Cell, Vol. 30, 517–528, September 1982, Copyright © 1982 by MIT

A Regulatory Upstream Promoter Element in the Drosophila Hsp 70 Heat-Shock Gene

Hugh R. B. Pelham
M. R. C. Laboratory of Molecular Biology
Hills Road
Cambridge CB2 2QH, England

Summary

Deletion mutants of the Drosophila hsp 70 heat-shock gene have been assayed in COS cells using a vector that contains an SV40 replication origin. COS cells are SV40-transformed monkey cells that support high-level replication of the vector. S1 mapping of transcripts shows that the hsp 70 gene is heat-inducible in these cells, whereas the herpes virus thymidine kinase gene is not. Residues −10 to −66 of the hsp 70 gene are apparently sufficient for heat-inducible promotion, and residues between −47 and −66 are necessary. This region upstream of the TATA box contains sequence features common to other heat-shock genes. Expression of the hsp 70 gene can be forced at low temperature by SV40 sequences that include the 72 bp repeat, but only if these are present on the 5′ side of the gene. It seems that the upstream element of the hsp 70 promoter is analogous to that of other promoters, but is only functional in heat-shocked cells.

Introduction

In response to an increase in temperature, cells from a wide range of organisms rapidly synthesize a small number of proteins (the heat-shock proteins or hsp's) while repressing the synthesis of most other proteins. This appears to be an ancient, evolutionarily conserved response to thermal stress, and it evidently serves a protective function. A similar response can be induced at constant temperature by other stressful agents, including certain metabolic inhibitors, ionophores and recovery from anoxia (reviewed by Ashburner and Bonner, 1979).

In Drosophila, heat shock has been shown to result in greatly increased transcription of seven different protein-coding genes (Ashburner and Bonner, 1979). Copies of all seven of the heat-shock genes from D. melanogaster have been cloned and their promoter regions sequenced (Karch et al., 1981; Torok and Karch, 1980; Ingolia and Craig, 1981; Ingolia et al., 1980; Holmgren et al., 1981).

The most highly conserved hsp is a 70 kd polypeptide (hsp 70). Most D. melanogaster strains have five copies of this gene per haploid genome, at two different cytogenetic loci (Ashburner and Bonner, 1979). Several copies of the gene have been sequenced; they are closely homologous to each other, and the homology extends for about 350 bases on the 5′ side of the transcribed portion of the genes (Karch et al., 1981; Ingolia et al., 1980). Recent genetic evidence indicates that no more than 479 bases of flanking sequence is required for activity of one of the genes in vivo, and all transcriptional and regulatory signals probably are contained in the homologous region (Udvardy et al., 1982). The genes contain no intervening sequences.

Recently it has been shown that expression of cloned copies of the hsp 70 gene introduced into the genome of mouse tissue culture cells can be controlled by heat shock, indicating a considerable conservation of the regulatory signals (Corces et al., 1981). Regulation must involve interaction of specific cellular factors with the gene because the temperature required to induce activity is characteristic of the mouse cells, not of the flies from which the gene was originally obtained.

I describe experiments designed to locate the regulatory signals of the hsp 70 gene and to define the level at which regulation occurs. The assay I have used involves the introduction of plasmids containing an SV40 origin of replication into COS cells. These are monkey cells containing a defective integrated SV40 genome; they produce T antigen constitutively, and thus support high-level replication of the input DNA (Gluzman, 1981; Mellon et al., 1981).

The results can be interpreted in the light of what is known about eucaryotic promoters. These can be divided into at least two functional elements. One contains the TATA box which is involved in defining the position at which RNA polymerase II initiates transcription both in vivo (Grosveld et al., 1982; Benoist and Chambon, 1981; Grosschedl and Birnstiel, 1980a) and in vitro (Corden et al., 1980; Grosveld et al., 1981). The other is an upstream element, somewhat variable in position, that is required in vivo for efficient use of the TATA box (Benoist and Chambon, 1981; Dierks et al., 1981; Grosschedl and Birnstiel, 1980a, 1980b; Grosveld et al., 1982; Mellon et al., 1981; McKnight et al., 1981; Moreau et al., 1981). The evidence suggests that heat-shock genes have an upstream element with specific sequence characteristics that directs efficient use of the TATA box in heat-shocked cells, but is nonfunctional under normal conditions.

Results

Characteristics of the COS Cell System

The plasmids used in this work are diagrammed in Figure 1; further details are given in Figure 8 and Experimental Procedures. The basic vector is a pBR322 derivative that lacks the "poison sequences" that inhibit replication in animal cells (Lusky and Botchan, 1981). To this is joined a 240 bp fragment of SV40 DNA that contains the origin of replication, but has only a partial copy of one of the 72 bp repeats, and thus does not contain an efficient promoter (Benoist and Chambon, 1981). The standard plasmid pHT1

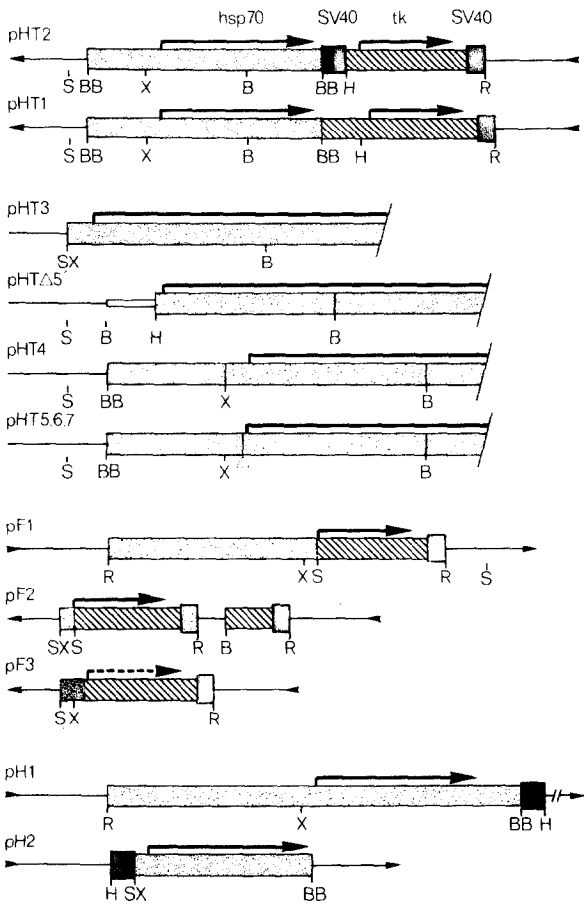

Figure 1. Structure of Plasmids

Thin lines: procaryotic vector sequences; diagonal stripes: tk sequences; stipples: hsp 70 gene and flanking sequence. Darker stippled areas are derived from plasmid 132E3, lighter ones from plasmid 56H8. Wider shaded boxes: SV40 sequences, with the 72 bp repeat region in black. Arrows above genes indicate transcripts. For further details see Figures 5 and 8 and Experimental Procedures. Note that pHT1 is the second plasmid shown. Only the 5′ part of the hsp 70 gene, enlarged twofold, is shown for plasmids pHT3–7 and the pHTΔ5′ series; the rest of the plasmid is in each case identical to pHT1. The thin box in the pHTΔ5′ plasmids represents 5S DNA spacer sequences, or in the case of pHTΔ5′-66* sequences from upstream of the tk gene.

The plasmids are shown linearized arbitrarily in the vector sequences; arrows on these indicate their orientation, pointing in the direction from the Eco RI site toward the Sal I site in the Tet^R gene. Selected restriction sites are abbreviated as follows: B = Bam HI; BB = Bam HI fused to Bgl II; H = Hind III; R = Eco RI; S = Sal I; SX = Sal I fused to Xho I; X = Xho I. Plasmid pF2 was constructed as part of a separate experiment, and the extra tk sequence shown on the right of the plasmid is irrelevant to the results presented here.

(shown on the second line of Figure 1) contains also a complete hsp 70 gene, including 1100 bases of 5′ flanking sequence, and the herpes virus thymidine kinase (tk) gene as an internal control.

