MOLECULAR AND CELLULAR BIOLOGY, Aug. 1989, p. 3457–3463
0270-7306/89/083457-07$02.00/0

Vol. 9, No. 8

Regulatory Domains of the *Gmhsp17.5-E* Heat Shock Promoter of Soybean†

EVA CZARNECKA,* JOE L. KEY,‡ AND WILLIAM B. GURLEY

Department of Microbiology and Cell Science, 1059 McCarty Hall, University of Florida, Gainesville, Florida 32611

Received 30 December 1988/Accepted 9 May 1989

Promoter domains required for in vivo transcriptional expression of soybean heat shock gene *Gmhsp17.5-E* were identified by insertion-deletion mutagenesis with transgenic expression monitored in *Agrobacterium tumefaciens*-incited tumors of sunflower. Removal of the TATA-distal domain from position −1175 to position −259 had little effect on overall activity. The four regions contributing to promoter activity identified by this study all map within 244 base pairs from the start of transcription. The most distal *cis*-acting element of major significance was located from −244 to −179 and contains a conserved TATA-dyad motif centered at −220. Sequences from −179 to −40 comprise the TATA-proximal domain and include an AT-rich region and two sites containing heat shock consensus elements (HSEs). Deletion of the HSE centered at −93 (site 2) severely reduced transcriptional activity. Heat-inducible expression was also eliminated by internal deletion of either the TATA motif or the overlapping HSEs at site 1, indicating that each of these regions is also a major determinant of promoter activity.

The strong correlation between the appearance of heat shock (HS)-specific RNAs and the rapid increase of HS proteins (HSPs) suggests that the expression of HS genes is regulated predominantly at the transcriptional level. In soybean, HS RNA is barely detectable in non-heat-shocked plants but rapidly increases to 15,000 to 20,000 copies per cell after 2 h of treatment at 40°C (30). Sequence analysis of the 5′ flanking region of HS genes reveals a strong conservation in promoter structure among plants and other eucaryotes (8, 23, 28, 31), suggesting that the regulatory mechanisms responsible for thermal induction of transcription are highly conserved among higher eucaryotes. Evidence for the conservation of HS promoter function was demonstrated by the expression of the *Drosophila melanogaster hsp70* promoter in mammalian (5, 6, 21, 26) and amphibian (37) cells and in regenerated tobacco plants (33). Deletion studies, insertion of synthetic oligonucleotides, and analysis of sequence homology identified a region (HS consensus element [HSE]) 13 to 28 base pairs (bp) upstream from the TATA motif that determines thermoinducibility of the *D. melanogaster hsp70* promoter (25, 37). The HSE in *D. melanogaster* is 5′-CTnGAAnnTTCnAG-3′ (25, 26) and lies within a region that interacts specifically with the HS transcription factor (HSF) (24, 38).

HSE-like sequences, in addition to poorly defined upstream sequences, have been implicated in the transcriptional regulation of plant HS genes. Two overlapping HSE-like elements from soybean HS gene *Gmhsp17.3-B* were shown in transgenic tobacco to confer heat inducibility to a minimal promoter (CaMV 35S; TATA only) fused to a reporter gene (34). Gene *Gmhsp17.5-E* 5′-flanking sequences were deleted to position −95 in a previous study with a drastic reduction of thermoinducible activity (15). The −95 deletion mutation contained a TATA motif and one pair of overlapping HSE-like sequences positioned 18 bp upstream from TATA (site 1) and was still weakly heat inducible (five- to 10-fold). In plant HS genes, sequences other than the TATA-proximal HSEs have been shown to be required for full expression of the promoter. Deletion mutagenesis (1, 15) of two soybean HS genes encoding low-molecular-weight HSPs indicated that sequences upstream of the TATA-proximal HSEs also contribute to the amplitude of thermal induction. In the closely related *Gmhsp17.3-B (hs6871)*, an AT-rich upstream sequence (−81 to −306 bp from the start of transcription) has been shown to possess enhancerlike properties (1).

In this study, we have refined our earlier deletion analysis of the *Gmhsp17.5-E* promoter (15) by mapping four discrete *cis*-acting DNA elements controlling thermoinducible activation of transcription. Transgenic expression of a series of 5′ and internal substitution mutations was assessed by transcript mapping with a T-DNA-based vector system to introduce the altered promoters into crown gall tumors of sunflower. Precise measurement of transcriptional activity was assured by using a homologous reference gene as an internal standard (20) and by pooling RNA from a large number of tumors. The TATA-proximal HSEs centered at −60 (HSE site 1) and −93 (HSE site 2) and a region containing a conserved TATA-dyad motif were shown to make a positive contribution towards the magnitude of thermal induction.

MATERIALS AND METHODS

Construction of mutants. The gene *Gmhsp17.5-E* was subcloned as a 1.9-kilobase-pair *Bgl*II fragment into pUC8J. The resulting plasmid, designated pUC8J:*Gmhsp17.5-E*, contained 1,175 bp of the 5′-flanking sequence, 866 bp of the complete protein coding region, and 35 bp of the 3′-flanking region. The cloning vector pUC8J was derived from pUC8 (36) by the insertion of a *Bgl*II linker into the multiple cloning site between the *Bam*HI and *Xma*I restriction sites.

