction digestion for subcloning (30) into pUC9 (host strain JM83) (51). An approximately 2.4-kilobase (kb) *Eco*RI fragment isolated from GmhsλM was subcloned and designated pM/EE2.4 (abbreviated pM). A 1.7-kilobase *Eco*RI insert subcloned from GmhsλL was designated pL/EE1.7 (abbreviated pL).

DNA sequence determination. The reactions for DNA sequence determination were those described by Maxam and Gilbert (31), except that formic acid was used for the A+G reaction. Protocols for sequencing procedures and the use of long (100-cm) sequencing gels were as described by Barker et al. (5). Computer analyses of DNA and protein sequences were performed with computer programs made available by J. Pustell and F. Kafatos (41) and modified for an HP 1000 (Hewlett-Packard Co., Palo Alto, Calif.) by M. Clegg and J. McClendon (University of Georgia, Athens).

Nuclease S1 hybrid protection mapping. A *Bam*HI site located within the coding region of all three HS genes served as a reference point for mapping the 5′ and 3′ termini of the HS transcripts according to the procedure of Favaloro et al. (17). The 5′ terminus of each gene was mapped with an appropriate DNA restriction fragment as hybridization probe that extended several hundred base pairs (bp) upstream from the 5′ end-labeled (31) *Bam*HI site and included the presumed cap site and TATA-like regions. The 3′ termini were mapped with DNA restriction fragments that extended downstream from the 3′ end-labeled (30) *Bam*HI site to include a few hundred bp distal to the deduced translational termination codon. Poly(A)$^+$ RNA (1 μg) isolated from control (28°C) and HS (40°C) soybean seedlings according to Czarnecka et al. (12) was hybridized (for 12 to 18 h) to end-labeled probe DNA at various temperatures ranging from 42 to 56°C to determine the optimum for hybrid formation. Nuclease S1 digestions were performed at 50 to 200 μ/ml for 30 min at 15°C. Protected hybrids were precipitated with isopropanol and detected by autoradiography after fractionation by electrophoresis on 6% polyacrylamide-urea sequencing gels.

RESULTS

Characterization of HS genomic clones. Soybean HS genomic clones were isolated from a λ_{1059} library by screening with HS cDNA inserts as described above. Approximately 40 independent clones were initially purified, each containing from 15 to 20 kilobases of genomic sequences. Our general strategy for mapping HS genes within large genomic inserts was first to identify the region of cDNA homology and start DNA sequencing in both directions from that point. Once restriction sites internal to HS mRNA were identified by their presence in the cDNA-homologous portion of the gene, nuclease S1 hybrid protection mapping was performed to position the 5′ and 3′ termini of the transcript. In several cases, the cDNA used originally to select a particular lambda clone did not show 100% homology to the genomic DNA. This apparent discrepancy is explained by cross-hybridization between the various members of closely related multigene families as was demonstrated previously in cDNA hybrid selection and in vitro translation studies (12, 23, 44). In all cases discussed here, the cDNA was colinear, with allowances for mismatch, to genomic sequences, indicating the absence of introns in these HS genes. The corresponding proteins for these genomic clones, based on cDNA hybrid select and translation, ranged in molecular weight from 15 to 27 kDa. The sequences of two selected representatives of the 15- to 18-kDa gene family are presented here.

Figure 1 shows a partial restriction map and strategies used for the determination of DNA sequences of genomic subclones pL and pM. The nucleotide sequence of each gene, along with its predicted amino acid sequence, is presented in Fig. 2. For comparative purposes, pE, the sequence of another HS gene in this size class, reported recently by Czarnecka et al. (13), will also be discussed. For

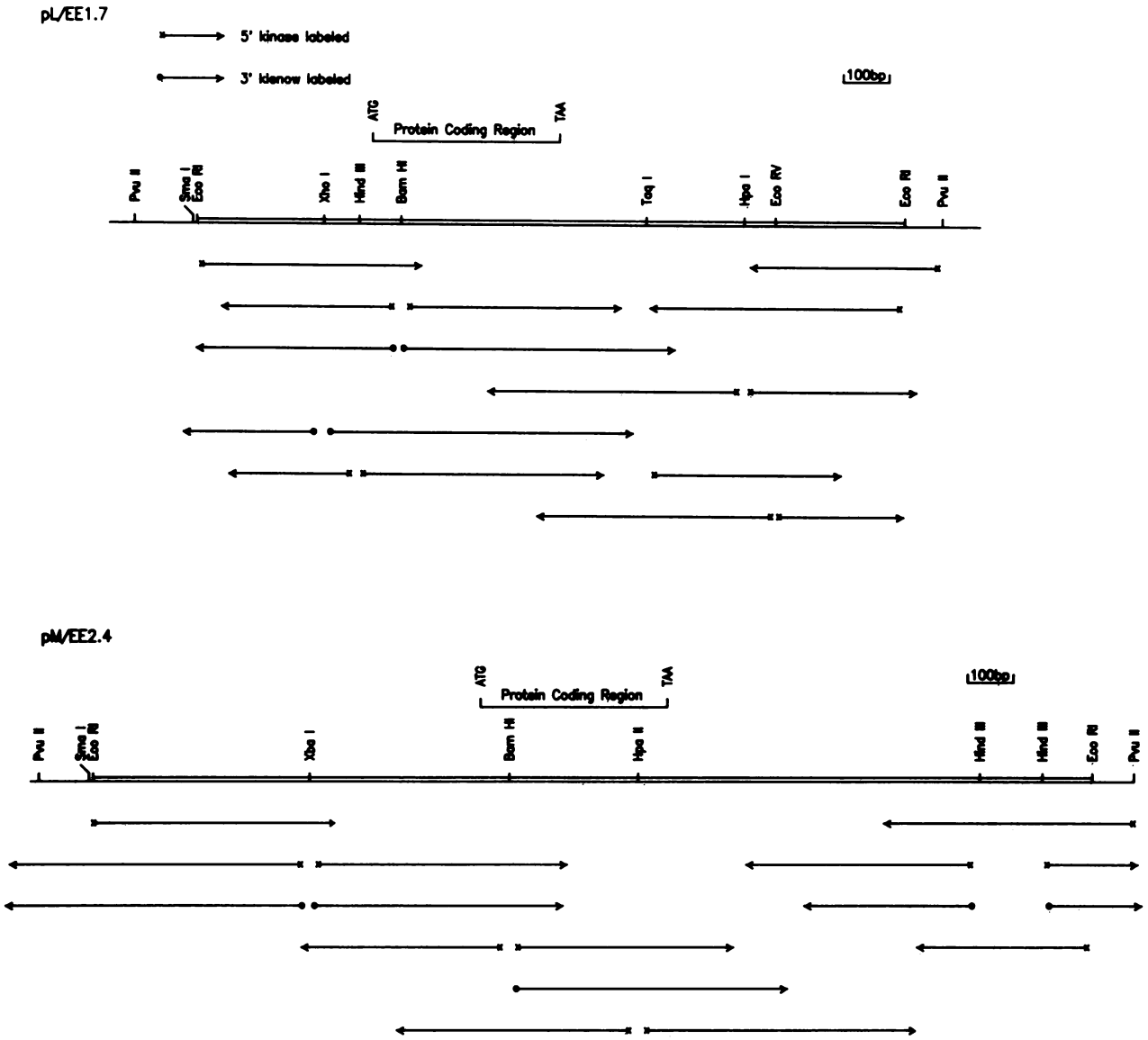

FIG. 1. Restriction map and sequencing strategy for *Eco*RI subclones pL/EE1.7 and pM/EE2.4 containing the HS genes *Gmhsp17.6-L* and *Gmhsp17.5-M*, respectively. Double line of restriction map indicates soybean genomic DNA; single line indicates plasmid vector pUC9. The strategies of restriction sites used for nucleotide sequence determination are shown below. The locations and orientations of protein coding regions are indicated above the restriction maps.

each of these genes there is close agreement between the molecular weight of the protein predicted from the deduced amino acid sequence and the molecular weights observed in hybrid selection and in vitro translation experiments (25, 44).