The system was characterized using the S1 mapping technique (Berk and Sharp, 1977; Weaver and Weissman, 1979) to detect and quantitate cytoplasmic tk transcripts. DNA was introduced into the cells in the presence of DEAE-dextran, a method that for these cells is simpler and at least as efficient as calcium phosphate coprecipitation. The ultimate concentration of transcripts was not affected by varying the amount of input DNA between 2 and 20 μg per 75 cm^2 flask, and was achieved within 30 hr of exposure of the cells to DNA; the transcripts remained at this level for at least another 10–15 hr. Plasmids containing two copies of the SV40 replication origin replicated faster, the transcript level saturating by about 20 hr, but the final abundance of transcripts was not significantly higher than that obtained with plasmids containing a single origin (see below). Plasmids lacking SV40 sequences yielded no detectable transcripts under these conditions. In subsequent experiments, flasks of cells were transfected with 2–5 μg of DNA, and the cells were harvested 35 hr after removal of the DNA.

Because the 72 bp repeat region of SV40 has been implicated as a long-range cis-acting enhancer of the expression of other genes (Banerji et al., 1981; Moreau et al., 1981), the properties of plasmid pHT1 were compared to those of pHT2, which has an extra copy of the SV40 origin including the complete 72 bp repeat region (see Figure 1). (The partial copy of the 72 bp repeat present on all the plasmids is not sufficient for enhancer activity, as shown by Banerji et al., 1981.) Figure 2A shows the results obtained by S1 mapping the 5′ ends of tk transcripts obtained from these plasmids. The probe was a 131 bp Eco RI–Bgl II DNA fragment, ^{32}P-labeled at the Bgl II site with polynucleotide kinase (see Figure 2). Hybridization of authentic tk mRNA to this fragment protects a labeled fragment of about 54–56 bases from S1 nuclease, and with probe added in excess, the amount of label recovered in the protected fragments reflects the abundance of tk transcripts (McKnight et al., 1981).

In addition to the expected tk transcripts, a number of transcripts originating upstream of the normal 5′ end were observed in COS cells, indeed, these were often more abundant than the correct transcripts. Figure 2A also shows that the presence of the second SV40 replication origin and 72 bp repeat apparently increased the overall abundance of tk transcripts (at least in non-heat-shocked cells). This result was not, however, obtained reproducibly. The effect seen in Figure 2A may be due to the second origin of replication; this alone can increase the copy number of the plasmid at early times, and hence the level of transcripts (data not shown). The variability between experiments probably reflects slight differences in the rate of plasmid replication in the COS cells, as well as some inherent variability of the S1 assay. In general, the 72 bp repeat appeared to have little effect on tk gene expression with the plasmid constructions and assay conditions that I have used. This was true whether or not the hsp 70 gene was also present on the plasmid.

The extra 5′ ends of the tk transcripts map upstream from the TATA box, and some are within the essential upstream region defined by deletion mapping of the

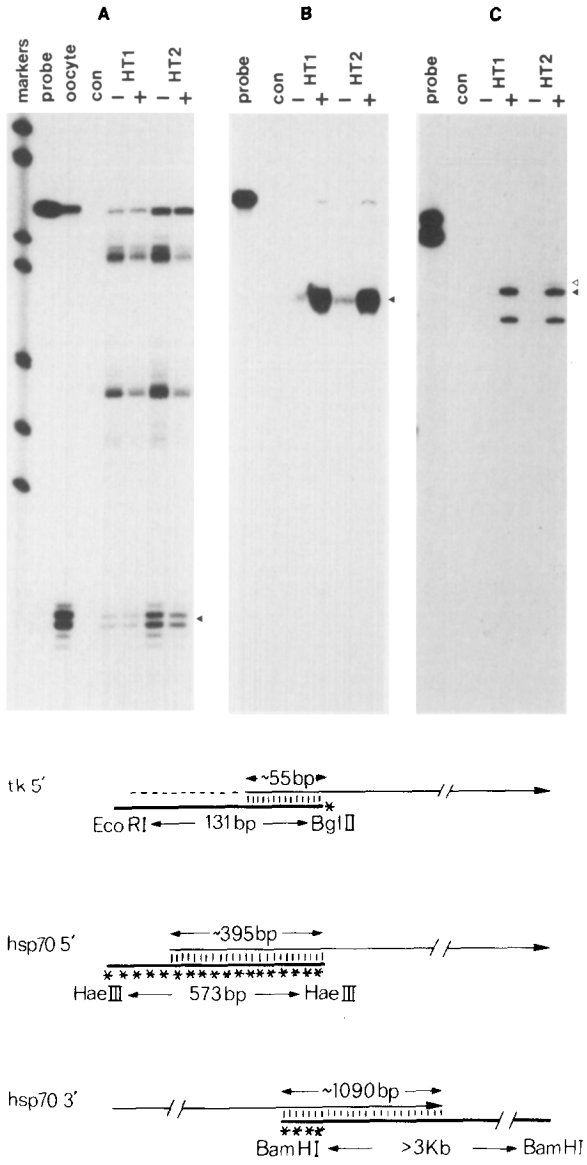

Figure 2. S1-Mapping of tk and hsp 70 Transcripts Synthesized in COS Cells

The same RNA samples were hybridized to the 5′ tk probe (A), the 5′ hsp 70 probe (B) and the 3′ hsp 70 probe (C). The hybrids were digested with nuclease S1 and analyzed on gels as described in Experimental Procedures. Autoradiographs of portions of the gels are shown. Markers visible in (A) are fragments of 162, 149, 124, 112, 92, 78 and 69 bp (pBR322/Hpa II). Also shown for reference is the S1 map of tk transcripts obtained after injection of the gene into Xenopus oocytes. Cells were transfected with pHT1 or pHT2 (see Figure 1) and either heat shocked (+) or not (–). Control RNA (con) was from cells that were mock-transfected and heat-shocked. The line drawing outlines the S1 mapping schematically (not to scale). Thin lines represent mRNA, with arrowheads at the 3′ end, thick lines represent the DNA probes, with asterisks indicating the position of ^{32}P label. The positions of the expected protected fragments on the gels are indicated by solid triangles in A–C. The open triangle in C indicates a band derived from readthrough transcripts which map to the point of sequence divergence between probe and template. The hsp 70 3′ probe contains two hybridizing species and yields two protected fragments as explained in the text. Note that panels A–C are not directly comparable; the probes were not of the same specific activity, the gels were of different acrylamide concentrations, and the autoradiographic exposures differ.

tk promoter (McKnight et al., 1981). They have been found consistently, whether or not there are SV40 or hsp 70 sequences upstream from the tk gene, and they are also synthesized in vitro in HeLa cell extracts (P. Farrell and H. Pelham, unpublished observations). This suggests that the extra 5′ ends correspond to true transcriptional starts, but their significance remains unclear.

Expression of the Hsp 70 Gene Is Heat-Inducible in COS Cells

Hsp 70 transcripts were detected using a 5′ S1 probe consisting of a 573 bp Hae III fragment continuously labeled in the mRNA complementary strand (see Figure 2 and Experimental Procedures). Authentic mRNA from heat-shocked Drosophila cells protects a ~400-base fragment from S1 nuclease (data not shown; see also Torok and Karch, 1980; Ingolia et al., 1980). There was also some weak cross-hybridization with monkey cell heat-shock RNA, but the protected fragments were less than 200 bases long and did not interfere with the assay.

Figure 2B shows that incubating the COS cells at 42°C from 30 to 35 hr after transfection increased the abundance of hsp 70 transcripts about 50-fold, and these transcripts had the expected 5′ ends. In contrast, heat shock slightly reduced the abundance of tk transcripts (Figure 2A). Since most of these were synthesized prior to the heat shock, it is impossible to say whether tk transcription continues at 42°C. However, in many experiments the level of at least the larger tk transcripts was reduced by heat shock; this may indicate a reduction in the rate of transcription and some turnover of preexisting RNA during heat shock.