A series of 5′ and 3′ deletions spanning the 5′-flanking region was obtained by BAL 31 nuclease digestions according to the specifications of the manufacturer. (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). Plasmid pUC8J:*Gmhsp17.5-E* DNA was linearized with either *Pst*I for 5′ deletions starting at position −1175 or *Bam*HI for 3′

* Corresponding author.
† Florida Experiment Station journal series number 10008.
‡ Present address: Department of Botany, University of Georgia, Athens, GA 30602.

deletions starting in the coding region at position +158. Linearized plasmids were treated with BAL 31 nuclease, and the resulting deletion endpoints were converted into *Sal*I sites through linker addition. For construction of internal deletion and insertion mutations, appropriate 5′ and 3′ deletion clones were joined by ligation at the *Sal*I linker sites. Isolated 3′ deletion fragments (*Hin*dIII-*Sal*I) were subcloned into the *Hin*dIII-*Sal*I sites of the matching 5′ deletion fragments that were previously subcloned into pUC19. All deletions and substitutions were confirmed by DNA sequencing (29).

Construction of the HS reference gene vector. For precise quantification of promoter mutations, a test and reference copy of *Gmhsp17.5-E* were incorporated into the intermediate (shuttle) vector pW9 (4). The reference gene consisted of *Gmhsp17.5-E* with 1,175 bp of 5′-flanking sequences and a 15-bp internal insertion-duplication in the untranslated leader sequence (Fig. 1A), which served to distinguish test and reference gene transcripts by S1 nuclease analysis. The reference gene shuttle vector, pW9-EC (Fig. 1B), was completed by the cloning of the reference gene (*Asp*718 fragment, 1.9 kilobase pairs) into the *Kpn*I site of pW9. The 5′ and internally deleted test genes of *Gmhsp17.5-E* were then substituted for the small *Hin*dIII-*Bgl*II fragment of pW9-EC to form a series of double-gene shuttle vectors as shown in Fig. 1B. The expression levels of the reference gene as assessed by S1 nuclease mapping were identical in both vector orientations (data not shown).

Triparental matings and preparation of tumor RNA. Double-gene shuttle vectors containing various *Gmhsp17.5-E* promoter mutants were transferred from *Escherichia coli* LE392 (Rec$^-$) into *Agrobacterium tumefaciens* 15955 (Str^r) (4) by triparental conjugation (13). Inoculation of sunflower seedlings, tumor growth and harvesting, and temperature treatment conditions have been previously described (15). In order to average out the variability in expression between individual tumors, RNA from 200 to 300 tumors was pooled for analysis of each promoter mutation. Methods for the purification of total RNA and the fractionation of poly(A$^+$) RNA from control and HS tumors have been described in previous reports (7, 15).

S1 nuclease mapping and primer extension analysis of transcripts. S1 nuclease mapping was used routinely to establish the transcript levels of both *Gmhsp17.5-E* test and reference genes. Analyses were performed with 250 ng of sunflower tumor poly(A$^+$) RNA as described (7, 15), with modification of the hybridization conditions according to Murray (22). The RNA:DNA hybridization step was performed for 4 h at 45°C in 10 μl of a solution containing 3 M sodium trichloroacetate, 50 mM PIPES [piperazine-*N*,*N*′-bis (2-ethanesulfonic acid)], pH 7.0, and 10 mM EDTA. The hybridization probe for S1 nuclease mapping was the 263-bp *Eco*RI-*Bam*HI fragment from the reference gene (−90 to +173) (Fig. 1A), which was 5′ end labeled at the *Bam*HI site located at +173. The reference gene DNA was selected for use as the probe in order to ensure that the looped-out region (leader insert) in the DNA:RNA hybrid between the radiolabeled probe and test gene RNA would consist of DNA to provide S1 nuclease a large target for cleavage. Primer extension analysis was performed (Fig. 2) as described previously (9) by using a synthetic primer (5′-CTAGGGAG ATTGTAATG-3′) complementary to sequences from positions +51 to +67 in the reference gene or from +36 to +52 in the wild-type gene.

Transcriptional activities of the test genes were quantified by comparison of radioactivity in gel electrophoresis bands