5′ Terminus of HS mRNAs. The 5′ termini of the HS mRNAs were positioned on the genomic sequences by nuclease S1 hybrid protection mapping with end-labeled DNA hybridization probes (17). Multiple 5′ termini were observed for both pL and pM transcripts (Fig. 3), whereas the closely related gene pE has a single 5′ terminus (13). For pL, there were two protected fragments of approximately equivalent intensity with lengths of 169 and 172 bp. S1 protection analysis for pM also yielded two fragments with lengths of 165 and 170 bp. The initiation sites for each of these genes are composed of short direct repeat sequences, both of which are used for initiation, except in the case of pE, in which only the TATA distal site is utilized. The redundant initiation sites for pL, pM, and pE are CATCATC, AAACGAAACG, and TCGTCCTCGTC, respectively, with two adenines occupying position +1 for pL and pM transcripts and a single guanine for the pE transcript. The 5′ leader sequences for these three gene transcripts are predicted to range from 82 to 96 bp. Some comparative features of the low-molecular-weight HS genes and proteins are presented in Table 1.

Analysis of the 3′ nontranslated region. DNA sequences of the 3′ nontranslated portion of the genes showed regions of extensive homology among the soybean genes based on homology matrix analyses (data not shown). The most notable regions of common homology reside around nucleotides 559 to 631, 651 to 692, and 721 to 730. Comparison of the 3′ end of soybean HS genes showed no consistent homology with the corresponding region of *Drosophila* HS genes.

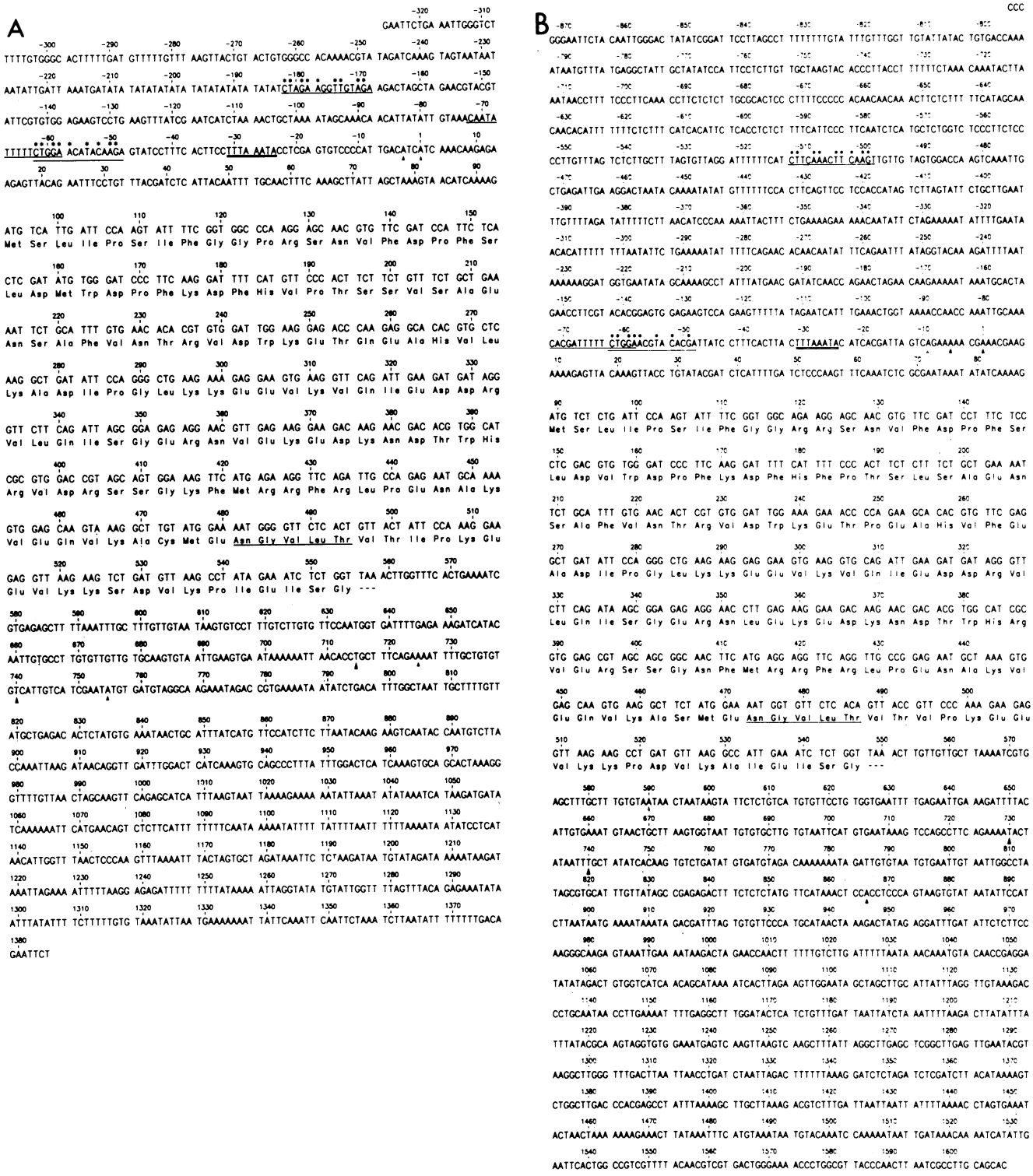

FIG. 2. The complete nucleotide sequences and deduced amino acid sequences of *Gmhsp17.6-L* (A) and *Gmhsp17.5-M* (B). For consistency, nucleotides are numbered from the distal start site for transcription, but both cap sites are indicated by arrows. The *Drosophila* HS consensus is underlined with asterisks denoting nucleotides homologous to the core inverted repeat at 90%. The TATA-like motif is bold underlined. Arrows denote termini of mRNA as determined by nuclease S1 hybrid protection analysis. The highly conserved amino acids found in the hydrophobic region of *Drosophila* and soybean low-molecular-weight HS proteins is underscored.