The 3′ ends of the hsp 70 transcripts were also located by S1 mapping (Figure 2C). The probe consisted of the entire plasmid 132E3, which contains two copies of the hsp 70 gene. This was labeled at the Bam HI sites in the centers of the genes (see Figure 8B). Because the gene sequences diverge on the 3′ side of the coding region, two protected fragments were obtained; one ends at the point of divergence, the other at the true 3′ end of the transcript. The latter mapped to the expected position within the hsp 70 sequence, but the large size of the protected fragment precluded accurate mapping. As expected, abundant protected fragments were only obtained with RNA from heat-shocked cells.

These results establish that expression of the Drosophila hsp 70 gene can be controlled in monkey cells, even when the gene is present on a plasmid at many thousands of copies per cell and is adjacent to a gene that is expressed at low temperature. The level of induction of hsp 70 mRNA (typically 20 to 50 fold) is comparable to the induction of monkey heat-shock RNA under the same conditions, as detected by cross-hybridization with the Drosophila probe (data not shown). The presence of the SV40 72 bp repeat at

the 3′ end of the hsp 70 gene (in pHT2 but not in pHT1) had no effect on this control or on the final level of transcripts obtained (Figure 2).

Fusion of the Hsp 70 Gene to the tk Gene

The results described above do not rule out the possibility that control of hsp 70 gene expression is at a posttranscriptional level in COS cells; for example, the transcripts might be selectively degraded at low temperature. To address this question, fusions of the hsp 70 promoter region to the tk gene were constructed (see Figure 1) and tested (Figure 3). The first of these fusions, pF1, contains about 2 kb of the 5′ flanking region of a hsp 70 gene up to position −10 joined via a synthetic linker to a Hae III site 8 bases upstream from the tk initiation site, the distance from TATA box to start site being exactly maintained. (Throughout this paper positions are numbered relative to the transcription start point of the gene concerned.) The hsp 70 gene used in this construction was derived from the 87A7 cytogenetic locus; it has a sequence very similar to the 87C1 gene used previously out to about position −350. In subsequent constructions, parts of both genes were used for convenience, but because of small differences in the untranslated portion of the mRNAs it was necessary always to use the appropriate probe for S1 mapping.

The fused gene in plasmid pF1 was clearly heat-inducible, transcription initiating at the normal tk start site and also a few bases upstream (Figure 3). In fact, transcripts synthesized during a 5 hr heat shock were considerably more abundant than those normally produced from the tk promoter. At low temperature the only transcripts detectable were initiated upstream from the fusion point, and thus mapped to the point of divergence of the S1 probe from the template. A further construction (pF2) containing only residues −10 to −186 of the heat-shock gene was equally inducible, although the level of readthrough transcripts was higher (Figure 3). Since tk mRNA is clearly stable at low temperature, this result implies that the principal control of hsp 70 gene expression in COS cells is at the level of transcription initiation, and that the promoter and regulatory sequences are contained within the −10 to −186 5′ flanking region.

A third fusion (pF3, see Figures 1 and 5) contains the entire tk gene including the TATA box and surrounding sequences, out to a Hha I site at position −50, but lacks the important upstream promoter region which lies between residues −52 and −109 (McKnight et al., 1981). This gene is joined to 5′ flanking sequences from the hsp 70 gene, extending from one of the overlapping Hha I sites at positions −64/−66 out to position −400. The resultant hybrid gene was transcribed extremely poorly in COS cells at both high and low temperatures (Figure 3). Prolonged exposure of the autoradiogram revealed very faint bands corresponding to the normal tk transcripts (including the upstream starts between −30 and −50), but these were completely unaffected by heat shock. Thus the sequences upstream of −63 in the hsp 70 gene are not sufficient to replace functionally the upstream region of the tk promoter at either 37°C or 42°C.

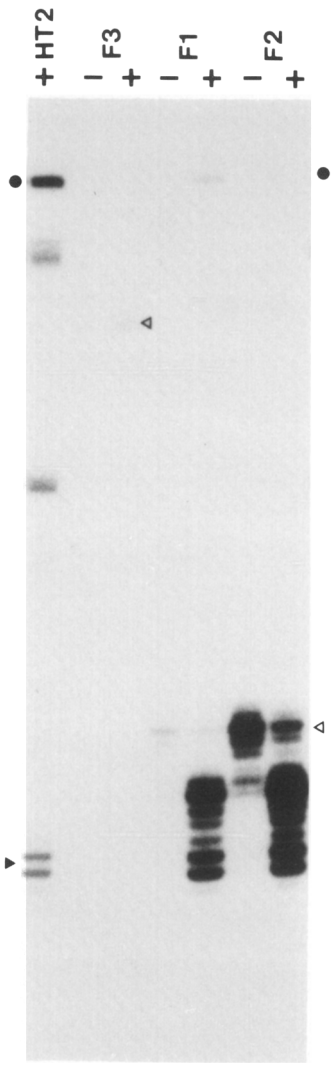

Figure 3. 5′ S1 Mapping of Transcripts from the tk Gene Fused to the hsp 70 Promoter

Solid circles indicate the position of intact probe; open triangles indicate the position corresponding to upstream transcripts mapping to the point of sequence divergence between probe and template. Solid triangles indicate fragments corresponding to initiation at the correct tk sequence. Cells were transfected with pF1, pF2 or pF3, and either heat-shocked (+) or not (−). A sample from Figure 2 (pHT2) is included for reference. The sequence at the junction of hsp 70 and tk DNA in pF1 and pF2 is ... TATAAATAGAGGCGCTTCGTCGACCCCTCGAACA*CCGAG ..., where the underlined bases are derived from the tk gene and the asterisked A residue probably corresponds to the first base of authentic tk mRNA. The junction in pF3 is described in the text.

Deletion Mapping the Heat-Shock Promoter

To locate a DNA sequence essential for the heat-shock response, sequential 5′ deletions were prepared with Bal 31 nuclease, synthetic Hind III linkers ligated to the remaining DNA, and the plasmid reconstructed as shown in Figure 1 (pHTΔ5′ series). These plasmids were introduced into COS cells, and the hsp 70 transcripts were detected by 5′ S1 mapping. Figure 4 shows that deletion to position −186, −108, −97, −68 or −66 did not prevent the normal heat induction. A deletion to position −44 gave a very low level of transcripts after heat shock, comparable to the normal low temperature level, while a deletion to position −28, within the TATA box, gave no detectable correct 5′ ends at all.

To confirm a consistent efficiency of transfection and RNA recovery, each RNA sample was also hybridized to the tk probe. As expected, the level of tk transcripts was similar in all cases (examples are shown in Figure 4). A further internal control is provided by transcripts that originate upstream of the hsp 70 promoter. With the 5′ S1 probe, these map to the point of sequence divergence between probe and template (Figure 4). The abundance of these transcripts was not greatly affected by heat shock, but a slight induction (about 2-fold) was sometimes observed even though the concentration of tk transcripts in the same samples remained constant. The mechanism of this apparent induction may be posttranscriptional; it is clearly distinct from the true transcriptional induction. Thus even though mutants such as Δ5′-44 show a very weak apparent induction (Figure 4), I have considered them uninducible because the ratio of correct to upstream starts is not increased by heat shock.

The construction of deletion mutants places novel DNA sequences adjacent to the promoter, and it is possible that these might substitute in some way for the deleted sequences. In these constructions I have used a DNA fragment from the spacer region of Xenopus 5S DNA in the hope that this would be completely irrelevant to a polymerase II promoter. As a control, however, the remaining hsp 70 sequences in Δ5′-66 were also joined to another DNA fragment from 200 to 700 bp upstream of the tk gene (see Experimental Procedures). This sequence has no effect on tk transcription in Xenopus oocytes (McKnight et al., 1981). The resultant plasmid, pHTΔ5′-66*, behaved identically to pHTΔ5′-66 (Figure 4). Thus induction of the hsp 70 promoter in COS cells requires DNA sequences between −44 and −66, but sequences farther upstream from this appear completely dispensable.