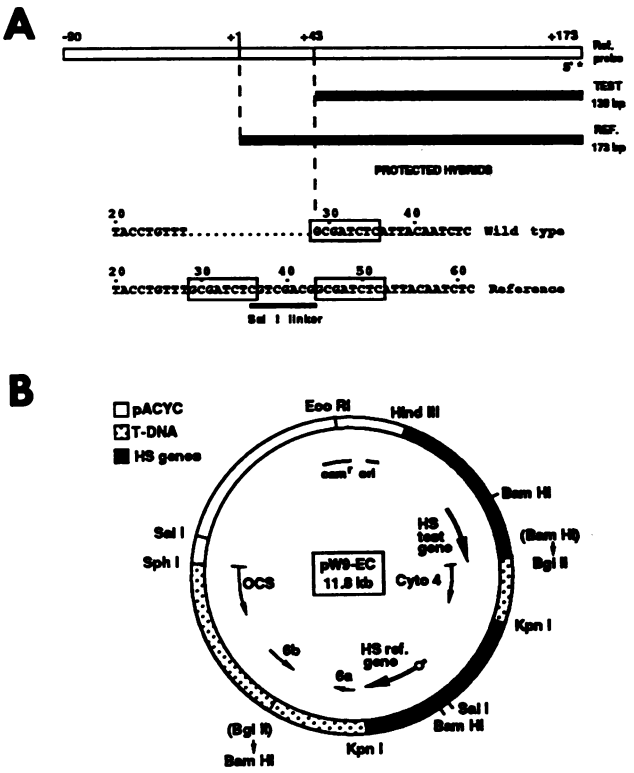

FIG. 1. (A) DNA sequence of altered 5′ untranslated region of HS reference gene. The DNA reference probe (□) used in S1 nuclease mapping analysis was an *Eco*RI-*Bam*HI fragment (−90 to +173) 5′ end labeled at the *Bam*HI site (+173). ■, S1 nuclease-protected hybrids; the shorter hybrid (130 bp) was derived from test RNA, and the longer hybrid (173 bp) was derived from reference RNA. The *Sal*I linker at position +37 was used to ligate a 3′ deletion mutation with a 5′ deletion mutation in order to recreate the reference gene with an altered leader sequence. Boxed sequences are duplicated on each side of the *Sal*I linker within the reference and wild-type leaders. Position +43 of the reference probe correlates with position +28 of the wild-type RNA, where the divergence in sequence homology between the two occurs. (B) Schematic diagram of the double-gene shuttle vector pW9-EC. The reference and the test HS genes were ligated into the pW9 intermediate shuttle vector consisting of T-DNA and pACYC 184 (4). Cam^r, Chloramphenicol resistance gene; ori, origin of replication; 6a and 6b, genes encoding T-DNA$_L$ transcripts; Cyto 4, isopentenyl pyrophosphatase transcript (cytokinin synthesis). *Bgl*II and *Bam*HI sites of T-DNA were changed by linker addition by *Bam*HI and *Bgl*II, respectively.

corresponding to test and reference gene transcripts (10). The positions of protected DNA fragments were determined by autoradiography of mapping gels. Excised gel slices containing protected fragments were dissolved in Insta-gel (Packard Instrument Co., Inc., Rockville, Md.), and radioactivity was determined by liquid scintillation counting. All values used for calculation of promoter activity were corrected for background radioactivity, which was determined by counting gel slices from portions of the gels containing no detectable bands. All values for relative transcription levels (RTLs) represent an average of three or more independent mapping experiments. The experimental variation was, on average, ±5% of the wild type.

When analyzed by S1 nuclease mapping, the undeleted test promoter appeared to have only 70% of the activity of the reference gene, although both copies of the promoter had identical 5′ flanking sequences. Primer extension analysis

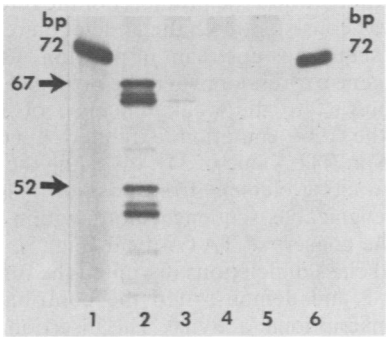

FIG. 2. Primer extension mapping of 5′ termini of *Gmhsp17.5-E* wild-type and reference gene transcripts. Lanes show results obtained using poly(A$^+$) RNA (1 μg) from sunflower tumors incited with *A. tumefaciens* 15955 Str^r (control; lanes 4 and 5) and transconjugant *A. tumefaciens* containing soybean wild-type (1175 bp of 5′ flanking sequences) HS test and reference genes (lanes 2 and 3). Prior to RNA extraction, tumors were either heat stressed at 40°C (lanes 2 and 4), or kept at 28°C (lanes 3 and 5). End-labeled *Hae*III restriction fragments of bacteriophage ϕX174 DNA were used as size markers in lanes 1 and 6.

(Fig. 2), however, showed that the full-length test and reference gene promoters had equal activity. All RTL values determined by S1 nuclease analysis were, therefore, normalized under the assumption that 70% activity of the test gene equaled full expression.

RESULTS

Activity of 5′ deletion mutants. Analysis of in vivo transcriptional activity of the 5′ deletion mutants indicated that the 5′ flanking sequences of *Gmhsp17.5-E* contain multiple *cis*-acting elements. Sequences 5′ to position −259 had very little influence on promoter activity; systematic 5′ deletions within this region decreased transcriptional activity by only 25% (Fig. 3 and 4). Removal of an additional 80 bp 5′ to position −179 (Δ5′-179) decreased transcriptional activity to 58%. This decrease may have been due to the removal of the overlapping secondary TATA-dyad motif (5′-TATAAAGA ATTTC-3′) centered at −220, which is also found in other small hsp genes of soybean (23) and *D. melanogaster* (32). Deletion of an additional 84 bp (Δ5′-95) resulted in an RTL value of 17, indicating the presence of one or more important *cis*-acting elements located between −179 and −95. The region from position −95 to −73 seemed to be neutral since the progressive removal of these sequences (Δ5′-90, Δ5′-80, Δ5′-73) did not significantly affect RTL values. Further 5′ deletion to position −61 (Δ5′-61) effectively removed the low-homology HSE (6 of 10 bp match) that overlaps the high-homology HSE1 (9 of 10 bp match) of site 1 (Fig. 5). The Δ5′-61 deletion mutant contained a single HSE at site 1 (one base deleted) and the TATA motif but was almost totally inactive (with an RTL value of 3). All 5′ deletion mutants to position −73 were heat inducible, indicating that at least one region conferring heat inducibility was located within the 73 bp of the transcription initiation site. Subsequent 5′ deletions that removed sequences downstream from −61 did not display significant transcriptional activity.