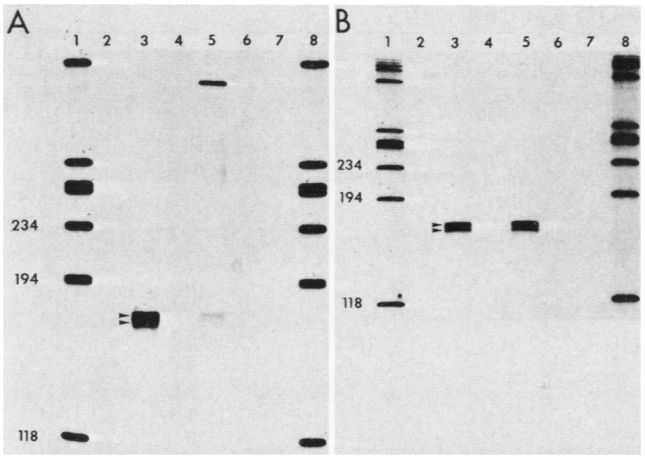

FIG. 3. Nuclease S1 mapping of 5′ termini of soybean HS genes *Gmhsp17.5-L* (A) and *Gmhsp17.5-M* (B). An internal *Bam*HI site was 5′ end labeled and used as a hybridization probe with 1 μg of poly(A)$^+$ RNA isolated from seedlings incubated at control (28°C) and HS (40°C) temperatures. Lanes: 2, 4, and 6, 28°C RNA; 3, 5, and 7, 40°C RNA. Temperatures of hybridization were: 42°C for lanes 2 and 3, 45°C for lanes 4 and 5, and 48°C for lanes 6 and 7. Bands (denoted by arrows) indicating 5′ termini are at 172 and 169 bases for *Gmhsp17.5-L* (A), and 170 and 165 bases for *Gmhsp17.5-M* (B). The high-molecular-weight (500-bp) band in lane 5 represents intact probe. Sizes (in bp) of φX174-*Hae*III DNA marker fragments are indicated in the margins.

The location of 3′ termini for transcripts of pL and pM were established with nuclease S1 hybrid protection studies. Multiple 3′ termini were observed for RNA homologous to both genes. Figure 4 demonstrates the utilization of two major (575 and 550 bases), two moderate (590 and 560 bases), and three minor (740, 710, and 660 bases) 3′ termini corresponding to transcripts from *Gmhsp17.6-L*. Mapping of the 3′ termini of transcripts homologous to *Gmhsp17.5-M* resulted in major bands at 580 and 570 bases, a moderately abundant band at 706 bases, and five to six minor bands from 780 to 520 bases. The presence of numerous minor bands is likely due, at least in part, to the complexity and interrelatedness of this multigene family, where some homology has been demonstrated in the 3′ nontranslated region. The possibility of hybridization bands arising from transcripts from different alleles of the same gene cannot be eliminated, but to reduce confusion arising from various degrees of cross-hybridization, we altered the criterion of RNA-DNA duplex formation by conducting hybridizations for nuclease S1 mapping at various temperatures. Autoradiographic bands that did not disappear or change in relative intensity with increasing stringency of hybridization were used to establish the location of presumptive 3′ termini or poly(A) addition sites. Thus, by this criterion, bands of 410, 395, and 380 for pL and 430, 420, 410, and 400 for pM are very likely the result of hybridizations with related but different members of this complex family of mRNAs. Many of the 3′ termini of pL and pM are located within a distance of 35 nucleotides downstream from a sequence similar to the mammalian consensus polyadenylation signal AATAAA (40) (for example, pL bands at 740, 710, and 660 bases; pM bands at 706 and 650 bases); however, many termini have no apparent correlation with the presence of a consensus-like polyadenylation sequence.

The 5′ flanking sequences. Once the 5′ terminus of the RNA was determined, the 5′ flanking sequences were searched for putative eucaryotic transcriptional regulatory elements. For the HS genes of *D. melanogaster*, the best-characterized promoter elements include an AT-rich motif known as the TATA box (Goldberg-Hogness box) and the HS consensus sequence CTgGAAtnTTCtAGA (38). A comparison of the 5′ flanking sequences of four soybean HS genes (including HS6871 [46]) revealed a striking degree of similarity in sequences between −23 and −72 bp, which include both a TATA-like motif and an upstream region with considerable homology to the *Drosophila* HS consensus. The TATA-like motif, TTTAAATA, was present in each of these soybean HS genes from 27 to 31 bp upstream from the transcription start sites and from 110 to 132 bp 5′ to the initiation codon for translation. Upstream, 31 and 41 bp from the 5′ end of the TTTAAATA motif, are the 5′ ends of two overlapping 15-bp sequences with homology to the *Drosophila* HS consensus promoter (38). Homology is 79% for both the overlapping consensus sequences of pE; 64 and 79% for the upstream and TATA-proximal overlapping consensus sequences, respectively, in pL; 71 and 64% respectively in pM; and 64 and 71%, respectively, in *HS6871* (46). The TATA-proximal HS consensus also demonstrated high homology to the HS core inverted-repeat consensus sequence identified in *D. melanogaster* (CTnGAAnnT TCnAG), with 90% homology for pE and pL and 80% for pM and *HS6871*. Secondary regions of high homology to the *Drosophila* HS consensus also occurred much further upstream from the TATA-proximal, or primary, HS element in each of the soybean genes. The *HS6871*, pL, and pM genes all had secondary HS elements with 90% homology to the core inverted repeat. These secondary consensus sequences were centered 86, 95, 106, and 137 bp upstream from the 5′ end of the TATA-like motif in *HS6871*; 145 bp upstream in *Gmhsp17.6-L*; and 472 bp from TATA in *Gmhsp17.5-M*. Gene *Gmhsp17.5-E* contained a single secondary consensus sequence 332 bp upstream from TATA with 80% homology to the *Drosophila* core inverted repeat.

Strong similarities among the soybean low-molecular-weight HS protein genes became more evident when approximately 300 nucleotides 5′ to the translation start codon of the soybean HS-protein genes were aligned (Fig. 5). The overall sequence homologies were: 68% between pM and pL, 74% between pE and pM, and 74% between pE and pL. A region of even higher homology (84 to 88%) lay between the 42 nucleotides from the 5′ end of the TTTAAATA motif through the 5′ end of the HS consensus sequence. The region of highest sequence heterogeneity was located between the 3′ end of the TTTAAATA motif and the CAP site. Over this 24-nucleotide sequence, the homologies between pE and pM, pL and pM, and pL and pE were only 29, 42, and 67%, respectively.

Alternating purine-pyrimidine stretches have the potential under certain conditions to form Z-DNA (35). Several regions of alternating purines and pyrimidines were present in the 5′ flanking sequences of genes *Gmhsp17.6-L* and *Gmhsp17.5-E*. In *Gmhsp17.6-L*, a group of 15 alternating pairs of AT occurred adjacent to a secondary HS consensus centered at position −198. In *Gmhsp17.5-E*, short clusters were centered at positions −130 and −106 immediately upstream from a sequence showing 78% (11 of 14) homology to the simian virus 40 enhancer core.