Further Characterization of the Upstream Promoter Element

Three additional mutants help to define the important sequences between −44 and −66. Their sequences

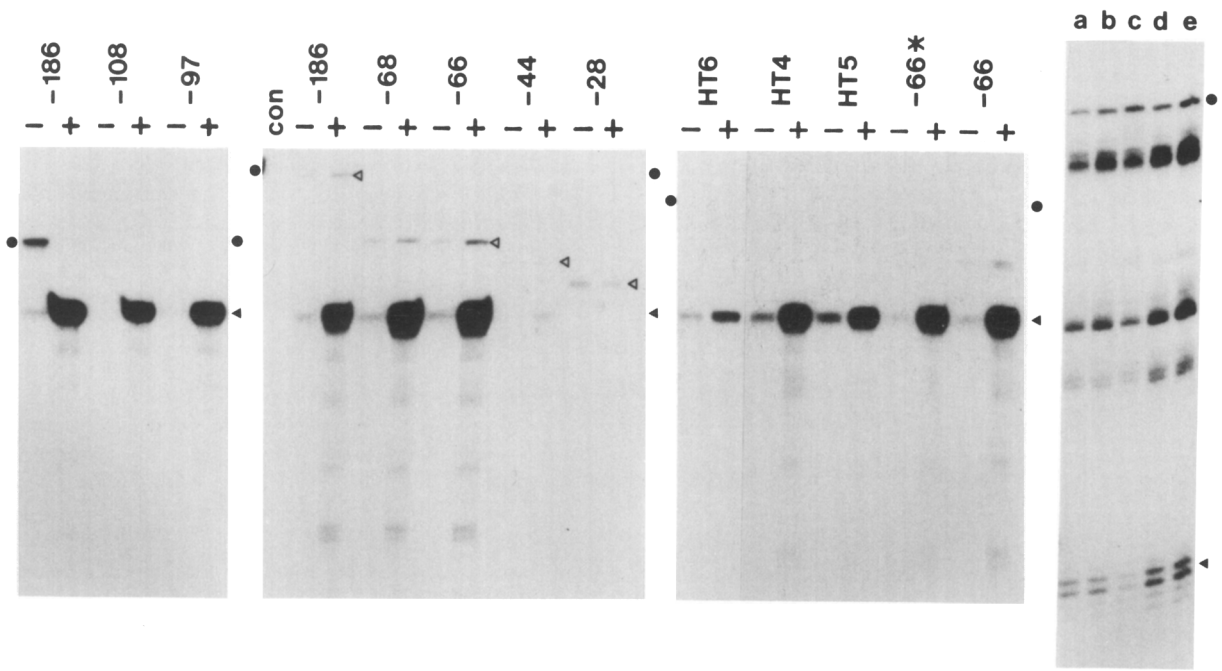

Figure 4. 5′ S1 Mapping of Transcripts from Deletion Mutants of the hsp 70 Gene

Symbols have the same meaning as those in Figure 3. Parts of four separate gels are shown. The plasmids tested were members of the pHTΔ5′ series, and also pHT4 (a control), pHT5 and pHT6. The critical sequences of these are shown in Figure 5. See text for an explanation of plasmid pHTΔ5′-66*. Control RNA was from heat-shocked mock-transfected cells. Tracks a, b, c, d and e show mapping of the tk transcripts in the RNA samples derived from heat-shocked cells that had been transfected with pHTΔ5′-186, -68, -66, -44 and -28, respectively.

in this region are shown in Figure 5 (see also Figure 1). One, pHT5, has residues −45 through −50 replaced by a decameric Hind III linker. This introduces four base changes and moves the −47 to −66 region four bases farther from the TATA box. Despite these changes, pHT5 was still clearly heat-inducible (Figure 4), although its activity was somewhat reduced; in duplicate experiments the abundance of induced transcripts varied from 40% to more than 70% of the control, pHT4 (see Figure 1). Comparison with pHTΔ5′-66 suggests that at most 18 bp in the region upstream of −45 are essential (Figure 5), and the distance of this sequence from sequences downstream of −45 is not critical.

Two other deletions were obtained after S1 digestion of Hind III–cut pHT5 and were identified by DNA sequencing. In one, pHT6, residues −45 through −69 were replaced by a single C residue, although the upstream sequences show some homology to those deleted (Figure 5). This plasmid produced very few transcripts after heat shock (Figure 4). Although it showed an apparent low level of induction (about 3-fold), the upstream transcripts also appeared inducible (not visible in Figure 4), and as discussed earlier this may not reflect a true transcriptional induction. The second mutant, pHT7, lacks a further 2 bp (Figure 5), and had equally low activity (data not shown).

These results confirm the conclusion derived from the inactivity of pF3 that hsp 70 sequences upstream of −66 are themselves unable to stimulate normal levels of transcription from an intact TATA box region. They also suggest that at least some of the 9 bases in the −47 to −66 region that differ between pHT5 and pHT6 are essential for efficient promoter function.

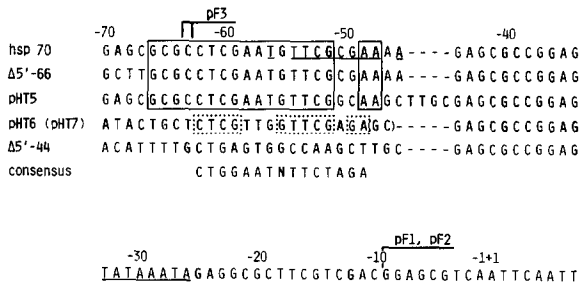

Figure 5. Sequences of the 5′ Flanking Region of the hsp 70 Gene and Some Mutants

The basic structures of the mutant plasmids are shown in Figure 1. The sequence shown is that of the mRNA sense strand from the gene derived from plasmid 56H8 (Karch et al., 1981). The corresponding region of the gene from plasmid 132E3 differs only at positions −3 (A) and −12 (T). Transcription is initiated at position +1; the TATA box and the inverted repeat at position −50 are underlined. The points of fusion with the tk gene in pF1, pF2 and pF3 are indicated above the sequence. Boxes enclose those sequences that are present in all the genes that show good heat-inducibility but are absent or disrupted in those that do not (see Figure 4). Dotted boxes indicate the bases in pHT6 that are homologous to the parent gene in this region. For an explanation of the consensus sequence see Figure 7 and the Discussion.

Forced Expression of the Hsp 70 Gene at Low Temperature

A reasonable interpretation of the data presented above is that the hsp 70 gene has a perfectly normal TATA box and transcription initiation site, but has an upstream region that is only functional in heat-shocked cells. A prediction from this would be that expression of the gene at low temperature could be achieved by provision of a functional upstream element, such as that contained in the SV40 promoter region.

Figure 2 shows that the presence of SV40 sequences including the 72 bp repeat at the 3′ end of the hsp 70 gene (in pHT2) does not affect its expression; a similar result is shown in Figure 6 with a plasmid (pH1) containing a different hsp 70 gene and lacking the tk gene. However, a completely different result was obtained with plasmid pH2, which has the SV40 origin region on the 5′ side of the gene, with the late side of the 72 bp repeat some 200 bases from the start site (see Figure 1). This plasmid gave a substantial level of hsp 70 transcripts at 37°C (Figure 6), the total abundance of these transcripts being comparable to that obtained from other plasmids after heat shock. However, mapping of the transcripts shows that while many of them were initiated at the normal site, there was also considerable use of upstream start sites (Figure 6). Such upstream starts are detectable at an extremely low level with plasmids that lack the 72 bp repeat, but they are not heat-inducible. The preferential use of these upstream start sites in cells transfected with plasmid pH2 suggests that the SV40 sequences may act as a polymerase "entry site" in this plasmid (see Discussion).

Although plasmid pH2 retains all sequences normally required for heat induction, the presence of the SV40 sequences upstream apparently prevented this response. Superimposition of the normal heat-shock response on the pattern of transcripts obtained at 37°C would have been readily detected; in fact there was no change in either the absolute level of transcripts or in the ratios of normal to upstream starts after heat shock (Figure 6). This suggests that the SV40 seqence can have a repressive effect, perhaps via chromatin structure, at least 150 bases away. Alternatively, the SV40 sequences may simply remain active at high temperature, thus not actually repressing the hsp 70 promoter element, but making it functionally redundant.