In sunflower tumors, the *Gmhsp17.5-E* promoter was expressed at high levels in heat-shocked tissue and weakly in the control tissue (28°C), unlike soybean seedlings, in which *Gmhsp17.5-E* expression is not detected at 28°C (7). Although some basal activity was detected in the transgenic tumors, the level of thermal induction was estimated to be at least 10-fold. The occurrence of basal activity of the *Gmhsp17.5-E* promoter in tumors is unexplained but may be caused by the long-range influence of upstream elements present in the flanking T-DNA promoters of the vector. The

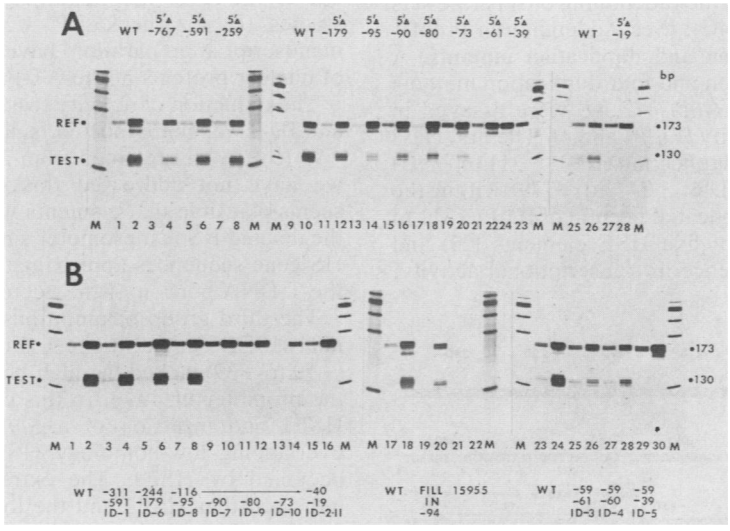

FIG. 3. S1 nuclease transcript mapping of 5′ termini of *Gmhsp17.5-E* wild-type (WT) and reference (REF) transcripts. (A) S1 nuclease analysis of 5′ deletion mutants. (B) S1 nuclease analysis of internal deletion-insertion mutants. The results of S1 mapping of poly(A$^+$) RNA from control (28°C) tumors are presented in the odd-numbered lanes, and those from HS (40°C) tumors are presented in the even-numbered lanes. With the reference gene as probe (Fig. 1), the 173-bp S1 nuclease-protected band was derived from hybrids formed with reference gene RNA and the 130-base band was derived from those formed with test gene RNA. The length of the protected wild-type (test gene) RNA: DNA hybrid corresponded to the predicted position of the 15-bp loop in the probe:RNA hybrid. Lanes WT (wild type) represent RNA from tumors containing full-length soybean test and reference genes (1,175 bp of 5′-flanking sequences). Lanes 21 and 22 in 3B show S1 nuclease mapping results with RNA isolated from sunflower tumors incited by *A. tumefaciens* 15955 Str^r (control; no soybean HS genes present), and lanes 19 and 20 show results with RNA from tumors containing the SV40 fill-in test gene. End-labeled *Hpa*II restriction fragments of pBR322 DNA ([in base pairs] 310, 271/281, 234, 194, and 118) were used as size markers in lanes M.

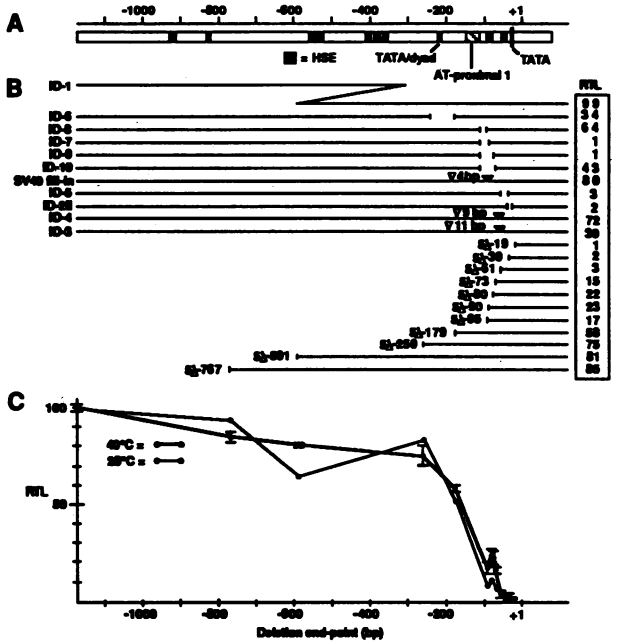

FIG. 4. Analysis of the effects of 5′ deletions on transcriptional activity of the *Gmhsp17.5-E* promoter. (A) The positions of potential *cis*-acting regulatory sequences, indicated by boxes. (B) RTLS were plotted from data shown in Fig. 3. Endpoints for the 5′ deletions are numbered from the transcriptional start site. (C) Analysis of the effects of deletions on transcriptional activity under control (○) and HS (●) conditions. TATA/dyad, TATA overlapping tetrameric palindrome; AT-proximal, AT-rich sequence; SV40, SV40-like enhancer core sequence.

activity plots of thermoinducible (40°C) and basal (28°C) activities closely paralleled, suggesting that both basal and thermoinducible activities differ only in the magnitude of activity and do not represent the functioning of separate sets of *cis*-acting elements (Fig. 4C). (See Addendum in Proof.)