Analysis of deduced amino acid sequences. The soybean gene sequences presented here each contained an uninterrupted open reading frame starting at the first ATG 3′ to the TTTAAATA motif. The molecular weights deduced from the single, uninterrupted open reading frames of these genes

TABLE 1. Some features of small HS genes and proteins

Organism and gene or protein	Sequences: TATA box-initiator codon distance (bp)	Sequences: Estimated leader sequence length (bp)	Sequences: % A+T content of first 200-bp extragenic sequence	% Base composition of leader sequences (A+T content)	Protein-coding sequence: Length of open reading frame (bp)	Protein-coding sequence: Length of encoded polypeptide (residues)	Protein-coding sequence: Mol wt of unmodified polypeptide chain	Termination codon
D. melanogaster[a]								
hsp22	284	ND[b]	57	70	525	174	19,705	TAG
hsp23	146	119 ± 3	47	67	561	186	20,630	TAG
hsp26	215	178 ± 3	66	67	627	208	22,997	TAA
hsp27	151	ND	61	69	642	213	23,620	TAA
Soybean[c]								
Gmhsp17.5-M	110	88 + 93	68[d]	68	459	153	17,544	TAA
Gmhsp17.6-L	123	93 + 96	69[e]	70	462	154	17,570	TAA
Gmhsp17.5-E	113	82	62	65	462	154	17,533	TGA
HS6871	132	104	64	65	459	153	17,345	TAA

were 17.6, 17.5, 17.5, and 17.3 kDa for genes *Gmhsp17.6-L*, *Gmhsp17.5-M*, *Gmhsp17.5-E*, and *HS6871*, respectively. The deduced molecular weights of these four genes are consistent with experimental determinations of molecular weights of 15 to 18 kDa based on in vivo-labeled HS proteins, in vitro translation of HS poly(A) RNA, and HS cDNA hybrid select and translated proteins (26, 44). This agreement between the deduced and observed molecular weights of these proteins, the comparison of partial-length cDNA sequences with genomic DNA sequences, and the transcript mapping studies indicate that these four HS genes do not contain introns.

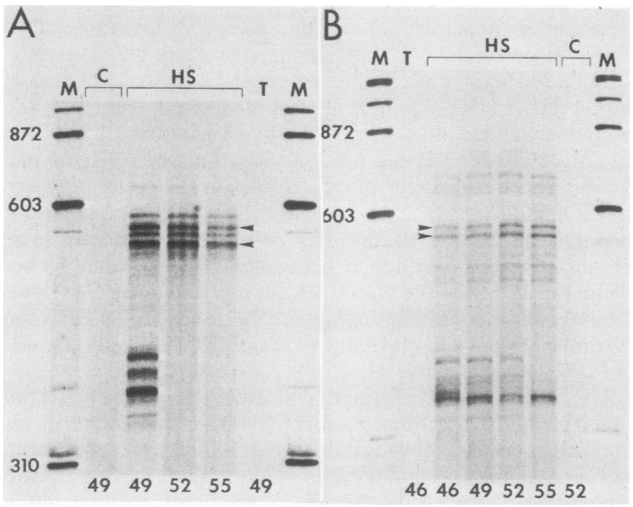

FIG. 4. Nuclease S1 mapping of the 3′ termini of soybean HS genes *Gmhsp17.5-L* (A) and *Gmhsp17.5-M* (B). An internal *Bam*HI site was 3′ end labeled and used as a hybridization probe with 1 μg of poly(A)$^+$ RNA isolated from seedlings incubated at control (28°C) and HS (40°C) temperatures. RNA samples are designated above the lanes as control (C), heat shock (HS), and yeast tRNA (T). Temperatures (in °C) of hybridization are indicated below the lanes. Major 3′ termini are indicated by bands (denoted by arrows) at 575 and 550 bases for *Gmhsp17.5-L* (A) and at 580 and 570 bases for *Gmhsp17.5-M* (B). Sizes (in bp) of ϕX174-*Hae*III DNA marker (M) fragments are indicated in the margins.

Amino acid sequence comparisons of the four soybean HS proteins revealed homologies of greater than 90% (Fig. 6). Of the nucleotide changes observed in the open reading frame, approximately two-thirds were silent substitutions. For example, of the 42 nucleotide differences between *Gmhsp17.5-M* and *Gmhsp17.6-L*, 28 were synonymous substitutions with 14 nucleotide changes leading to amino acid changes. Thirty-nine nucleotide changes were observed when genes *Gmhsp17.5-M* and *Gmhsp17.5-E* were compared; of these, 26 are synonymous substitutions, and 13 lead to amino acid changes. A comparison of *Gmhsp17.5-E* with *Gmhsp17.6-L* showed 31 nucleotide changes, of which 16 are synonymous substitutions, and 15 cause amino acid changes. Comparison of *Gmhsp17.5-E* with *HS6871* showed 34 synonymous substitutions out of 47 total changes. In addition, a single amino acid deletion is predicted in the sequences of *Gmhsp17.5-M* and *HS6871*. Tyrosine is not encoded in the proteins of *Gmhsp17.5-E* and *HS6861*, whereas cysteine is only encoded in the *Gmhsp17.6-L* sequence.

A comparison of the deduced amino acid sequences of six proteins in the small HS family from *C. elegans* (42), *D. melanogaster* (18), and *Xenopus laevis* (7) with the four low-molecular-weight HS proteins of soybean is presented in Fig. 7. Only the carboxy-terminal half of the proteins is shown, since a significant level of homology was not seen in the amino-terminal regions. Spaces represent adjustments in alignment to maximize homology. There were five positions where the same amino acid is used in all 10 of the proteins examined. Of 92 positions, 15 (16%) in the carboxy-terminal portion of the soybean genes had the identical amino acid found in at least three other proteins. Of 92 soybean positions, 38 (41%) were identical, with at least one protein from the other organisms. Separate comparisons of the composite amino acid sequences of *C. elegans* and *X. laevis* over the region presented with the composite of the four proteins of *D. melanogaster* show 39 and 37% homology, respectively. A similar comparison of the composite soybean amino acid sequences with *D. melanogaster* reveals a somewhat lower homology of 28%. Maximum alignment was obtained by assuming that a deletion occurred near the carboxy terminus in the soybean proteins. These results demonstrate significant homology in primary amino acid sequence between the

TABLE 1—*Continued*

Acidic amino acids (%)	Basic amino acids(% Arg + Lys)	Basic amino acids (% Arg + Lys + His)	Base composition (%) A	G	C	T	G+C	Base composition (in 100 bp) 3′ to the termination codon A	G	C	T	A+T
16.7	12.1	14.4	22	32	28	17	58	41	13	13	33	74
14.0	11.3	14.0	22	31	27	20	60	32	15	19	34	66
12.5	12.5	16.8	22	29	30	19	59	30	9	21	38	68
13.1	12.7	18.8	22	31	29	18	60	43	22	17	16	59
18.3	15.1	17.1	27	29	19	25	48	34	17	14	34	68
17.4	15.5	17.4	30	28	17	25	45	36	15	12	38	74
17.4	16.1	18.0	27	28	19	26	47	36	16	15	33	69
17.0	16.4	17.8	28	29	18	25	47	31	18	13	38	69

[a] *Drosophila* data from Southgate et al. (49).
[b] ND, Not determined.
[c] Soybean data for *Gmhsp17.5-E* and *HS6871* from Czarnecka et al. (13) and Schöffl et al. (46), respectively.
[d] M clone 600 bp of 5′ = 70% A+T; M clone 957 bp of 5′ = 69% A+T.
[e] L clone 417 bp of 5′ = 70% A+T.

low-molecular-weight family of HS proteins in soybean and the class of proteins present in animals known as the small HS proteins (18).