From these results, I conclude that hsp 70 mRNA is not exceptionally unstable in cells at 37°C; that the TATA box region is capable of efficiently directing polymerase initiation events at low temperature; and that a promoter function analogous to that which is regulated by heat shock can be provided at low tem-

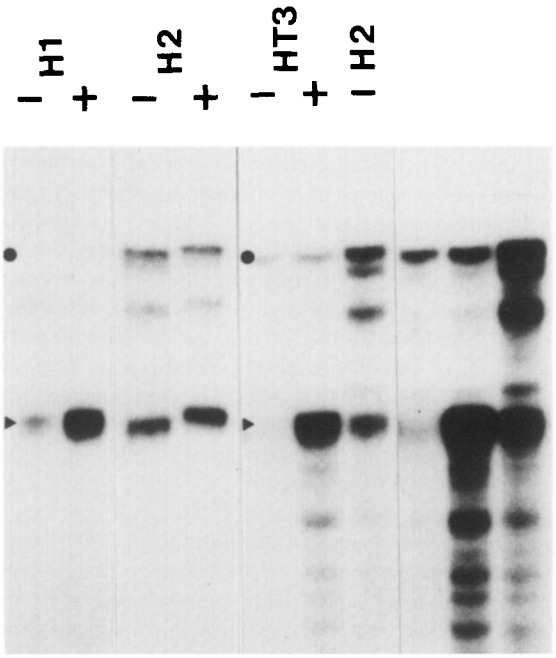

Figure 6. Effect of SV40 Sequences on hsp 70 Transcription

Transcripts were detected by 5′ S1 mapping, and symbols have the same meaning as in Figure 3. The SV40 72 bp repeat is on the 3′ side of the gene in plasmid pH1, on the 5′ side in plasmid pH2, and absent in plasmid pHT3 (see Figure 1). Results from two different experiments with plasmid pH2 are shown. The three tracks on the right are a longer exposure of the adjacent three tracks; the bands below the normal ~400 bp band reflect slight degradation of the RNA in this experiment.

perature by a piece of SV40 DNA that includes the 72 bp repeat.

Discussion

Functional Domains of the Hsp 70 Promoter

Sequences upstream of the hsp 70 coding region promote efficient transcription of the gene in monkey cells only when the cells are exposed to heat shock. This regulation is maintained despite the heterologous nature of the assay, with the gene present on a multicopy replicating plasmid, and whether or not it is adjacent to a gene (tk) that is active at low temperature. The results presented place limits on the sequences responsible for this function, and they suggest that the response to heat shock is an intrinsic property of the promoter itself.

The functional analysis of fusions between the hsp 70 and tk genes shows that heat-inducible promotion does not require any of the transcribed hsp 70 sequences, nor the first 10 bases of the 5′ flanking region. Analysis of 5′ deletion mutants indicates that sequences upstream of position −66 have no detectable function in COS cells. Thus the region −10 to −66 appears sufficient for regulated promotion of transcription. This region can be divided into two functional domains on the basis of this work and by analogy with other systems. The first of these consists of the TATA box and succeeding GC-rich sequence. The second is an element upstream of the TATA box that is required for efficient promotion in vivo.

The TATA Box Region

There is nothing particularly unusual about the TATA box region of the hsp 70 and other heat-shock genes. All the genes show homology to the sequence (A/G)C(C/A)GGCGC immediately following the TATA box, but many polymerase II promoters have a similar region. For example, equally strong homology to this sequence is found in the Drosophila histone H4 and alcohol dehydrogenase genes (Goldberg, 1979; H. Haymerle, personal communication).

Evidence from a number of systems indicates that the function of the TATA box region in vivo is to direct the polymerase to initiate transcription at a discrete site approximately 30 bases downstream from the TATA box itself (Grosschedl and Birnstiel, 1980a; Benoist and Chambon, 1981; Grosveld et al., 1982). It seems that the hsp 70 TATA box can perform this function equally well in normal and heat-shocked cells. First, the residual accurate transcription from the deletion mutant pHTΔ5′-44 is significantly higher than that from pHTΔ5′-28, which lacks the TATA box, but this transcription is not greatly affected by heat shock. The TATA box also directs a significant fraction of polymerases to the correct start site when transcription is forced at low temperature by SV40 sequences (Figure 6). Furthermore, this region directs initiation in whole-cell extracts prepared from non-heat-shocked HeLa cells with an efficiency comparable to that of other promoters, whether or not the upstream region is present (P. Farrell and H. Pelham, unpublished observations). These results suggest that the TATA box region plays a passive, nonregulatory role in the transcription of the hsp 70 gene.

The Upstream Promoter Element

The upstream element appears to lie between residues −47 and −66 (see Figure 5). Mutations that affect this region have a profound effect on promoter activity in heat-shocked cells. The element can be separated from the TATA box by a further 4 base pairs without abolishing promoter function.

Since the heat shock genes probably share a common induction mechanism, I have looked for similarities between the sequence of the hsp 70 upstream element and the corresponding position in other Drosophila heat-shock genes. An obvious feature of the hsp 70 sequence is the inverted repeat centered at position −51/−50. Larger inverted repeats are present in approximately the same position in 5 of the 6 other heat-shock genes, and when these repeats are aligned, significant sequence homology between the genes is apparent (Figure 7). This homology is largely

restricted to positions −48 through −62 on the hsp 70 sequence, within the region containing the functional element, and for these bases a consensus sequence CTGGAAT(N)TTCTAGA can be derived. The hsp 70 gene fits this consensus at 11 of the 14 positions; of these, three are altered in pHT6, which is at best only weakly inducible (see Results).

Ten of the 14 bases in the consensus sequence themselves form an inverted repeat (CT-GAA--TTC-AG), a property characteristic of many protein recognition sites on DNA. This inverted repeat is, however, offset from the center of the large inverted repeats; many of the bases in the latter do not match the consensus sequence or are outside it (Figure 7). It seems that there has been selection for two independent features: a sequence similar to the consensus, and a larger inverted repeat at the appropriate position nearby, regardless of the exact sequence of the latter. One can speculate that the larger inverted repeat might be a general feature that aids recognition of the functional sequence by a protein (for example, by influencing the position of a nucleosome). Alternatively, it might facilitate the function of such a protein, for example, by forming loops that would make a local unwinding of the DNA helix more energetically favorable. However, the activity of pHT5, albeit lower than that of the parental plasmid, indicates that perturbation of the inverted repeat in the hsp 70 gene can at least be tolerated.

Not all the heat-shock genes fit the general pattern closely. The hsp 68 gene fits the consensus sequence poorly, although a better homology exists on the 3′ side of the inverted repeat (Figure 7). The hsp 23 gene also does not fit the consensus in this region, although the homologous sequence CGA-GAAGTTTCGTG is present 100 bases upstream from the TATA box. It will be interesting to locate the functional sequences in this gene.

Despite these exceptions, the general correlation between the positions of the homologous sequences, the inverted repeats, and the hsp 70 functional region suggests that the consensus sequence and the inverted repeat together may constitute a functional domain. The variable distance between this region and the TATA box in the various genes is consistent with the flexibility of this distance in the hsp 70 gene (as exemplified by pHT5), and suggests that the upstream region is a physically independent functional unit. This model can be tested by mutational analysis of other heat-shock promoters.

Function of the Upstream Element

The upstream element of the hsp 70 promoter seems to be the site at which regulation of transcription occurs. The regulation is positive, in the sense that deletions in this region greatly reduce transcription during heat shock rather than causing constitutive expression of the gene. This can explain how control can be maintained despite the presence of the gene in thousands of copies per cell: there may be no specific repressor to be titrated out.