Activity of internal deletion and duplication mutants. A total of seven internal deletion and four duplication mutants in the 5′ flanking region of *Gmhsp17.5-E* were assayed in vivo for transcriptional activity (Fig. 3 and 4). The first class of mutants consisted of one duplication (ID-1, −311 to −591) and one internal deletion (ID-6, −244 to −179) within the TATA-distal region. The duplicated portion of ID-1 (−311 to −591) contained a cluster of five HSE elements (23) and showed no significant difference in transcriptional activity,

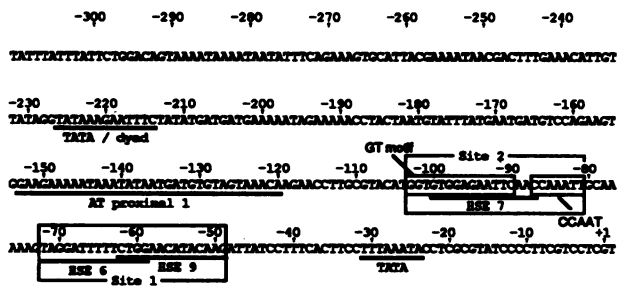

FIG. 5. Sequence of the 5′-flanking region of *Gmhsp17.5-E*. Sites 1 and 2, a potential CCAAT box, and the SV40 enhancer GT motif are boxed. HSEs (numbers indicate base pair homology of 10 bp specified), the AT-proximal 1 sequence, and the TATA-dyad are underlined. Numbers correspond to the start sites for transcription.

compared with the wild-type promoter of the reference gene. This result was consistent with the 5′ deletion results, which showed that sequences upstream of position −259 had little influence on gene expression when the proximal domain was intact. In contrast to the weak influence of TATA-distal sequences, the ID-6 mutation exhibited a 66% drop in activity (with an RTL value of 34), revealing the presence of a positive *cis*-acting element from position −244 to −179. The only recognizable sequence motif within the deleted region was the conserved TATA-dyad (Fig. 5).

The second class of deletions disrupted the function of the HSE at site 2 and demonstrated the contribution of this region to transcriptional activity. The insertion (designated simian virus 40 [SV40] fill-in) of 4 bp, filling in the *Eco*RI site in HSE2 at position −94, resulted in a 20% decrease in activity. The vital importance of site 2 was demonstrated more clearly by the results from a series of internal deletions designated ID-8 (−116 to −95), ID-7 (−116 to −90), ID-9 (−116 to −80), and ID-10 (−116 to −73), which progressively moved upstream sequences from −1175 to −116 closer to the start of transcription by removing sequences between −116 and TATA. The removal of 21 bp (ID-8) from −116 to −95 partially deleted HSE2 and caused a drop in transcriptional activity to 64%. Two other internal deletions, ID-7 (−116 to −90) and ID-9 (−116 to −80), completely removed site 2 and effectively abolished promoter activity (to an RTL value of 1).

When the upstream sequences from −1175 to −116 were joined by the addition of an 8-bp *Sal*I linker to the proximal promoter at position −73 (ID-10), partial activity was restored (to an RTL value of 43). Site 2 in construction ID-10 was occupied by the AT-rich sequences (designated AT-proximal 1) normally positioned from −120 to −153. One interpretation of this partial restoration of activity by the substitution of AT-rich sequences for the HSP at site 2 is that these AT-rich sequences may function as a *cis* element which is able to partially compensate for the missing HSE at site 2. We favor this hypothesis since in vitro footprinting studies (E. Czarnecka, P. C. Fox, and W. B. Gurley, manuscript in preparation) have shown a specific interaction of nuclear proteins at the AT-proximal 1 site.

The reduction of activity to 1% for internal deletions ID-7 and ID-9 was not expected, since 5′ deletions from position −95 to −73 still retained from 23 to 15% activity. Although we have not addressed this question experimentally, it seems plausible that elements within T-DNA may stimulate the deleted HS test promoter since the 5′ deletions move the HS gene sequences approximately 1 kilobase pair closer to the T-DNA portion of the vector.

The third group of mutations disrupted the TATA-proximal HSE1 located in site 1 (Fig. 5). Internal deletion ID-5 (−59 to −39) lacked the high-homology HSE1 and rendered the promoter inactive. In this construction, the excision of HSE1 and insertion of a *Sal*I linker (8 bp) moved the overlapping low-homology HSE to the position normally occupied by HSE1. The extremely low activity of ID-5 indicated that HSE1 and the low-homology HSE were not functionally interchangeable. The insertion of 9 bp (ID-4, −59 to −60) and 11 bp (ID-3, −59 to −61) between HSE1 and its overlapping HSE caused a drop in transcriptional activity to 72 and 30%, respectively. The abrupt decrease in activity seen between the insertion of 9 and 11 bp suggested that there was only a limited degree of flexibility in the relative positions of HSE1 and the overlapping HSE.