Hydropathy plots provide a means to evaluate common structural features of distantly related proteins (27). Hydropathy profiles of the four soybean proteins were nearly identical over the entire length because of the high degree of amino acid homology found among members of this closely related family of HS proteins. A comparison of a representative (*Gmhsp17.6-L*) of the soybean HS proteins with small HS proteins from other organisms gave a more complete view of the similarity in structural features among members of this diverse class of proteins. The hydropathy profiles shown in Fig. 8 present representative proteins from

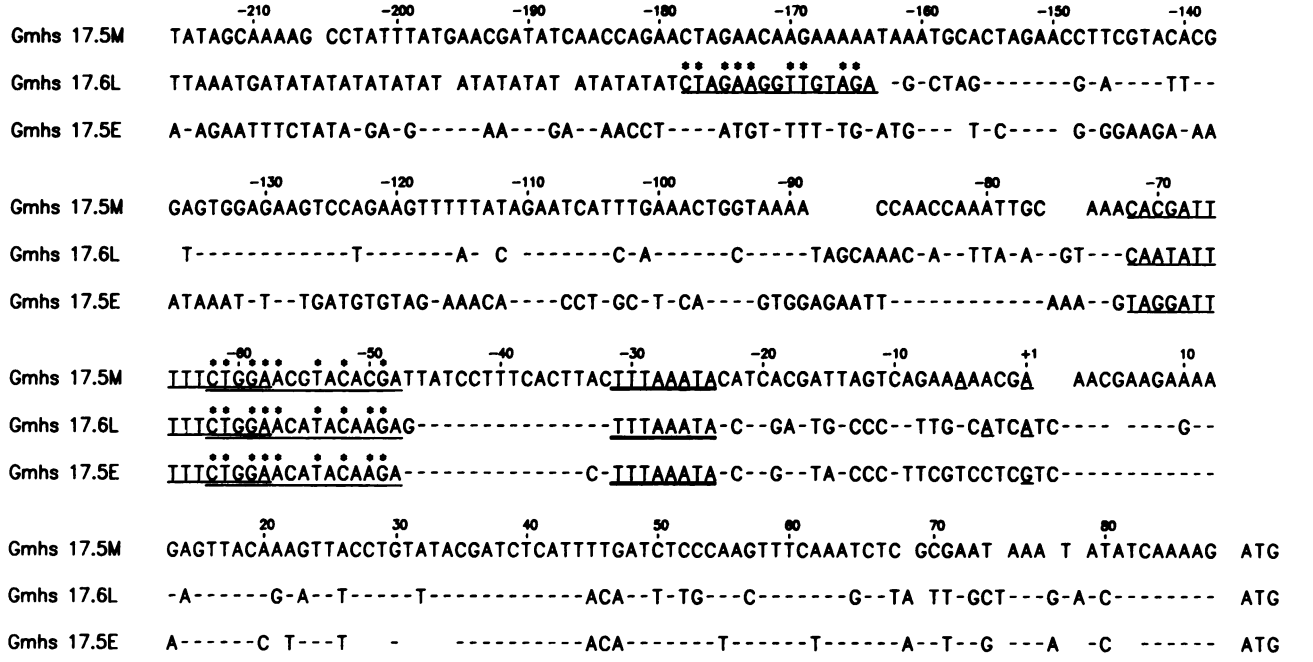

FIG. 5. Comparison of 5′ flanking region of soybean HS genes. The nucleotide sequences of *Gmhs17.5-M*, *Gmhs17.6-L*, and *Gmhs17.5-E* (13) are aligned with *Gmhs17.5-M* as a reference. Numbering is from the transcription start site distal to the TATA motif (heavy line) of *Gmhs17.5-M*. Broken lines indicate identical bases; spaces are included for maximum alignment. The HS consensus sequences are designated by light underlines. Asterisks denote nucleotide homology to HS core inverted repeat at a minimum of 80%.

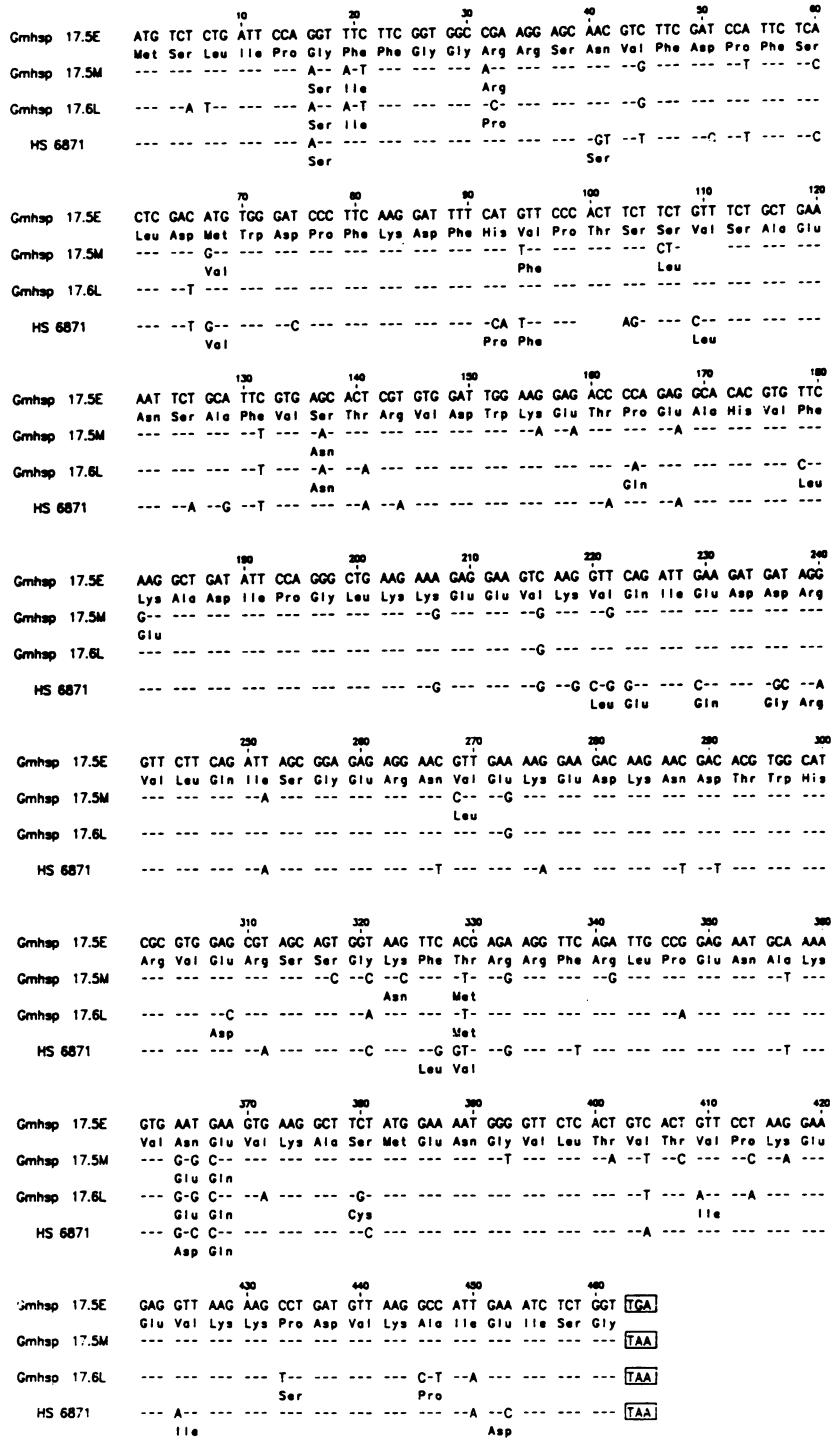

FIG. 6. Nucleotide sequence and deduced amino acid sequence comparisons of four soybean low-molecular-weight HS proteins. The sequence of *Gmhsp17.5-E* is used as reference. Dashed lines indicate identical nucleotides, whereas nucleotide changes are indicated by appropriate letters. Amino acid differences are listed below corresponding codons. Blank spaces in *Gmhsp17.5-M* and *HS6871* are predicted deletions which are added to maintain maximum homology. Sequence *Gmhsp17.5-E* is from Czarnecka et al. (13), and *HS6871* is from Schöffl et al. (46).