The role of the upstream region in heat-shocked cells appears to be analogous to that of the upstream regions of other polymerase II promoters, namely, to stimulate use of the TATA box region. This is also suggested by the observation that its function can be mimicked at low temperature by SV40 sequences containing the 72 bp repeat. When these are on the

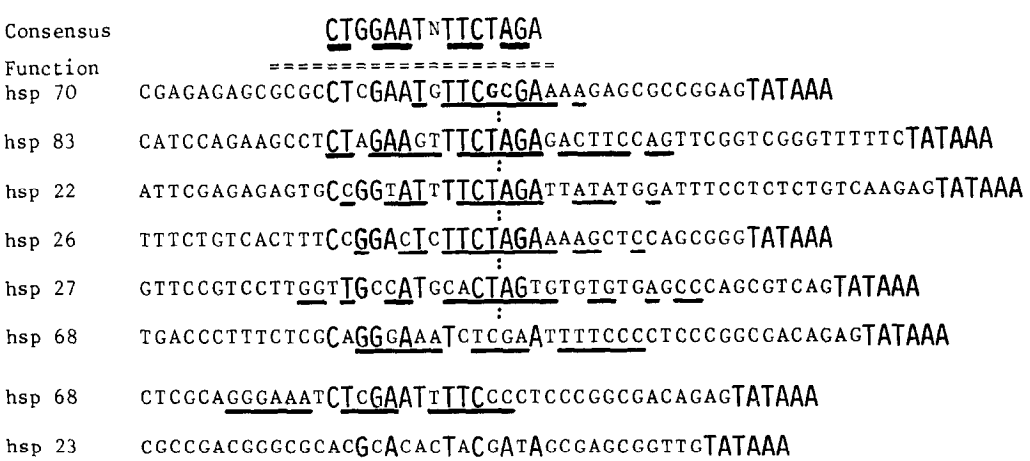

Figure 7. Homologies between the Drosophila Heat-Shock Genes in the Upstream Region

Sequences are from Karch et al. (1981), Holmgren et al. (1981) and Ingolia and Craig (1981). The sequences of the genes for hsp 70, hsp 83, hsp 22, hsp 26, hsp 27 and hsp 68 were aligned at the center of the major inverted repeats (underlined), and the consensus sequence was derived for the homologous region. The inverted repeat in the consensus sequence is also underlined. Dots indicate the dyad axis of the major inverted repeats. Lines above the hsp 70 sequence indicate the region that is essential for heat induction of this gene (see Figure 5). The bottom two sequences show an alignment of the hsp 68 gene that matches the consensus sequence more closely, and the best match that can be obtained with the hsp 23 gene in this region. There is no corresponding inverted repeat in the hsp 23 gene, and the significance of the weak homology is doubtful. The second base shown in the hsp 70 sequence is a G, not a C as reported previously (Karch et al., 1981).

5′ side of the gene, polymerase molecules tend to initiate transcription at preferred sites close to the SV40 sequence, but many of them initiate some 200 bp farther along the DNA, downstream from the TATA box. The ability of the SV40 72 bp repeat to induce initiation in such a manner supports the suggestion of Moreau et al. (1981) that it acts as an entry site in the chromatin from which polymerase II can search for a TATA box or preferred start site. It is not clear whether this activity is the same as the enhancer activity associated with the SV40 72 bp repeat. In assays that do not involve high-level replication of the DNA the enhancer sequences greatly stimulate (200-fold) transcription from normal, intact promoters in a way that is cis-acting but apparently independent of the distance of the enhancer from the promoter, and of its orientation (Banerji et al., 1981). Such an effect of the 72 bp repeat was not observed in the COS cell system with either the hsp 70 gene or the tk gene, presumably because the high copy number of the genes in these cells somehow makes enhancer sequences redundant.

The results presented here are compatible with a model in which the hsp 70 upstream element itself functions as a regulated entry site. It is not clear what such a site might be in physical terms, but one of its properties could be to organize the local chromatin structure such that an appropriate region of the DNA is exposed. The SV40 origin and 72 bp repeat region is evidently capable of organizing nucleosomes such that the 72 bp repeat is exposed (see Jakobovits et al., 1980; Saragosti et al., 1980, and references therein) and such forced organization might also affect the function of the hsp 70 upstream region when this is nearby. The hsp 70 genes in Drosophila cells also have an organized chromatin structure: there is a DNAase I–hypersensitive site that has been mapped to a diffuse region covering about 200 bp upstream from the transcription initiation site (Wu, 1980). This site is present before heat shock, and it has been suggested that it is a necessary but not sufficient prerequisite for transcription (Elgin, 1981). If this is true in the COS cell system, then any DNA sequences required for chromatin organization must also be present in the −10 to −66 region, and the upstream element becomes a candidate for a chromatin organizer at low temperature as well as a promoter element at high temperature. Analysis of the chromatin structure of mutant plasmids in COS cells should clarify this point.

COS cells are a heterologous environment for the Drosophila hsp 70 gene, and they respond to heat shock at a different temperature from flies (which are fully heat-shocked at 37°C). Also, the replicating plasmid system is a somewhat abnormal state for a cellular gene. DNA sequences (such as enhancers) that facilitate gene expression in the homologous chromosome may appear functionless when assayed in this way. It is thus possible that DNA sequences other than those described in this paper also contribute to the regulation of hsp 70 gene expression in Drosophila cells. However, it is not necessary to postulate the existence of such sequences to account for the heat-inducibility of the gene.

Finally, it may be useful to compare the upstream region of the hsp 70 promoter with the so-called CCAAT box. This is a conserved sequence found about 80 bp upstream from the start site of all globin genes and also some other genes (Efstratiadis et al., 1980). There is good evidence in the globin system that the CCAAT box is a functional upstream element (Dierks et al., 1981; Grosveld et al., 1982; Mellon et al., 1981). Despite this functional analogy, the CCAAT box differs from the hsp 70 upstream region in several respects: it is farther away from the TATA box, is not generally associated with an inverted repeat, and its activity is constitutive in COS cells. It remains to be seen whether sequences with properties similar to those of the hsp 70 upstream element are a common feature of inducible genes.

Experimental Procedures

Plasmid Constructions

All plasmids were grown in Escherichia coli HB101. The basic structures of the plasmids are shown in Figure 1 and details of some of them are shown in Figure 5. Figure 8C outlines the construction of pHT1. The starting plasmid was pMLRIIG (Lusky and Botchan, 1981). This is a pBR322 derivative (pML) in which bases 1092–2485 are deleted, with the RIIG fragment of SV40 (see Figure 8A) inserted into the Eco RI site via synthetic Eco RI linkers, the late side of the SV40 insert being close to the ampicillin resistance gene. This was digested with Bam HI and Hind III, and the herpes tk gene was inserted to form pXTK1. The tk gene was derived from the plasmid ptk/Δ3′-1.13 sequenced by McKnight (1980). The Hind III site in pXTK1 was removed by filling in and religation, and then a decameric Hind III linker was inserted at the Pvu II site present 200 bp upstream of the transcription initiation site in the tk gene, to form pXTK10 (Figure 8C). This plasmid still retains the origin fragment from SV40 (Figure 8A). The hsp 70 genes were obtained from the plasmids 56H8 and 132E3 (Figure 8B), large portions of which have been sequenced by Torok and Karch (1980) and Karch et al. (1981). The Bgl II fragment containing the left-hand gene of 132E3 was inserted into the Bam HI site of pXTK10 to form pHT1 (Figure 8C). This fragment contains about 1100 bases of 5′ flanking sequence and about 70 bases of 3′ flanking sequence (Karch et al., 1981).

Other plasmids are diagrammed in Figure 1. pHT2 was constructed by first replacing the Hind III–Bam HI fragment of pXTK10 with the origin + 72 bp rpt fragment of SV40 (see Figure 8A; the SV40 Pvu II site was converted to a Bam HI site with a CCGGATCCGG linker) and then inserting the hsp 70 Bgl II fragment. pHT3 was derived from pHT1 by deletion of the Sal I–Xho I fragment on the 5′ side of the hsp 70 gene (Figure 1); it retains 194 bp of 5′ flanking sequence. pHT4 was formed by substitution of the Xho I–Bam HI fragment of pHT1 with the corresponding fragment from 56H8 (Figure 8B).