The final deletion class contained a single internal deletion mutant (ID-2ii) that demonstrated the essential role of

TATA. ID-2ii lacked sequences from −40 to −19, which included TATA, and resulted in a complete loss of promoter function (with an RTL value of 2). The loss of activity resulting from the removal of TATA was comparable to results obtained upon removal of HSE1 (ID-5) at site 1 or HSE2 (ID-9 and ID-7) at site 2, suggesting that all three of these elements were major determinants of activity.

DISCUSSION

TATA-distal domain. The upstream region of *Gmhsp17.5-E* from −1175 to −179 (TATA-distal domain) contains several AT-rich areas and a total of seven HSE-like sequences. The presence of AT-rich regions is not unique to *Gmhsp17.5-E*; large blocks of AT-rich sequences are also present in the 5′-flanking sequences of other soybean HS genes (23, 27). At least one of these AT-rich sequence blocks (200 bp), upstream of the *Gmhsp17.3-B* gene of soybean, has been shown to contain a major determinant of promoter activity (1). Although there is circumstantial evidence suggesting that the AT-rich region from −153 to −120 in *Gmhsp17.5-E* contains an active *cis*-acting element, similar sequences in the TATA-distal domain seemed to have little transcriptional activity. The lack of evidence indicating that these sequences contribute to promoter strength in vivo is contrasted by the strong binding to this region in vitro obtained with crude nuclear extracts.

The TATA-distal domain also contains multiple regions with homology to the *Drosophila* HSE, but removal or duplication of these sequences had little effect on the thermoinducible activity of *Gmhsp17.5-E*. This apparent lack of activity of the distal HSEs was similar to that seen in vivo with the *Drosophila hsp70* promoter incorporated into a P-element vector (11). Although HSEs have been shown to act at long range when duplicated and placed upstream of the *Xenopus laevis hsp70* and human β-globin genes (3), distal HSEs are sometimes functionally redundant as in the case of the *Drosophila hsp70* gene (35). Redundant HSEs are able to substitute for proximal HSEs when the latter are removed or disrupted by mutagenesis.

A subdomain of major importance within the TATA-distal region is located from position −244 to −180 and contributes approximately 66% to promoter activity (ID-6). The only obvious consensus sequence located within this 65 bp is an upstream TATA and overlapping partial HSE dyad (5′-TATAAAGAATTTC-3′). This upstream TATA-dyad has not been shown to be involved in thermal induction in animal systems, but its conservation both in *Drosophila* genes (*hsp22*, *hsp26*, and *hsp70*) (32) and soybean HS genes (*Gmhsp17.5-M*, *Gmhsp17.5-L*, *Gmhsp17.3-B*) (23), together with the results of our internal and 5′ deletion analyses, suggests that it can contribute to the transcriptional activity of HS promoters. The involvement of the TATA-dyad in transcriptional regulation is also consistent with in vitro footprinting studies which show the TATA-dyad to be protected from DNase I digestion by proteins in crude nuclear extracts from soybean plumules (Czarnecka et al., manuscript in preparation).

TATA-proximal domain. Our analysis of internal and 5′ deletions within the TATA-proximal domain from −179 to the start of transcription indicated that this region is of primary significance in the thermoinducibility of *Gmhsp17.5-E*. The positioning of HSEs within this region is generally conserved among soybean HS genes sequenced to date (23). DNase I footprinting studies with restriction fragments mapping within the 5′ flanking region of *Gmhsp17.5-E* indicated that the entire TATA-proximal domain was bound in a DNA-protein complex in vitro by nuclear factors from soybean plumules (Czarnecka et al., manuscript in preparation). The apparent continuity of protein-DNA interaction sites from TATA to −153 and the density of potential control sequences in this area was the basis for our designation of this region as a single TATA-proximal domain.

The TATA-proximal domain of *Gmhsp17.5-E* contains two HS control regions located at sites 1 and 2 (Fig. 5). Site 1 consists of an overlapping pair of HSEs; one with a 6- and the other with a 9-bp match with the *Drosophila* HSE. Internal deletion of either site 1 (ID-5) or site 2 (ID-7) abolished transcriptional activity (Fig. 3). The requirement of a second HSE for full activity has also been seen in deletion studies with *Drosophila* (11, 35) and *Xenopus hsp70* (3) genes. The *Drosophila hsp70* HSE at site 1 (TATA proximal) has a five- to 12.5-fold-higher affinity for HSF than that of the low-homology HSE at site 2 (−92 to −66) (35). Topol et al. (35) have suggested that protein cooperation between sites 1 and 2 is involved in the binding of HSF to site 2. The binding of HSF to site 2 of Drosophila *hsp70* is postulated to serve as the switch that activates full thermal induction of the promoter. In the soybean *Gmhsp17.5-E* promoter, site 2 may serve a similar function; the binding of HSF to site 2 may, however, be regulated by more than simple HSF abundance.