a diverse group of small HS proteins. The soybean and *Drosophila* (49) plots were constructed from the deduced sequences of completely sequenced genomic clones, whereas the *Xenopus* (7) and *C. elegans* (42) plots are from published amino acid sequences derived from the DNA sequences of partial-length cDNA clones. The most prominent common feature of these profiles is the major hydrophilic peak centered around amino acid residue 95, which is flanked by a carboxy-proximal hydrophobic region. Although the hydropathic nature of this region seems to be

FIG. 7. Deduced amino acid sequences from the carboxy halves of 10 HS proteins. The sequences have been aligned to maximize homology with the four *Drosophila* small HS proteins (18). Numbers under the sequences correspond to the amino acid position in the composite soybean protein. The organisms of origin and molecular weights of the proteins are given at the carboxy termini. The sequences of *C. elegans* are a composite of two cDNAs (42), and the *Xenopus* sequences represent cDNAs X4 and X5 (7). Boxes indicated homology with the composite amino acid sequence of four soybean HS genes.

important, the actual amino acids used in this region showed considerable diversity among the organisms studied. This type of analysis also suggests a closer structural relationship between soybean and *Drosophila* proteins than is evident by a simple analysis of amino acid homology and suggests that the soybean proteins have more in common with *Drosophila* small HS proteins than with either *X. laevis* or *C. elegans*.

DISCUSSION

HS genes as multigene families. The relatively high abundance and distinctly lower molecular weight of the more than 20 15- to 18-kDa HS proteins of soybean initially suggested that perhaps these HS proteins could be unique to plants and perhaps regulated in a manner different from those of animals. Analysis of data presented here, however, together with other work (12, 23, 44, 45), demonstrates that the low-molecular-weight HS proteins of soybean represent multigene families analogous to the small HS groups first characterized in *D. melanogaster* (18). The largest soybean family, consisting of thirteen proteins, is partially described here and is represented by genomic clones pE2019, pL2005, pM2005, and pHS6871 and cDNA clones pFS2019, pFS2005, pFS1968, and pCE53 (13, 44–46). Additional families of small HS proteins include cDNA clone pCE75 (15-kDa class), pFS2033 (21- to 24-kDa class), and possibly pCE54 (27-kDa class) (13, 23, 44, 46). Other groups of small HS proteins are known to exist, including HS proteins which are transported into the chloroplasts (E. Vierling, unpublished data).

Coding region analysis. A comparison of the predicted amino acid sequences reveals approximately 90% homology among the soybean HS proteins analyzed to date. A majority of the nucleotide changes represent silent substitutions. A comparison of the base changes with regard to composition and location suggests a duplication mechanism for the origin of these HS genes. Likewise, the close linkage of the four low-molecular-weight HS genes in *D. melanogaster* at locus 67B suggests that these have a common origin and arose by gene duplication and inversion events (49). The very high homology between the coding regions of the low-molecular-weight HS genes of soybean, as well as multiple-cross homologies in the 5′ flanking regions, also suggests a common origin by duplication. However, regardless of origin, the maintenance of these multiple homologies may represent a regulatory refinement for the assurance of high expression of these genes under specific physiological conditions, e.g., HS.

Comparison with the *Drosophila* small HS proteins based simply on linear amino acid composition shows low but significant homology to the soybean HS proteins. However, a more general structural relationship is revealed by hydropathy analyses (27). The major hydrophilic peak, which is characteristic of the small HS proteins of *D. melanogaster* (49), is clearly present in the soybean proteins

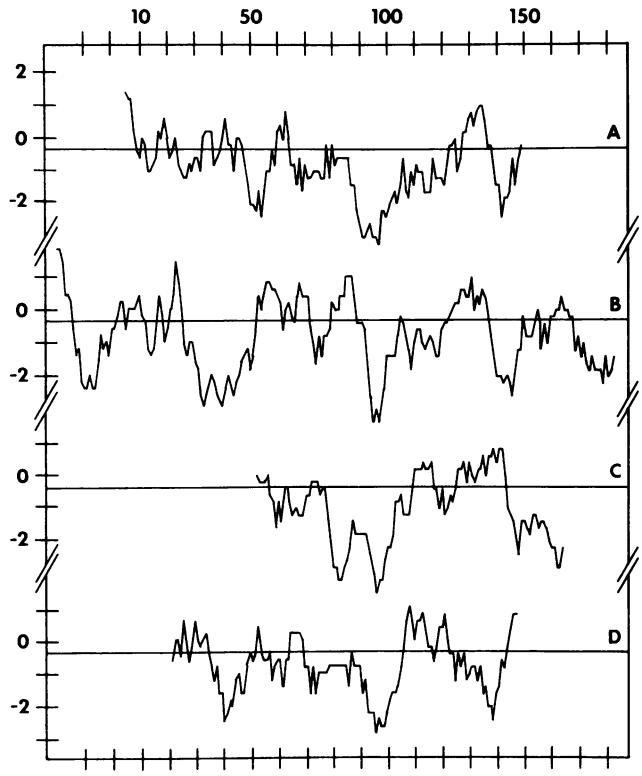

FIG. 8. Hydropathic profiles of deduced amino acid sequences of (A) soybean *Gmhsp17.6-L*, (B) *Drosophila hsp27* (49), (C) *Xenopus hsp30* (7), and (D) *C. elegans hsp16* (42). Plots were constructed by the method of Kyte and Doolittle (27) by progressively moving along the amino acid sequence and averaging the hydropathy index for nine amino acids. Points above the horizontal line correspond to hydrophobic regions, and points below this line are hydrophilic. The plots are aligned along the major hydrophilic peak which is characteristic of the small HS protein.

from position 95 to 105 (Fig. 8). A total of four common domains of similarity have been identified (13). A striking conservation of Asp-Gly-Val-Leu-Thr occurs in a very hydrophobic domain (positions 125 to 140; Fig. 8) in all the *Drosophila* small HS proteins, α-crystallin, and the soybean genes, except that in soybean an Asn replaces Asp. The overall similarities in the hydropathy profiles and a significant level of amino acid homology are suggestive of the conservation of functional domains among this broad group of proteins from highly diverged groups of organisms.

The selective localization of low-molecular-weight HS proteins in soybean to organellar fractions (e.g., nuclei, mitochondria, and ribosomes) during HS is likely the basis of thermoprotection (22). Selective localization is also known for some low-molecular-weight HS proteins of *Drosophila* (2, 3). It has been suggested (28, 46, 49) that the similarities in the protein domains derived from hydropathy analysis between the small HS proteins and the bovine α-crystallin lens protein may be related to common aggregation properties and possibly other types of protein-protein interactions involved in function and localization.