The pHTΔ5′ series was prepared as follows: 10 μg of Xho I–cut pHT4 DNA was incubated at 30°C with 3 units of Bal 31 nuclease (BRL) in the recommended buffer, and samples were taken after 15, 40, 65 and 90 seconds. Digestion was terminated by phenol extraction. DNA from the samples was pooled, ligated to Hind III linkers (GCAAGCTTGC) and digested with Hind III and Bgl II (which cuts within the tk gene). The Hind III–Bgl II gene-containing fragment was isolated on an agarose gel and recloned between the Hind III and Bgl II sites of plasmid pXTK12. This plasmid is similar to pXTK10 (Figure

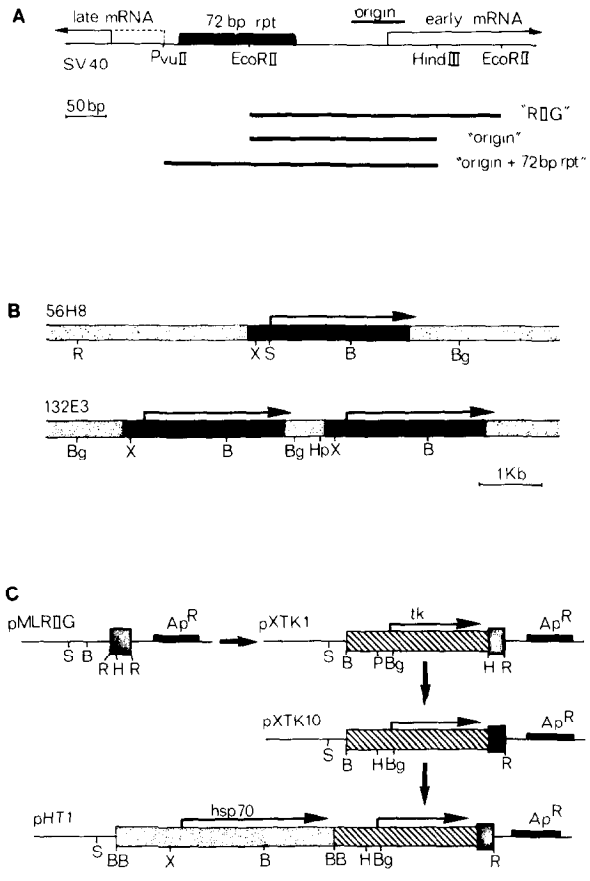

Figure 8. Origins of DNA Fragments Used in Plasmid Construction

(A) The SV40 origin region, showing the 72 bp repeats and the early and late promoters. Late transcripts have very heterogeneous 5′ ends. The fragments used are indicated and have the following coordinates (SV numbering system; see Tooze, 1981): RIIG, 160–5092; origin, 160–5171; origin + 72 bp rpt, 270–5171. (B) Relevant segments of Drosophila DNA from plasmids 56H8 and 132E3. Black boxes indicate the sequences that are strongly conserved between all three hsp 70 gene copies shown. Arrows above these indicate transcripts. (C) Steps in the construction of pHT1. Plasmids are shown arbitrarily linearized, and symbols have the same meaning as in Figure 1. For details see Experimental Procedures. Selected restriction sites are indicated in B and C, and are abbreviated as follows: B = Bam HI; Bg = Bgl II; BB = Bam HI joined to Bgl II; H = Hind III; Hp = Hpa I; P = Pvu II; R = Eco RI; S = Sal I; X = Xho I.

8C), except that the Hind III–Bam HI fragment has been replaced by a 360 bp Hind III–Hpa II fragment containing spacer sequences from a Xenopus 5S DNA clone (pXIs 11; for sequence see Peterson et al., 1980), modified by ligation of a 10 base Bam HI linker (BRL) to the filled-in Hpa II site. The final construction (Figure 1) is thus the same as pHT1, but with the sequences upstream of the hsp 70 gene deleted and replaced by the 5S spacer DNA. pHTΔ5′-66* has this DNA replaced by the Bam HI–Hind III fragment from upstream of the tk gene in pXTK10 (Figure 8C). Deletion mutants were mapped by accurate sizing of small (<50 bp) restriction fragments on sequencing gels, and the end points of pHTΔ5′-68, -66, -44 and -28 were confirmed by partial DNA sequencing (Maxam and Gilbert, 1977). pHTΔ5′-186 was prepared by ligating linkers directly to the filled-in Xho I site. The nomenclature of these mutants indicates the last base of the gene sequence remaining, even if this is in fact derived from the Hind III linker.

pHT5 was derived from pHT1 by replacing the Xho I–Bam HI fragment of the hsp 70 gene with two fragments: the Xho I–Nru I fragment from positions −190 to −50 of the hsp 70 gene, with a Hind III linker added to the Nru I end, and the Hind III–Bam HI fragment from pHTΔ5′-44. The junction sequence is shown in Figure 5. pHT6 and pHT7 were derived from pHT5 by digestion with Hind III, treatment with S1 nuclease and transformation of the linear molecules into E. coli HB101. Sequencing of the resultant clones showed two classes of deletion; the most common was that in pHT7, and is effectively the result of recombination between the GAGCGC sequences at −44 and −70 (see Figure 5).

pF1 contains the Eco RI–Sal I fragment from the 5′ side of the hsp 70 gene in 56H8 (Figure 8B) joined via an 8 base Sal I linker to the Hae III site 8 bp upstream from the transcription start site of the tk gene. The Sal I site is at position −12 of the hsp 70 gene; the junction sequence is given in the legend to Figure 3. The tk gene is joined to the SV40 origin fragment (Figure 8A) as in pHT1, and this tripartite piece of DNA is inserted at the Eco RI site of pML (see Figure 1). pF2 contains a similar insert, but the hsp 70 sequences extend only from the Sal I site to the Xho I site at −186, which is fused to the Sal I site of pML. The plasmid also contains a second SV40 origin fragment and a piece of the tk gene that extends from the Bam site upstream of the gene to a Hind III linker positioned at the start site of transcription (Figure 1). This fragment was derived from a deletion mutant constructed by S. McKnight, and it lacks all the transcribed sequences of the tk gene. pF2 was originally constructed for a different purpose, and the presence of this extra tk sequence is in fact irrelevant to the results presented in this paper.

The hsp 70 sequences in pF3 are derived from the right-hand gene in 132E3 (Figure 8B). They extend from the Hpa I site (converted to a Sal I site via an 8-base linker) about 400 bases upstream of the gene to one of the two overlapping Hha I sites at position −65 (see Figure 5). This sequence is very similar to that of the left-hand hsp 70 gene of 132E3, and is identical to it downstream of the Xho I site. The Sal I (Hpa I) site is joined to the Sal I site in pML, and the Hha I site is joined to the Hha I site at position −50 of the tk gene. The rest of the plasmid is identical to pXTK1 (Figure 8C).

pH1 is derived from pML by insertion of Drosophila and SV40 sequences between the Eco RI and Hind III sites (Figure 1). The Drosophila DNA consists of the Eco RI–Bgl II fragment from 56H8 (Figure 8B), which is joined to the late side of the origin + 72 bp rpt fragment of SV40 (Figure 8A), the SV40 Pvu II site having been converted to a Bam HI site by ligation of a linker (CCGGATCCGG). pH2 is a Bam HI–Hind III insert in pML (Figure 1). The Pvu II site on the SV40 origin + 72 bp rpt fragment was converted to a Sal I site by ligation of a linker (GGTCGACC). This was joined to the Xho I–Bgl II fragment containing the left-hand hsp 70 gene from 132E3 (Figure 8B). The Xho I site is at position −195 of this hsp 70 gene.

Transfection of COS Cells

Plasmid DNA was prepared by the rapid alkaline lysis method described by Ish-Horowicz and Burke (1981). High molecular weight RNA was removed by precipitation with 2.5 M LiCl, and the DNA was purified further by phenol extraction. Such DNA was as efficient at transfection as CsCl-gradient-purified DNA.