Inspection of the DNA sequence at site 2 of *Gmhsp17.5-E* suggests that its potential function as a switch may involve *cis* elements other than the single low-homology HSE. For example, an SV40 enhancer core sequence (19) (homology, 11 of 14 bp) extends from −103 to −90 overlapping the site 2 HSE. The *rubcS-3A* gene box II sequence (5′-GTGTGGT TAATATG-3′) (19) is also similar to the SV40 enhancer core and the GT motif within site 2 of *Gmhsp17.5-E*. The box II element binds a specific nuclear factor (GT-1) from pea cells (14) and acts as a negative element to prevent transcriptional expression of *rubcS-3A* in the dark (18). Another SV40-like GT motif (5′-GTGGTTTG-3′) in plant genes is present in the maize *Adh-1* promoter within the region involved in anaerobic regulation (12). This sequence appears to be a region of close contact between DNA and proteins in vivo as assessed by altered dimethyl sulfate reactivity. The GT motifs in *Gmhsp17.5-E* and *Adh-1* are immediately preceded 5′ by a short stretch of alternating pyrimidine-purine residues. Although it has not been directly tested, the presence of the GT motif near site 2 raises the possibility that the *Gmhsp17.5-E* promoter may be regulated through a variety of environmental signals.

Another potential *cis* element at site 2 is the CCAAT-box-like sequence (−87 to −81) immediately 3′ to the HSE. In other organisms, the CCAAT box has been shown to be involved in the regulation of HS genes that are also induced by non-HS conditions. For example, when the HSE was placed immediately upstream of the *Xenopus hsp70* CCAAT box, the promoter showed an elevated constitutive expression in oocytes, compared with the CCAAT box alone (2). It has been proposed that CCAAT-binding protein interacts with HSF to enhance its binding to the HSE in the absence of HS (2, 3). In the human *hsp70* gene, an inverted CCAAT box is centered at −65 and is required for basal- and serum-stimulated activity (39). Although we have not directly assessed the contribution of the CCAAT-like sequences to *Gmhsp17.5-E* promoter function, the close spacing between CCAAT and the HSE raises the possibility of protein-binding cooperation between the HSE and a CCAAT-binding protein at site 2.

TATA motif. In eucaryotes, including higher plants, a TATA motif is frequently present, positioned 21 to 35 bp upstream of the start of transcription. In animal genes, TATA has been shown to be an essential component of many promoters (10). In this respect, the soybean *Gmhsp17.5-E* promoter is a member of this common class of eucaryotic promoters which include HS promoters from animal systems as well (2). Our internal (ID-2ii) and 5′ deletion studies indicate that the TATA element (5′-TTTA AATA-3′) is absolutely required for transcriptional activity of the *Gmhsp17.5-E* promoter (Fig. 3) and that TATA alone is not sufficient for thermal activation.

Summary. We have identified four regions (TATA-dyad, site 2, site 1, and TATA) within the 5′-flanking sequences of *Gmhsp17.5-E* that make a positive contribution to heat-inducible transcription. A fifth site with potential for transcriptional control comprised of AT-rich sequences (AT-proximal 1) has been tentatively identified by in vitro DNA binding studies (Czarnecka et al., manuscript in preparation). The overall organization of the *Gmhsp17.5-E* promoter closely parallels that of *Drosophila hsp70* with analogous sites 1 and 2, and the AT-rich and TATA-dyad elements positioned similarly to sites 3 (HSE, −189 to −165) and 4 (HSE, −258 to −235) of the *Drosophila* gene (35). The function of sites 1 and 2 and the TATA-dyad may all be mediated by HSF binding caused by the HSE homology present within each of these regions; these elements are probably responsible for transcriptional activation of the gene upon exposure to a variety of stresses typically associated with the HS response. The TATA-dyad, however, also shows sequence similarity with the AT-proximal 1 region and may therefore interact with *trans*-acting factors that bind AT-rich sequences. As with the *Drosophila hsp70* promoter, it seems likely that the binding of transcription factors to site 2 may constitute the primary switch in activity seen upon thermal induction. Site 2 of *Gmhsp17.5-E* may also be a control point for other environmental or physiological influences caused by the juxtaposition of other potential *cis*-acting elements such as the CCAAT box and GT motif.

The identification of specific *trans*-acting proteins that contribute to the function of the promoter is the next phase in the development of a functional model describing the integration of multiple signals in the regulation of stress-induced transcription. The AT-proximal 1 site is very similar to an upstream sequence [5′-A T T (A/T) A A T-3′] in an embryo-specific soybean lectin gene (17), to two sites (5′-TNAATNNTTTATTT-3′) of protein-DNA interaction in the 5′-flanking sequences of a nodule-specific leghemoglobin gene (lbc_3) (16), and to a potential regulatory sequence (5′-TGTAATAATATATTTATATTTT-3′) in the β-conglycinin promoter (R. D. Allen, R. D. Lessard, and R. N. Beachy, Abstr. J. Cell. Biochem. Suppl. 12C, p. 223, abstr. no. LT09, 1988). The presence of AT-rich *cis* elements within the 5′-flanking sequences of a variety of highly regulated genes suggests that these sequences may represent a common class of promoter elements.

ACKNOWLEDGMENTS

We thank Sabita Bandyopadhyay and Deirdre Yawn for their excellent technical support and Dulce Barros and Paul Fox for critical reading of the manuscript.

This work was supported by contracts from Lubrizol Genetics, Inc., to W.B.G. and J.L.K.

ADDENDUM IN PROOF

A recent report by Park and Craig (H.-O. Park and E. A. Craig, Mol. Cell. Biol. **9**:2025–2033, 1989) indicates that in *Saccharomyces cerevisiae*, HSE2 plays an important role in the basal as well as heat-inducible expression of the *SSA1* gene. Two base substitutions in HSE2 caused a decrease in the basal activity of the *SSA1* promoter to 18% of the level of the wild-type promoter and altered kinetics of gene expression after heat shock.