HS promoter analysis. In addition to a TATA-like motif characteristic of many eucaryotic genes, soybean HS genes also show a great deal of similarity in the DNA sequence composition of the 5′ flanking region of HS genes of *D. melanogaster* and *X. laevis*. The best-characterized sequence element is the HS consensus core of *D. melanogaster*, CT-GAA--TTC-AG (38). In heterologous systems, in high copy number, this sequence has been shown to be sufficient for thermal activation of transcription when placed 13 or 19 nucleotides upstream to the TATA-box (39). Although the *Drosophila* HS consensus functions inductively in heterospecific systems, the temperature optimum for expression corresponds to the recipient cells (9, 11, 32, 33). In most cases, an 8 of 10 match with the HS consensus core is required for thermal induction. Exceptions to this rule suggest that a 6 of 10 match can function when additional copies of the consensus are present (1, 8, 39). In all four of the soybean genes discussed here, the TATA-proximal HS consensus element overlaps with a second HS consensus sequence with lower homology. Additional HS consensus elements with 8 of 10 homology match are identified further upstream in the soybean genes. These redundant elements are located at positions −169 to −182 nucleotides in pL; −499 to −512 nucleotides in pM; −358 to −371 nucleotides in pE; and −212 to −225, −221 to −234, −232 to −245, and −263 to −276 nucleotides in *HS6871*. This finding is not unexpected, since most HS genes in other organisms also contain multiple copies of the HS consensus core sequences located within 250 nucleotides of the start of transcription (8).

Two protein-binding sites are implicated in the activation of HS genes (53), and a specific activating protein has been described that binds the HS consensus region of *hsp82* gene chromatin in vitro (54). A specific HS transcription factor and a general transcription factor (the A factor) are required for active in vitro transcription of the *hsp70* gene by RNA polymerase II (37). The presence of redundant HS consensus sequences in soybean HS genes suggests the possibility of cooperative binding of multiple HS transcription factor proteins either to effect high levels of transcription or to modulate the level under somewhat different physiological states relative to stress or both. It has been proposed (8, 16) that redundant HS consensus core elements are required for efficient thermal induction when the gene is in low copy number and must compete with other HS genes for transcriptional factors or when a suboptimal promoter configuration exists because of poor spacing or homology.

A decameric palindromic sequence in *D. melanogaster* is found 5′ distal to the HS consensus element in similar positions in *hsp26* and *hsp70* and in slightly modified form in *hsp22* (49). It is intriguing to note that this decameric palidromic sequence is partially homologous to the HS core inverted repeat (AGAAATTTCT; CTnGAAnnTTCnAG). An analysis for homology to this decameric palindromic sequence showed 70 to 80% homology in the soybean HS genes in numerous locations. Interestingly, in the pL clone, 70% homology to this decameric palindromic sequence is located four times in the sequence 5′ to the TTTAAATA motif. In each case, this sequence is within or partially within a sequence of at least 60% homology to an HS consensus sequence (centered at −35, −94, −144, and −154 bp from the 5′ end of the TTTAAATA motif). The 5′ end of the sequence centered at −94 is 28 bp upstream from a second potential TATA-like sequence of TAAAATA. A similar analysis of the pM clone 5′ sequence shows 80% homology with the *Drosophila* decameric palindrome at −88, −329, −350, and −621 bp from the TTTAAATA motif. There are numerous matches of 70% homology, but none of these are located within or adjacent to an HS consensus sequence. This element is also located in the upstream region of *Gmhsp17.5-E* (−186 and −354) and *HS6871* (−35, −87, −138, and −310). The element at −310 in *HS6871* is 40

bp upstream from a TTTAAATA motif. Although this decameric palindromic sequence has not been shown to be functionally significant in the thermal activation of transcription in animal systems, the presence of a similar sequence in soybean HS genes suggests evolutionary conservation of this element and therefore a possible role in modulating HS gene expression.

In *D. melanogaster*, numerous other stress treatments are effective in inducing HS proteins (4). Most of these other stresses are not effective in soybean; however, arsenite and, to a lesser extent, cadmium induce poly(A^+) RNAs homologous to some HS cDNAs (12). In many organisms, heavy metals, such as cadmium, induce the synthesis of metal-binding proteins, the metallothioneins, for which specific regulatory elements have been identified in the 5′ flanking regions of these genes (19). Sequence homology searches for these elements located four regions of 75% homology, for example, in the 5′ flanking sequence of pM. Several homologies of greater than 70% were located within other soybean HS genes. Since some HS genes in soybean are induced by heavy metals (12), these homologies are suggestive of possible functional significance.

In the soybean HS genes, slight homology to the canonical sequence 5′-GG$^{C}_{T}$CAATCT′-3′ or CCAAT box (6) is centered 53 and 50 bp upstream from the 5′ end of the TTTAAATA motif of pE and pM. Higher homology is located 145 bp upstream from the TTTAAATA motif in *HS6871* (46), but no corresponding homology is found in pL. The CCAAT box homology present in some of the soybean genes may only be coincidental, since the *Drosophila* low-molecular-weight HS-protein genes show no obvious homology to the CCAAT box sequence (49).

DNA sequence analysis of the low-molecular-weight HS genes of soybean supports the view that the molecular mechanisms involved in the thermal induction of HS genes are highly conserved among eucaryotes. In addition to the HS consensus core, other DNA sequences, such as the metal response element, simian virus 40 enhancer core, and potential Z-DNA stretches have been identified in the 5′ flanking regions. The presence of these homologies suggests a general conservation in DNA sequences that control transcription in eucaryotes and raises the possibility that these soybean HS genes may be subject to a variety of controls in addition to HS. The continued structural analyses of HS genes and HS proteins coupled with in vitro mutagenesis transformation or expression experiments should provide a better basis for understanding the HS genes of plants.

ACKNOWLEDGMENTS

We thank Jerry Slightom and Yu Ma for providing the λ_{1059} genomic library. We acknowledge the technical assistance of Kenlock Westberry III. We thank Burlyn Michel for writing and modifying computer plotting programs and Joyce Kochert for graphics assistance. We thank Janice Kimpel, Elizabeth Vierling, and John Walker for their helpful comments in preparing the manuscript.

This work was supported by a research contract from Agrigenetics Research Associates, Limited.