COS 7 cells (Gluzman, 1981) were maintained in Dulbecco's MEM with 10% calf serum, and transfected when slightly subconfluent. Each 75 cm^2 flask was exposed to 2–5 μg of DNA in 3 ml of medium (without serum) containing 200 μg/ml DEAE-dextran for 8 hr as described by Sompayrac and Danna (1981). Similar results were obtained with 30 min exposure to DNA in 2 ml containing 500 μg/ml DEAE-dextran. The DNA solution was then replaced with fresh medium (with serum) and the cells maintained at 37°C. For heat shock, the flasks were placed in a water bath at 42–43°C from 30 to 35 hr after transfection. Cells were then trypsinized, spun down and lysed on ice in 140 mM NaCl, 5 mM KCl, 3 mM $MgCl_2$, 25 mM Tris-HCl (pH 7.5), 1% NP40. Nuclei were removed by centrifugation, and RNA was prepared from the supernatant by phenol extraction in the presence of 1% SDS, followed by isopropanol precipitation.

S1 Mapping

For tk transcripts, the probe was the 131 bp Eco RI–Bgl II fragment that spans the transcription start (McKnight, 1980), labeled at the Bgl II site with polynucleotide kinase before Eco RI digestion, and purified by agarose gel electrophoresis. For the 5′ end of the hsp 70 transcripts, probe was prepared from 132E3 or 56H8. One microgram of plasmid DNA was digested with Xho I in 33 mM Tris–acetate (pH 8), 66 mM KOAc, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mg/ml gelatin. Five units of T4 DNA polymerase (BRL) were added and incubation continued for 30 min at 37°C. Then 40 μCi α-^{32}P-dATP (400 Ci/mmole) and 100 μM each of dCTP, dTTP and dGTP were added, and 30 min later 250 μM dATP was added. After a 30 min chase, Hae III was added. Finally, the labeled 568 bp (56H8) or 573 bp (132E3) fragment was isolated by agarose gel electrophoresis. Digestion of this with other restriction enzymes showed that it was uniformly labeled by this procedure.

A similar method was used to prepare 3′ probe from 132E3. In this case the initial digestion was with Bam HI, the Hae III digestion was omitted and the entire plasmid was used as probe after phenol extraction.

One tenth of the total cytoplasmic RNA from a flask of cells (about 10 μg) was mixed with the appropriate probe in 20 μl of 80% formamide, 10 mM PIPES (pH 6.5), 1 mM EDTA, 0.4 M NaCl, heated to 85°C for 5 min and hybridized overnight at 50°C. The hybrids were diluted with 200 μl of S1 buffer (225 mM NaCl, 75 mM $Na_{0.5}H_{0.5}OAc$ (pH 4.5), 0.7 mM $ZnSO_4$), digested with 100–200 units of S1 nuclease (BRL) for 10 min at 37°C or 30 min at 22°C, mixed with 50 μl 5× stop buffer (100 mM EDTA, 0.4 M Tris, pH 9.5), ethanol-precipitated and analyzed on sequencing gels.

Acknowledgments

I thank S. McKnight and M. Botchan for providing plasmids and J. Gurdon, A. Travers, R. Laskey and M. Bienz for helpful comments on the manuscript.

Received May 7, 1982; revised June 17, 1982

References

Ashburner, M. and Bonner, J. J. (1979). The induction of gene activity in Drosophila by heat shock. Cell *17*, 241–254.

Banerji, J., Rusconi, S. and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Benoist, C. and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature *290*, 304–310.

Berk, A. J. and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell *12*, 721–732.

Corces, V., Pellicer, A., Axel, R. and Meselson, M. (1981). Integration, transcription, and control of a Drosophila heat shock gene in mouse cells. Proc. Nat. Acad. Sci. USA *78*, 7038–7042.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P. and Uedinger, L. (1980). Promoter sequences of eucaryotic protein-coding genes. Science *209*, 1406–1414.

Dierks, P., van Ooyen, A., Mantei, N. and Weissman, C. (1981). DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin mRNA with the correct 5′ terminus. Proc. Nat. Acad. Sci. USA *78*, 1411–1415.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. and Proudfoot, N. J. (1980). The structure and evolution of the human β-globin gene family. Cell *21*, 653–668.

Elgin, S. C. R. (1981). DNAase I-hypersensitive sites of chromatin. Cell *27*, 413–415.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell *23*, 175–182.

Goldberg, M. (1979). Sequence analysis of Drosophila histone genes. Ph.D. thesis, Stanford University, Stanford, California.

Grosschedl, R. and Birnstiel, M. L. (1980a). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Nat. Acad. Sci. USA *77*, 1432–1436.

Grosschedl, R. and Birnstiel, M. L. (1980b). Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequences are essential for promotion of H2A histone gene transcription in vivo. Proc. Nat. Acad. Sci. USA *77*, 7102–7106.

Grosveld, G. C., de Boer, E., Shewmaker, C. K. and Flavell, R. A. (1982). DNA sequences necessary for transcription of the rabbit β-globin gene in vivo. Nature *295*, 120–126.

Grosveld, G. C., Shewmaker, C. K., Jat, P. and Flavell, R. A. (1981). Localization of DNA sequences necessary for transcription of the rabbit β-globin gene in vitro. Cell *25*, 215–226.

Holmgren, R., Corces, V., Morimoto, R., Blackman, R. and Meselson, M. (1981). Sequence homologies in the 5′ regions of four Drosophila heat-shock genes. Proc. Nat. Acad. Sci. USA *78*, 3775–3778.

Ingolia, T. D. and Craig, E. A. (1981). Primary sequence of the 5′ flanking region of the Drosophila heat shock genes in chromosome subdivision 67B. Nucl. Acids Res. *9*, 1627–1642.

Ingolia, T. D., Craig, E. A. and McCarthy, B. J. (1980). Sequence of three copies of the gene for the major Drosophila heat shock induced protein and their flanking regions. Cell *21*, 669–679.

Ish-Horowicz, D. and Burke, J. F. (1981). Rapid and efficient cosmid cloning. Nucl. Acids Res. *9*, 2989–2998.

Jakobovits, E. B., Bratosin, S. and Aloni, Y. (1980). A nucleosome-free region in SV40 minichromosomes. Nature *285*, 263–265.

Karch, F., Torok, I. and Tissières, A. (1981). Extensive regions of homology in front of the two hsp 70 heat shock variant genes in Drosophila melanogaster. J. Mol. Biol. *148*, 219–230.

Lusky, M. and Botchan, M. (1981). Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. Nature *293*, 79–81.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA *74*, 560–564.

McKnight, S. L. (1980). The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucl. Acids Res. *8*, 5949–5964.

McKnight, S. L., Gavis, E. R., Kingsbury, R. and Axel, R. (1981). Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell *25*, 385–398.

Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981). Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host–vector system. Cell *27*, 279–288.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P. and Chambon, P. (1981). The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucl. Acids Res. *9*, 6047–6068.

Peterson, R. C., Doering, J. L. and Brown, D. D. (1980). Characterization of two Xenopus somatic 5S DNAs and one minor oocyte-specific 5S DNA. Cell *20*, 131–141.

Saragosti, S., Moyne, G. and Yaniv, M. (1980). Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. Cell *20*, 65–73.

Sompayrac, L. M. and Danna, K. J. (1981). Efficient infection of monkey cells with DNA of simian virus 40. Proc. Nat. Acad. Sci. USA *78*, 7575–7578.

Tooze, J., ed. (1981). DNA tumour viruses. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Torok, I. and Karch, F. (1980). Nucleotide sequences of heat shock activated genes in Drosophila melanogaster. I. Sequences in the regions of the 5′ and 3′ ends of the hsp 70 gene in the hybrid plasmid

56H8. Nucl. Acids Res. *8*, 3105–3123.

Udvardy, A., Sumegi, J., Toth, E. C., Gausz, J., Gyurkovics, H., Schedl, P. and Ish-Horowicz, D. (1982). Genomic organization and functional analysis of a deletion variant of the 87A7 heat shock locus of Drosophila melanogaster. J. Mol. Biol. *155*, 267–280.

Weaver, R. F. and Weissman, C. (1979). Mapping of RNA by a modification of the Berk–Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucl. Acids Res. *7*, 1175–1193.

Wu, C. (1980). The 5′ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature *286*, 854–860.