LITERATURE CITED

1. **Baumann, G., E. Raschke, M. Bevan, and F. Schöffl.** 1987. Functional analysis of sequences required for transcriptional activation of a soybean heat shock gene in transgenic tobacco plants. EMBO J. **6**:1161–1166.
2. **Bienz, M.** 1986. A CCAAT box confers cell-type-specific regulation on the Xenopus *hsp70* gene in oocytes. Cell **46**:1037–1042.
3. **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.
4. **Bruce, W. B., and W. B. Gurley.** 1987. Functional domains of a T-DNA promoter active in crown gall tumors. Mol. Cell. Biol. **7**:59–67.
5. **Burke, J. F., and D. Ish-Hrowicz.** 1982. Expression of *Drosophila* heat shock genes is regulated in Rat-1 cells. Nucleic Acids Res. **10**:3821–3830.
6. **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.
7. **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.
8. **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.
9. **Czarnecka, E., R. T. Nagao, J. L. Key, and W. B. Gurley.** 1988. Characterization of *Gmhsp26-A*, a stress gene encoding a divergent heat shock protein of soybean: heavy-metal-induced inhibition of intron processing. Mol. Cell. Biol. **8**:1113–1122.
10. **Dierks, P., A. van Ooyen, M. Cochran, C. Dobkin, J. Reiser, and C. Weissmann.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**:695–706.
11. **Dudler, R., and A. A. Travers.** 1984. Upstream sequences necessary for optimal function of the hsp 70 promoter in transformed flies. Cell **38**:391–398.
12. **Ferl, R. J., and H. S. Nick.** 1987. *In vivo* detection of regulatory factor binding sites in the 5′ flanking region of maize *Adh-1*. J. Biol. Chem. **262**:7947–7950.
13. **Fraley, R. T., S. G. Rogers, R. B. Horsch, P. Sanders, J. Flick, S. Adams, M. L. Bittner, L. Fink, J. S. Fry, G. R. Galluppi, S. B. Goldberg, N. L. Hoffman, and S. C. Woo.** 1983. Expression of bacterial genes in plant cells. Proc. Natl. Acad. Sci. USA **80**:4803–4807.
14. **Green, P. J., S. A. Kay, and N.-H. Chua.** 1987. Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. EMBO J. **6**:2543–2549.
15. **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.
16. **Jensen, Ø. E., K. A. Marcker, J. Schell, and F. J. de Bruijn.** 1988. Interaction of a nodule specific, *trans*-acting factor with distinct DNA elements in the soybean leghaemoglobin lbc_3 5′ upstream region. EMBO J. **7**:1265–1271.
17. **Jofuku, K. D., J. K. Okamuro, and R. B. Goldberg.** 1987. Interaction of an embryo DNA binding protein with a soybean lectin gene upstream. Nature (London) **328**:734–737.
18. **Kuhlemeier, C., R. Fluhr, P. J. Green, and N.-H. Chua.** 1987.

Sequences in the pea *rbc-3A* gene have homology to constitutive mammalian enhancers but function as negative regulatory elements. Genes & Dev. **1**:247–255.

19. **Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA **79**:6453–6457.
20. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein coding gene. Science **217**:316–325.
21. **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.
22. **Murray, M. G.** 1986. Use of sodium trichloroacetate and Mung Bean nuclease to increase sensitivity and precision during transcript mapping. Anal. Biochem. **158**:165–170.
23. **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.
24. **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.
25. **Pelham, H. R. B.** 1982. A regulatory upstream promoter element in the *Drosophila* hsp 70 heat-shock gene. Cell **30**:517–528.
26. **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.
27. **Raschke, E., G. Baumann, and F. Schöffl.** 1988. Nucleotide sequence analysis of soybean small heat shock protein genes belonging to two different multigene families. J. Mol. Biol. **199**:549–557.
28. **Rochester, D. E., J. A. Winter, and D. M. Shah.** 1986. The structure and expression of maize genes encoding the major heat shock protein, hsp 70. EMBO J. **5**:451–458.
29. **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.
30. **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.
31. **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.
32. **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.
33. **Spena, A., and J. Schell.** 1987. The expression of a heat-inducible chimeric gene in transgenic tobacco plants. Mol. Gen. Genet. **206**:436–440.
34. **Strittmatter, G., and N.-H. Chua.** 1987. Artificial combination of two cis-regulatory elements generates a unique pattern of expression in transgenic plants. Proc. Natl. Acad. Sci. USA **84**:8986–8990.
35. **Topol, J., D. M. Ruden, and C. S. Parker.** 1985. Sequences required for *in vitro* transcriptional activation of a *Drosophila hsp70* gene. Cell **42**:527–537.
36. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13am7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
37. **Voellmy, R., and D. Rungger.** 1982. Transcription of a *Drosophila* heat shock gene is heat-induced in *Xenopus* oocytes. Proc. Natl. Acad. Sci. USA **79**:1776–1780.
38. **Wu, C.** 1984. Activating protein factor binds *in vitro* to upstream control sequences in heat shock gene chromatin. Nature (London) **311**:81–84.
39. **Wu, B., R. Kingston, and R. Morimoto.** 1986. Human *HSP70* promoter contains at least two distinct regulatory domains. Proc. Natl. Acad. Sci. USA **83**:629–633.