LITERATURE CITED

1. **Ayme, A., R. Southgate, and A. Tissieres.** 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.
2. **Arrigo, A. P., and C. Ahmad-Zadeh.** 1981. Immunofluorescence localization of a small heat shock protein (hsp23) in salivary gland cells of *Drosophila melanogaster*. Mol. Gen. Genet. **184**:73–79.
3. **Arrigo, A. P., S. Fakan, and A. Tissieres.** 1980. Localization of the heat shock-induced proteins in *Drosophila melanogaster* tissue culture cells. Dev. Biol. **78**:86–103.
4. **Ashburner, M., and J. J. Bonner.** 1979. The induction of gene activity in *Drosophila* by heat shock. Cell **17**:241–254.
5. **Barker, R. F., K. B. Idler, D. V. Thompson, and J. D. Kemp.** 1983. Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine Ti plasmid pTi 15955. Plant Mol. Biol. **2**:335–350.
6. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
7. **Bienz, M.** 1984. Developmental control of the heat shock response in *Xenopus*. Proc. Natl. Acad. Sci. USA **81**:3138–3142.
8. **Bienz, M.** 1985. Transient and developmental activation of heat-shock genes. Trends Biochem. Sci. **10**:157–161.
9. **Bienz, M., and H. R. B. Pelham.** 1982. Expression of a *Drosophila* heat-shock protein in *Xenopus* oocytes: conserved and divergent regulatory signals. EMBO J. **1**:1583–1588.
10. **Blattner, F. R., A. Blechl, K. Denniston-Thompson, H. E. Faber, J. E. Richards, J. E. Slightom, P. W. Tucker, and O. Smithies.** 1978. Cloning human fetal γ globin and mouse α-type globin DNA: preparation and screening of shotgun collections. Science **202**:1279–1284.
11. **Corces, V., A. Pellicer, R. Axel, and M. Meselson.** 1981. Integration, transcription and control of a *Drosophila* heat shock gene in mouse cells. Proc. Natl. Acad. Sci. USA **78**:7038–7042.
12. **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.
13. **Czarnecka, E., W. B. Gurley, R. T. Nagao, L. 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.
14. **DiDomenico, B. J., G. E. Bugaisky, and S. Lindquist.** 1982. Heat shock and recovery are mediated by different translational mechanisms. Proc. Natl. Acad. Sci. USA **79**:6181–6185.
15. **DiDomenico, B. J., G. E. Bugaisky, and S. Lindquist.** 1982. The heat shock response is self-regulated at both the transcriptional and posttranscriptional levels. Cell **31**:593–603.
16. **Dudler, R., and A. A. Travers.** 1984. Upstream elements necessary for optimal function of the hsp70 promoter in transformed flies. Cell **38**:391–398.
17. **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.
18. **Ingolia, T. D., and E. A. Craig.** 1982. Four small *Drosophila* heat shock proteins are related to each other and to mammalian α-crystallin. Proc. Natl. Acad. Sci. USA **79**:2360–2364.
19. **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 glucorticoid hormones induce human metallothionein-II_A gene. Nature (London) **308**:513–519.
20. **Karn, J., S. Brenner, L. Barnett, and G. Cesareni.** 1980. Novel bacteriophage λ cloning vector. Proc. Natl. Acad. Sci. USA **77**:5172–5176.
21. **Kelley, P. M., and M. J. Schlesinger.** 1982. Antibodies to two major chicken heat shock proteins cross-react with similar proteins in widely divergent species. Mol. Cell. Biol. **2**:267–274.
22. **Key, J. L., E. Czarnecka, C. Y. Lin, J. Kimpel, C. Mothershed, and F. Schöffl.** 1983. A comparative analysis of the heat shock response in crop plants, p. 107–118. *In* D. D. Randall, D. G. Blevins, R. L. Larson, and B. J. Rapp. (ed.), Current topics in plant biochemistry and physiology, vol. 2. University of Missouri, Columbia, Mo.
23. **Key, J. L., W. B. Gurley, R. T. Nagao, E. Czarnecka, and M. A. Mansfield.** 1985. Multigene families of soybean heat shock proteins. *In* L. van Vloten-Doting, G. Groot, and T. Hall (ed.),

NATO Advanced Studies Institute molecular form and function of the plant genome. Plenum Publishing Corp., New York.
24. **Key, J. L., J. A. Kimpel, C. Y. Lin, R. T. Nagao, E. Vierling, E. Czarnecka, W. B. Gurley, J. K. Roberts, M. A. Mansfield, and L. Edelman.** 1985. The heat shock response in soybean, p. 161–179. *In* J. L. Key and T. Kosuge (ed.), Cellular and molecular biology of plant stress, vol. 22. Alan R. Liss, Inc., New York.
25. **Key, J. L., J. A. Kimpel, E. Vierling, C. Y. Lin, R. T. Nagao, E. Czarnecka, and F. Schöffl.** 1985. Physiological and molecular analyses of the heat shock response in plants, p. 327–348. *In* B. G. Atkinson and D. B. Walden (ed.), Changes in eukaryotic gene expression in response to environmental stress. Academic Press, Inc., New York.
26. **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.
27. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
28. **Lin, C. Y., J. R. Roberts, and J. L. Key.** 1984. Acquisition of thermotolerance in soybean seedlings. Plant Physiol. **74**:152–160.
29. **Linquist, S.** 1980. Varying patterns of protein synthesis in *Drosophila* during heat shock: implications for regulation. Dev. Biol. **77**:463–479.
30. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:449–560.
32. **McMahon, A. P., T. J. Novak, R. J. Britten, and E. H. Davidson.** 1984. Inducible expression of a cloned heat shock fusion gene in sea urchin embryos. Proc. Natl. Acad. Sci. USA **81**:7490–7494.
33. **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.
34. **Nagao, R. T., D. M. Shah, V. K. Eckenrode, and R. B. Meagher.** 1981. Multigene family of actin-related sequences isolated from a soybean genomic library. DNA **1**:1–9.
35. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303**:674–678.
36. **O'Connor, D., and J. T. Lis.** 1981. Two closely linked transcription units within the 65B heat shock puff locus of *D. melanogaster* display strikingly different regulation. Nucleic Acids Res. **9**:5075–5092.
37. **Parker, C. S., and J. Topol.** 1984. A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. Cell **37**:273–283.
38. **Pelham, H. R. B.** 1982. A regulatory upstream promoter element in the *Drosophila* hsp70 heat-shock gene. Cell **30**:517–528.
39. **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.
40. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
41. **Pustell, J., and F. C. Kafatos.** 1982. A convenient and adaptable package of DNA sequence analysis programs. Nucleic Acids Res. **10**:51–59.
42. **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.
43. **Schlesinger, M. J., M. Ashburner, and A. Tissieres.** 1982. Heat shock: from bacteria to man. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. **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.
45. **Schöffl, F., and J. L. Key.** 1983. Identification of a multigene family for small heat shock proteins in soybean and physical characterization of one individual gene coding region. Plant Mol. Biol. **2**:269–278.
46. **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.
47. **Shah, D. M., D. E. Rochester, G. G. Krivi, C. M. Hironaka, T. J. Mozer, R. T. Fraley, and D. C. Tiemeier.** 1985. Structure and expression of maize hsp 70 gene, p. 181–200. *In* J. L. Key and T. Kosuge (ed.), Cellular and molecular biology of plant stress, vol. 22. Alan R. Liss, Inc., New York.
48. **Slightom, J. L., A. E. Blechl, and O. Smithies.** 1980. Human fetal G_γ- and A_γ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.
49. **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.
50. **Southgate, R., M.-E. Mirault, A. Ayme, and A. Tissieres.** 1985. Organization, sequences, and induction of heat shock genes, p. 3–30. *In* B. G. Atkinson and D. B. Walden (ed.), Changes in eukaryotic gene expression in response to environmental stress. Academic Press, Inc., Orlando, Fla.
51. **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.
52. **Vierling, E., and J. L. Key.** 1985. Ribulose 1,5-bisphosphate carboxylase synthesis during heat shock. Plant Physiol. **78**:155–162.
53. **Wu, C.** 1984. Two protein-binding sites in chromation implicated in the activation of heat-shock genes. Nature (London) **309**:229–234.
54. **Wu, C.** 1984. Activating protein factors binds *in vitro* to upstream control sequences in heat-shock gene chromation. Nature (London) **311**:81–84.